Clinical Interventions in Aging

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Machine Learning-Based Prediction of Post-PKP Frailty: A Retrospective Cohort Study

Dingjun Xu*, Ziwei Fan*, Zhiyuan Li*, Mengxian Jia, Xiang Fang, Yizhe Shen, Quan Zhou, Changnan Xie, Honglin Teng

Department of Orthopedics (Spine Surgery), The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, 325035, People's Republic of China

*These authors contributed equally to this work

Correspondence: Changnan Xie; Honglin Teng, Email 541017595@qq.com; tenghonglin@wzhospital.cn

Background: Frailty and osteoporotic vertebral compression fractures (OVCFs) exhibit bidirectional causality, yet the impact of percutaneous kyphoplasty (PKP) on frailty progression remains unclear. This study developed machine learning (ML) models to predict post-PKP frailty and identify key predictors.

Methods: A retrospective cohort of 4599 PKP patients was categorized into frailty/non-frailty groups based on two-year follow-up. Variables included preoperative baseline data, imaging parameters (fracture number/segments, Genant classification, T2 hyperintensity), clinical characteristics (osteoporosis severity, Visual Analogue Scale scores, residual low back pain [LBP]), and surgical details. After data splitting (4:1 ratio), features were selected to train and optimize ML models, with performance evaluated via area under the curve (AUC). The ML model with the best performance was selected as our final model while using it for external validation. SHAP analysis determined predictor contributions.

Results: Key features (residual LBP, Genant classification, etc) informed model development. Hyperparameter optimization enhanced performance, with Extreme Gradient Boost achieving superior prediction (AUC 0.950, 95% CI 0.934–0.965). The model still maintains a good performance in the external test set, with an AUC of 0.845 (95% CI 0.805–0.884). SHAP identified residual LBP, Genant classification, and postoperative recumbency duration as top predictors.

Conclusion: ML models effectively predict post-PKP frailty, highlighting modifiable risk factors. Standardized anti-osteoporosis therapy, residual LBP prevention, and reduced postoperative recumbency may mitigate frailty risk.

Keywords: frailty, machine learning, osteoporosis, PKP, prognostic prediction

Introduction

Frailty, a clinical syndrome characterized by diminished physiological reserves and multisystem dysfunction, compromises the capacity to withstand minor stressors, thereby predisposing individuals to adverse health outcomes.[1,2] Although prevalent across all age groups, frailty demonstrates age-dependent epidemiology, with incidence rates positively correlating with advancing age.[3] Global demographic projections estimate the population aged ≥65 years will reach 2 billion by 2050, suggesting an impending escalation in frailty burden.[4] Substantial evidence links frailty to critical adverse outcomes including falls, increased hospitalization rates, malignancies, and mortality, which collectively deteriorate elderly health and impose substantial strain on healthcare systems.[5–9] Notably, frailty represents a dynamic and potentially reversible state except during terminal decline phases, underscoring the imperative to identify modifiable risk factors for targeted prevention and mitigation strategies.[10]

Osteoporotic vertebral compression fractures (OVCFs), a prevalent geriatric condition, have emerged as a critical public health concern due to their substantial morbidity burden in elderly populations.[11] Substantial evidence identifies frailty as a significant predictor of osteoporotic fractures in older adults.[12] Furthermore, OVCFs exhibit an accumulative effect on both frailty incidence and progression, establishing a bidirectional causal relationship between these two clinical

Received: 9 May 2025
Accepted: 4 September 2025
Published: 11 September 2025

entities.[13] However, the differential impacts of post-OVCF intervention strategies on frailty development remain under-explored in current clinical research.

Since its inception in the 1990s, percutaneous kyphoplasty (PKP) has been refined through decades of clinical practice and is now established as a gold-standard intervention for OVCFs.[14,15] PKP achieves rapid analgesia, stabilizes fractured vertebral bodies, and promotes early mobilization, thereby reducing immobilization-related complications such as pneumonia and thromboembolism.[16–18] Nevertheless, procedure-associated adverse events—including refractures, cement leakage, neural compromise, and adjacent-segment vertebral fractures—pose non-negligible risks.[19–22] Currently, whether perioperative PKP affect frailty development remains unknown.

Machine learning (ML) has been widely implemented in disease prognosis prediction, with recent advances showing superior performance in forecasting spinal surgery outcomes compared to traditional statistical models.[23] This study applied machine learning to identify key perioperative factors influencing post-PKP frailty development, providing actionable insights for optimizing surgical approach selection and technique modification to reduce frailty incidence.

Methods

Patients and Study Design

The study conducted a retrospective analysis of a dataset obtained from our hospital, spanning from April 2013 to April 2023. The external validation dataset covers April 2017 to April 2022 from Wenzhou Hospital of Integrated Traditional Chinese and Western Medicine. All patients in this study signed an informed consent form prior to undergoing procedure. And ethics approval was obtained from Ethics Committee in Clinical Research (ECCR) of the First Affiliated Hospital of Wenzhou Medical University (KU2025-R062) and Wenzhou Hospital of Integrated Traditional Chinese and Western Medicine (2025-L134; for external validation). The study was conducted according to the Declaration of Helsinki.

Inclusion criteria were as follows: (1) OVCFs resulting from low-energy trauma (including sprains, heavy lifting, or falls) with MRI-confirmed acute vertebral fractures; (2) T score of bone mineral density (BMD) < -2.5; (3) Patients underwent PKP surgical intervention.

Exclusion criteria were as follows: (1) history of spinal surgery; (2) presence of frailty at admission; (3) pathological vertebral fractures secondary to malignancies or infections; (4) fractures with spinal cord compression accompanied by neurological deficits (eg, numbness and/or muscle weakness); (5) incomplete medical records; (6) cognitive impairment precluding independent communication; (7) incomplete follow-up data.

Baseline Data and Primary Outcome

In the study, population characteristics [age, sex, weight, height, body mass index (BMI), smoking, drinking, hypertension, diabetes, and hyperlipidemia], perioperative imaging features (severe osteoporosis, the number of fractured vertebrae, segmental classification of spine fractures, Genant classification, T2 hyperintensity in soft tissue, restoration of anterior vertebral height, and kyphosis, preoperative and immediate postoperative Visual Analogue Scale (VAS) scores, anesthetic features [American Society of Anesthesiologists Physical Status Classification System (ASA classification), type of anesthesia, and anesthesia duration], and surgical and perioperative information (surgical time, intraoperative blood loss, volume of bone cement used, bone cement extravasation, duration of postoperative recumbency, and duration of hospitalization). In addition, residual low back pain (LBP) was evaluated according to following criteria during the one-year follow-up period: defined as persistent pain occurring in the original pain site postoperatively, with a VAS score of ≥ 4.

The 2-year interval sufficiently meets both PKP prognostic assessment criteria and longitudinal frailty monitoring requirements.[24–27] So the primary outcome was postoperative frailty after PKP at two-years follow-up. Frailty was diagnosed according to the Fried criteria.[1]

Radiographic Characteristics

(1) Severe osteoporosis
Dual-energy X-ray absorptiometry (DXA)-derived T-scores exceeding −3.5.[28]
(2) Genant classification[29]
(3) T2 hyperintensity in soft tissue
Hyperintensity in the dorsal soft tissues at or below the fracture level was observed on preoperative T2-weighted MRI sequences.
(4) Restoration of anterior vertebral height
The anterior vertebral body height of the fractured vertebra was measured on sagittal radiographs both preoperatively and postoperatively, with the difference quantified as the restoration.
(5) Kyphosis
Kyphosis angle was quantified on sagittal radiographs by measuring the angle between the superior endplate of the adjacent superior vertebra and the inferior endplate of the adjacent inferior vertebra. Kyphosis was defined as a postoperative-to-3-month angular change >10°.

Feature Selection

The guidelines for developing ML models were followed for the present study.[30] The patient cohort that underwent PKP was divided into training and testing subsets with a ratio of 4:1. Consequently, we obtained a Dataset 1 and a Dataset 2, with 3679 and 920 patients respectively.

To alleviate the potential bias caused by our limited sample size, the least absolute shrinkage and selection operator (Lasso) regression was employed to select features.[23] Features were selected based on the λ.1-s.e. (standard error) criterion.[31] And λ serves as the penalty parameter, determining the degree to which the function is shrunk. Lasso regression coefficient of each variable was shown in Supplementary Table 1.

Model Training and Testing

The model was established using five ML algorithms below: XGBoost (Extreme Gradient Boost), LightGBM (Light Gradient Boosting Machine), AdaBoost (Adaptive Boosting), GBDT (Gradient Boosting Decision Tree), and SVM (Support Vector Machine) with the RBF (radial basic function) kernel. Based on the area under the receiver operating characteristics curve (AUC-ROC), the optimal hyperparameters were optimized by grid search to improve the predictive performance of our models.[23,32] And the hyperparameters for each model were shown in Supplementary Table 2.

The testing set was used to evaluate the predictive power of each model. These assessments were performed using various metrics, including AUC-ROC, accuracy, sensitivity, specificity, negative predictive value (NPV), positive predictive value (PPV). In brief, discrimination was assessed using AUC-ROC for binary classification.[33] Calibration of the ML models was evaluated by calibration plots. The Brier score was recognized as a measurement of the models' discrimination and calibration.[34] Additionally, the net benefit of each model was analyzed using decision curve analysis.[35] Ultimately, utilizing the ML model that demonstrated the best predictive performance, we ranked and visualized the contribution of each predictor to the model's predictive performance using Shapley Additive Explanations (SHAP) value analysis.[36]

Statistical Analysis

In the study, mean ± SD was used to represent continuous variables and frequency [proportion (%)] was used to represent categorical variables. Intergroup comparisons were employed using Student's unpaired *t*-test and Chi-Square tests. Comparisons with values of $P < 0.05$ were considered statistically significant. Lasso regression was conducted using R programming language (4.2.3). And AdaBoost, GBDT, and SVM were trained using Python (scikit-learn = 1.1.3). XGBoost, LightGBM, and SHAP value analyses were respectively trained using Python (xgboost = 2.0.1), Python (lightgbm = 3.2.1), and Python (shap = 0.43.0).

Result

Patient Data

This study enrolled 4599 patients from a single-center cohort and 561 from an external validation cohort who underwent PKP. Comprehensive demographic profiles and clinical characteristics of both cohorts are summarized in Table 1.

Table 1 Patients' Demographic and Clinical Characteristics

Variable	Mean ± SD or Frequency [Proportion (%)]		*P*
	Single-Center Data (n = 4599)	External Data (n = 561)	
Age (years)	70.990±6.152	70.746±5.888	0.374[a]
Sex			0.737[b]
Male	1118 (24.310)	140 (24.955)	
Female	3481 (75.690)	421 (75.045)	
Weight (kg)	58.050±11.027	57.296±8.824	0.119[a]
Height (cm)	157.887±7.575	158.617±7.450	0.032[a]
BMI (kg/m^2)	23.954±4.587	22.882±3.907	0.004[a]
Smoking			0.496[b]
Yes	357 (7.763)	39 (6.952)	
No	4242 (92.237)	522 (93.048)	
Drinking			0.322[b]
Yes	331 (7.197)	34 (6.061)	
No	4268 (92.803)	527 (93.939)	
Hypertension			0.139[b]
Yes	2078 (45.184)	235 (41.889)	
No	2521 (54.816)	326 (58.111)	
Diabetes			0.584[b]
Yes	847 (18.417)	98 (17.469)	
No	3752 (81.583)	463 (82.531)	
Hyperlipidemia			0.017[b]
Yes	456 (9.915)	38 (6.774)	
No	4143 (90.085)	523 (93.226)	
Perioperative clinical characteristics			
VAS score	6.282±1.083	6.201±1.082	0.095[a]
Severe osteoporosis			0.547[b]
Yes	1597 (34.725)	202 (36.007)	
No	3002 (65.275)	359 (63.993)	
The number of fractured vertebrae	1.102±0.449	1.064±0.315	0.056[a]
Segmental classification of spine fractures			0.003[b]
Thoracic vertebrae	313 (6.806)	38 (6.774)	
Lumbar vertebrae	631 (13.720)	110 (19.608)	
Thoracolumbar spine	3492 (75.930)	394 (70.232)	
Thoracic and lumbar vertebrae	163 (3.544)	19 (3.387)	
Genant classification			0.173[b]
Grade 1	2652 (57.665)	341 (60.784)	
Grade 2	1344 (29.224)	161 (28.699)	
Grade 3	603 (13.112)	59 (10.517)	
ASA classification			0.986[b]
Grade I	617 (13.416)	74 (13.191)	
Grade II	2961 (64.384)	363 (64.706)	
Grade III	1021 (22.200)	124 (22.103)	
T2 hyperintensity in soft tissue			0.113[b]
Yes	1648 (35.834)	182 (32.442)	
No	2951 (64.166)	379 (67.558)	

(*Continued*)

Table 1 (Continued).

Variable	Mean ± SD or Frequency [Proportion (%)]		*P*
	Single-Center Data (n = 4599)	External Data (n = 561)	
Perioperative information			
Type of anesthesia			0.016[b]
Regional	2649 (57.599)	353 (62.923)	
General	1950 (42.401)	208 (37.077)	
Anesthesia duration (min)	50.940±33.040	50.920±29.755	0.989[a]
Surgical time (min)	53.708±34.163	57.762±45.819	0.043[a]
Intraoperative blood loss (mL)	12.203±7.293	12.645±7.000	0.174[a]
Volume of bone cement used (mL)	3.743±0.756	3.705±0.710	0.231[a]
Bone cement extravasation			0.522[b]
Yes	1870 (40.661)	236 (42.068)	
No	2729 (59.339)	325 (57.932)	
Duration of postoperative recumbency (days)	1.596±0.953	1.490±0.917	0.114[a]
Duration of hospitalization (days)	6.352±4.427	6.462±4.049	0.575[a]
Restoration of anterior vertebral height (cm)	3.163±0.991	3.044±1.010	0.007[a]
Kyphosis			0.532[b]
Yes	932 (20.265)	120 (21.390)	
No	3667 (79.735)	441 (78.610)	
Postoperative VAS score	2.383±0.864	2.321±0.795	0.083[a]
Residual LBP			0.018[b]
Yes	401 (8.719)	66 (11.765)	
No	4198 (91.281)	495 (88.235)	
Primary outcome			
Frailty			<0.001[b]
Yes	821 (17.852)	192 (34.225)	
No	3778 (82.148)	369 (65.775)	

Notes: Mean ± SD is used to represent continuous variables and frequency [proportion (%)] is used to represent binary categorical variables. The statistical methods indicated by the superscripted letters were as follows: [a]Student's unpaired *t*-test; [b]Chi-Square test.
Abbreviations: SD, standard deviation; BMI, body mass index; VAS, Visual Analogue Scale; ASA classification, American Society of Anesthesiologists Physical Status Classification System; LBP, low back pain.

Baseline Predictive Performance

Based on the λ.1-s.e. criterion, seven features were selected using Lasso regression (Figure 1 and Supplementary Table 1) for the subsequent model developing. The resulting features encompassed residual LBP, severe osteoporosis, Genant classification, T2 hyperintensity in soft tissue, the number of fractured vertebrae, and the duration of postoperative recumbency.

The default hyperparameters for each model were shown in Supplementary Table 2. The comparative analysis of the baseline predictive performance of the ML models in the validation set (10-fold cross-validation) was shown in Supplementary Figure 1 and Supplementary Table 3. As shown in Figure 2 and Table 2, the XGBoost model demonstrated the superior baseline predictive performance in testing sets, achieving an AUC-ROC [95% confidence interval (CI)] of 0.946 (0.930–0.963), an accuracy of 0.882, a sensitivity of 0.885, a specificity of 0.881, a PPV of 0.647, a NPV of 0.969, and a Brier score of 0.072. Furthermore, the calibration curve of the XGBoost model was fitting the perfect situation (Supplementary Figure 2). And the decision curves indicated that XGBoost, GBDT, and SVM models performed well in clinical usefulness (Supplementary Figure 3).

The Performance of ML Models with the Optimal Hyperparameters

The hyperparameters of each model were optimized by grid search. And the optimized hyperparameters for each model were shown in Supplementary Table 2. The comparative analysis of the ML predictive performance after hyperparameter optimization in the validation set (10-fold cross-validation) was shown in Supplementary Figure 4 and Supplementary Table 4. Interestingly,

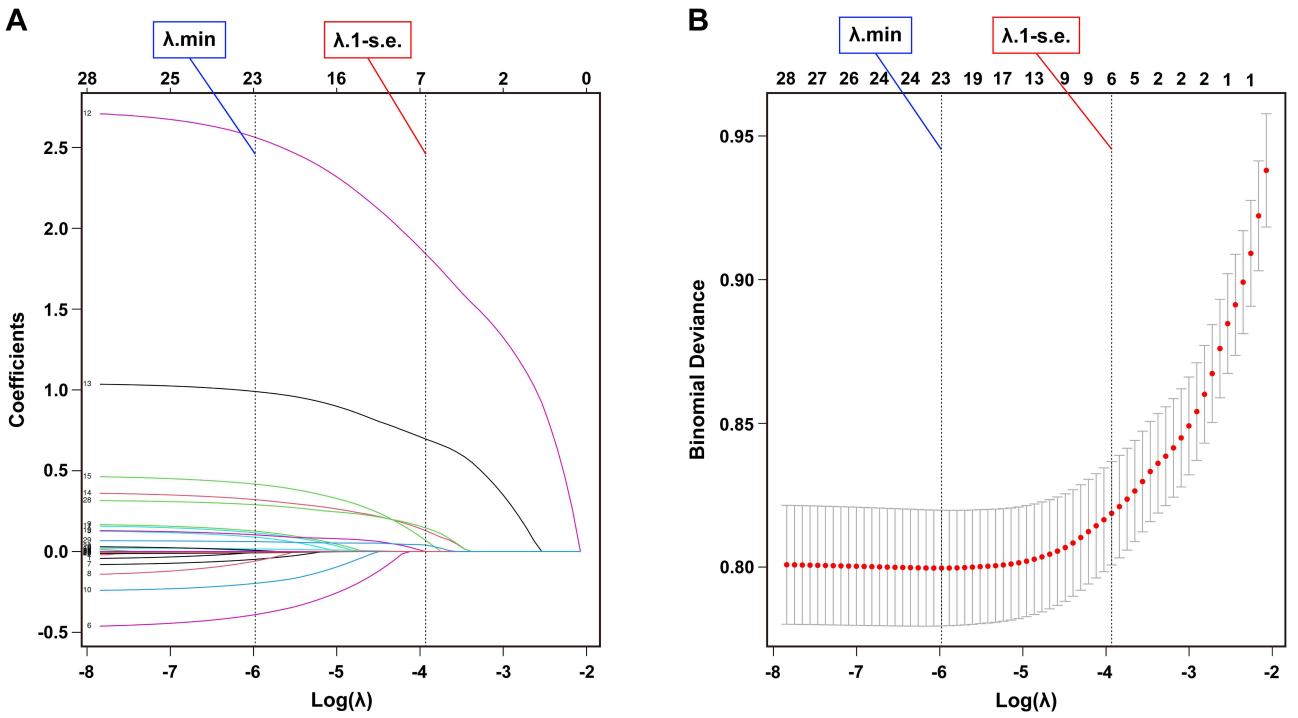

Figure 1 (**A**) The coefficient profiles of the 29 features were analyzed by Lasso regression. (**B**) Based on the λ.1-s.e. criterion, there were seven features selected using Lasso regression. The two dotted vertical lines were drawn at the optimal scores by minimum criterion (λ.min) and λ.1-s.e. criterion, respectively. At the λ.1-s.e. criterion, the selected features included residual LBP, severe osteoporosis, Genant classification, T2 hyperintensity in soft tissue, the number of fractured vertebrae, and the duration of postoperative recumbency.
Abbreviations: Lasso, the least absolute shrinkage and selection operator; s.e., standard error; LBP, low back pain.

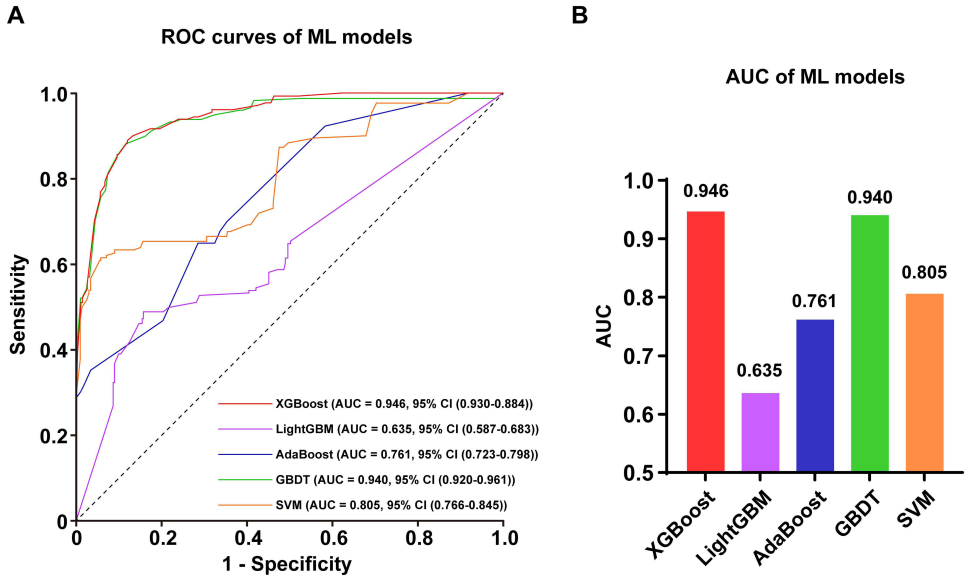

Figure 2 The baseline predictive performance for ML models in the testing dataset. (**A**) The ROC curves for ML models. (**B**) The AUCs for ML models.
Abbreviations: ML, machine learning; ROC, receiver operating characteristics; AUC, the area under curve; XGBoost, Extreme Gradient Boost; LightGBM, Light Gradient Boosting Machine; AdaBoost, Adaptive Boosting; GBDT, Gradient Boosting Decision Tree; SVM, Support Vector Machine.

after hyperparameter optimizing, the XGBoost model still demonstrated the best predictive performance in the testing set, with an AUC-ROC (95% CI) of 0.950 (0.934–0.965), an accuracy of 0.884, a sensitivity of 0.890, a specificity of 0.882, a PPV of 0.651, a NPV of 0.970, and a Brier score of 0.066 (Figure 3 and Table 3). And the calibration curve of the hyperparameter-optimized

Table 2 Comparison of Brier Score, AUC-ROC, Accuracy, Sensitivity, Specificity, PPV, NPV of the Baseline Prediction Performance for ML Models in Testing Sets

Model	Brier score	AUC-ROC (95% CI)	Accuracy	Sensitivity	Specificity	PPV	NPV
XGBoost	0.072	0.946 (0.930–0.963)	0.882	0.885	0.881	0.647	0.969
LightGBM	0.095	0.635 (0.587–0.683)	0.773	0.489	0.843	0.434	0.870
AdaBoost	0.191	0.761 (0.723–0.798)	0.702	0.648	0.715	0.360	0.892
GBDT	0.072	0.940 (0.920–0.961)	0.883	0.885	0.882	0.649	0.969
SVM	0.284	0.805 (0.766–0.845)	0.878	0.615	0.943	0.727	0.909

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

XGBoost model was very close to the ideal calibrated curve (Supplementary Figure 5). As shown in Supplementary Figure 6, XGBoost, LightGBM, GBDT, and SVM models possessed excellent clinical benefit.

The performance of the hyperparameter-optimized XGBoost model, trained as described, remained stable in the external test set (AUC: 0.845, 95% CI (0.805–0.884)) (Figure 4 and Table 4). Moreover, in the external test set, the calibration curve of the hyperparameter-optimized XGBoost model also showed strong alignment with the ideal curve (Supplementary Figure 7) and the model still possessed excellent clinical benefit (Supplementary Figure 8).

In addition, based on the hyperparameter-optimized XGBoost model, we identified the importance ranking of each Lasso regression-selected feature using the SHAP value analysis. The results were shown in descending order as follows: residual LBP, Genant classification, the duration of postoperative recumbency, severe osteoporosis, T2 hyperintensity in soft tissue, and the number of fractured vertebrae (Figure 5).

Discussion

In recent years, linear regression and nomograms have been extensively utilized for clinical prognostic prediction and have demonstrated favorable performance in predicting postoperative outcomes following PKP.[22,37] Machine learning, which employs extensive datasets to optimize algorithmic performance, demonstrates superior reliability, objectivity, and reproducibility when processing large-scale clinical data[38] Emerging evidence suggests that machine learning algorithms

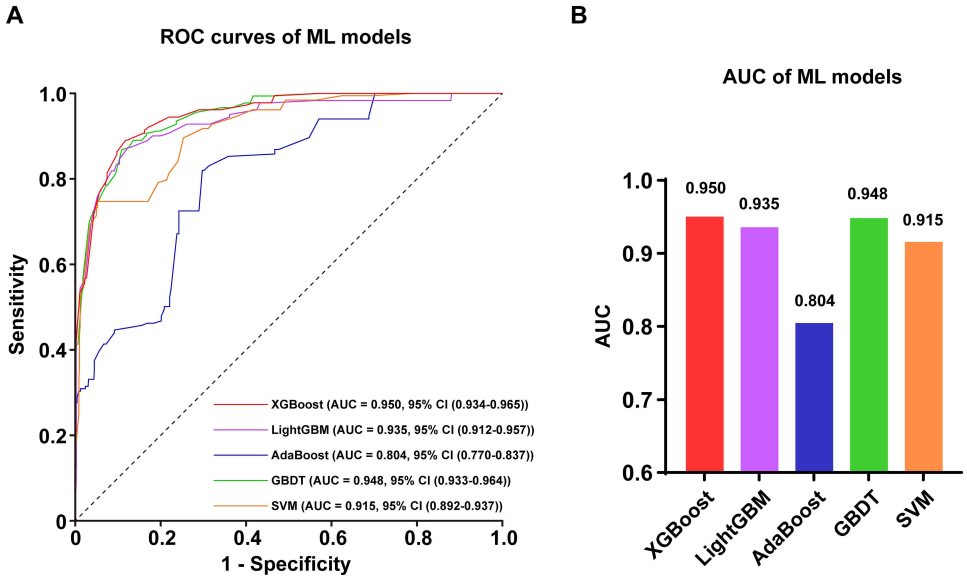

Figure 3 The predictive efficiency of ML models with the optimized hyperparameters using the grid search. (**A**) The ROC curves for hyperparameter-optimizing ML models. (**B**) The AUCs for hyperparameter-optimizing ML models.

Abbreviations: ML, machine learning; ROC, receiver operating characteristics; AUC, the area under curve; XGBoost, Extreme Gradient Boost; LightGBM, Light Gradient Boosting Machine; AdaBoost, Adaptive Boosting; GBDT, Gradient Boosting Decision Tree; SVM, Support Vector Machine.

https://doi.org/10.2147/CIA.S537151

Table 3 Comparison of Brier Score, AUC-ROC, Accuracy, Sensitivity, Specificity, PPV, NPV of Each ML Model After Optimizing Hyperparameters Using Grid Search in Testing Sets

Model	Brier score	AUC-ROC (95% CI)	Accuracy	Sensitivity	Specificity	PPV	NPV
XGBoost	0.066	0.950 (0.934–0.965)	0.884	0.890	0.882	0.651	0.970
LightGBM	0.113	0.935 (0.912–0.957)	0.896	0.819	0.915	0.703	0.953
AdaBoost	0.191	0.804 (0.770–0.837)	0.725	0.819	0.702	0.404	0.940
GBDT	0.068	0.948 (0.933–0.964)	0.870	0.890	0.864	0.618	0.970
SVM	0.078	0.915 (0.892–0.937)	0.905	0.731	0.949	0.778	0.935

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

outperform conventional statistical methods in predicting clinical outcomes following PKP.[39] However, previous studies failed to implement variable selection prior to model development, potentially leading to overfitting and multicollinearity issues. LASSO regression effectively addresses these limitations. In this study, we employed LASSO regression with the λ.1-standard error (λ.1-s.e.) criterion for feature selection, followed by predictive model construction using the identified variables. Our findings identify six significant predictors of postoperative frailty: residual LBP, Genant classification, duration of postoperative recumbency, severe osteoporosis, T2 hyperintensity in paraspinal soft tissues, and number of fractured vertebrae. This model facilitates personalized postoperative rehabilitation protocols tailored to individual patients, providing evidence-based guidance to interprofessional rehabilitation teams to mitigate frailty risk in PKP populations.

Notably, residual LBP was identified as the most significant predictor of frailty following PKP, consistent with previous studies demonstrating the association between LBP and frailty.[40] Residual LBP persistently impairs post-operative recovery and disrupts sleep patterns, with prolonged pain and insomnia contributing to anxiety and depression.[41] Furthermore, LBP alters gait characteristics, reduces movement coordination, and induces kinesiophobia. Pavel et al demonstrated that LBP may compromise diaphragmatic function, thereby affecting trunk stability.[42] These

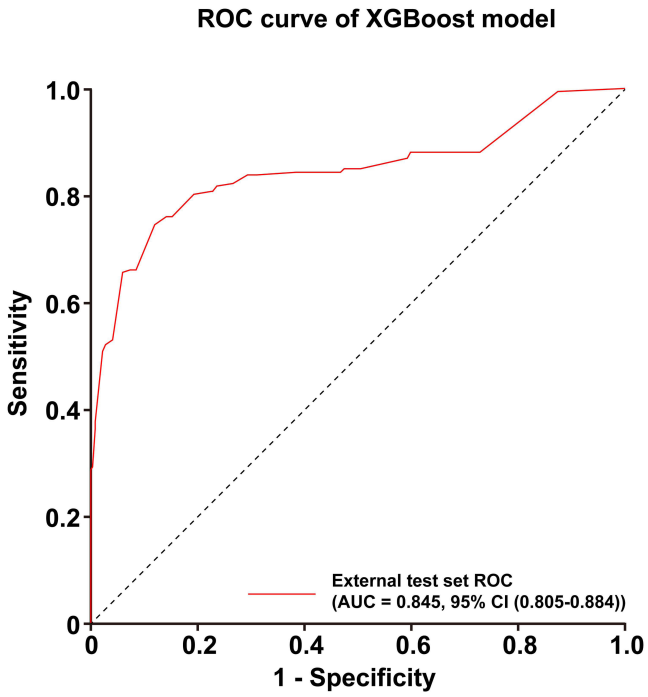

Figure 4 The ROC curve of the hyperparameter-optimized XGBoost model for predicting post-PKP frailty in the external test set.
Abbreviations: ROC, receiver operating characteristics; AUC, the area under curve; XGBoost, Extreme Gradient Boost; CI, confidence interval.

Table 4 The Performance of Hyperparameter-Optimized XGBoost Model for Predicting Post-PKP Frailty in the External Test Set

Brier Score	AUC-ROC (95% CI)	Accuracy	Sensitivity	Specificity	PPV	NPV
0.142	0.845 (0.805–0.884)	0.824	0.760	0.856	0.734	0.873

Abbreviations: AUC-ROC, the area under the receiver operating characteristics curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; XGBoost, Extreme Gradient Boost.

factors collectively reduce physical activity levels and increase sedentary behavior, ultimately elevating frailty risk.[40] Progressive frailty inevitably diminishes health-related quality of life (HRQoL).[43] We emphasize that enhanced post-operative pain management and prevention of residual LBP could reduce frailty incidence and provide long-term clinical benefits for patients.

The Genant classification and number of fractured vertebrae, as preoperative indicators, reflect the severity of vertebral fractures. Since its introduction in 1993,[29] the Genant classification has been widely utilized for assessing OVCFs. Higher grades indicate more severe vertebral deformities, characterized by greater buckling of the vertebral cortex or endplates.[44] Multilevel vertebral fractures suggest extensive trauma to the spinal system. Severe Genant classifications and a higher number of fractured vertebrae correlate with poorer surgical outcomes, increased incidence of kyphosis, and significant associations with frailty.[45] Previous studies have also demonstrated the relationship between severe osteoporotic fractures, multilevel vertebral fractures, and frailty, which aligns with our findings.[46] Notably, this study found no correlation between kyphosis progression and frailty, potentially because single-level kyphosis progression does not reflect overall spinal deformity.

T2 hyperintensity in soft tissue typically indicates muscle edema, hemorrhage, or inflammatory exudates, reflecting functional impairment of the lumbar musculature. Although PKP allows early postoperative ambulation, muscle injury may restrict patient mobility and prolong recovery. Additionally, prolonged bed rest during hospitalization increases frailty risk, underscoring the importance of early mobilization to prevent disuse atrophy.

Substantial evidence confirms that osteoporosis exacerbates frailty syndrome in older adults through systemic chronic inflammation, nutritional alterations, and endocrine system dysregulation. Chronic inflammation acts as a pivotal determinant of frailty, exerting direct effects while also mediating indirect impacts via intermediate mechanisms.[47,48] Consequently, timely anti-osteoporosis therapy is imperative for patients with severe osteoporosis following PKP.[49] Beyond pharmacological interventions, balanced physical activity and dietary modifications constitute essential components for preventing or reversing frailty. For patients with established frailty, treatment strategies should be tailored based on specific frailty domains.[50]

While our study demonstrates the potential of ML in predicting postoperative frailty, several limitations should be acknowledged. First, the study was retrospective in nature, and the data were collected from a single institution, which

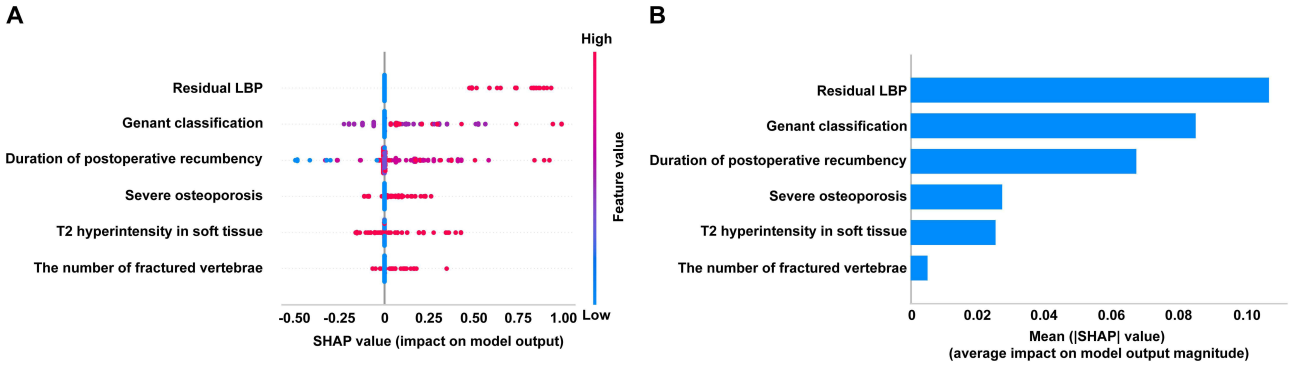

Figure 5 A SHAP analysis of the XGBoost model with the optimized hyperparameters. (**A**) A visualization for SHAP values of the Lasso-regression-selected features in each sample. The dots in red represented the higher SHAP values, while the blue ones represented the lower values. (**B**) The selected features were listed in descending order by the average absolute values of SHAP.
Abbreviations: SHAP, Shapley Additive Explanations; LBP, low back pain.

https://doi.org/10.2147/CIA.S537151

may limit the generalizability of the findings. We cannot exclude potential spectrum bias and selection bias in the tertiary medical center cohort, while unmeasured perioperative rehabilitation variables may potentially confound frailty outcomes. Future studies should validate our findings in larger, multicenter, and prospective cohorts. Second, the study focused on a specific PKP for OVCFs. The predictive performance of ML models may vary for other surgical interventions or patient populations. Third, while we identified key predictors of postoperative frailty, the underlying mechanisms remain unclear. Moreover, the internal dataset exhibited class imbalance that may introduce potential biases in ML model training and optimization—a limitation to be systematically addressed in our subsequent research. At last, this study employed a binary classification of frailty status (frail/non-frail) without stratifying the pre-frail subgroup for comparative analysis. Future research should explore the biological pathways linking these predictors to frailty development.

Conclusion

Our findings indicated that the application of ML algorithms, particularly the hyperparameter-optimized XGBoost model, can effectively predict postoperative frailty in patients underwent PKP. Furthermore, the key predictors were identified through SHAP value analysis. Standardized anti-osteoporosis treatment, prevention of residual postoperative low back pain, and reduction of duration of postoperative recumbency are critical factors in avoiding the occurrence of frailty following PKP.

Acknowledgments

We sincerely thank Dr. Guangxi Ma from Wenzhou Hospital of Integrated Traditional Chinese and Western Medicine for providing the external validation dataset.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Disclosure

All authors report no conflicts of interest in this work.

References

1. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56(3):M146–56. doi:10.1093/gerona/56.3.m146
2. Lipsitz LA. Dynamics of stability: the physiologic basis of functional health and frailty. *J Gerontol A Biol Sci Med Sci*. 2002;57(3):B115–25. doi:10.1093/gerona/57.3.b115
3. Fried LP, Ferrucci L, Darer J, Williamson JD, Anderson G. Untangling the concepts of disability, frailty, and comorbidity: implications for improved targeting and care. *J Gerontol A Biol Sci Med Sci*. 2004;59(3):255–263. doi:10.1093/gerona/59.3.m255
4. Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. *Lancet*. 2013;381(9868):752–762. doi:10.1016/s0140-6736(12)62167-9
5. Kojima G. Frailty as a predictor of future falls among community-dwelling older people: a systematic review and meta-analysis. *J Am Med Dir Assoc*. 2015;16(12):1027–1033. doi:10.1016/j.jamda.2015.06.018
6. Kojima G. Frailty as a predictor of hospitalisation among community-dwelling older people: a systematic review and meta-analysis. *J Epidemiol Community Health*. 2016;70(7):722–729. doi:10.1136/jech-2015-206978
7. Aaldriks AA, van der Geest LG, Giltay EJ, et al. Frailty and malnutrition predictive of mortality risk in older patients with advanced colorectal cancer receiving chemotherapy. *J Geriatr Oncol*. 2013;4(3):218–226. doi:10.1016/j.jgo.2013.04.001
8. Shamliyan T, Talley KM, Ramakrishnan R, Kane RL. Association of frailty with survival: a systematic literature review. *Ageing Res Rev*. 2013;12(2):719–736. doi:10.1016/j.arr.2012.03.001
9. Ensrud KE, Kats AM, Schousboe JT, et al. Frailty phenotype and healthcare costs and utilization in older women. *J Am Geriatr Soc*. 2018;66(7):1276–1283. doi:10.1111/jgs.15381
10. Puts MTE, Toubasi S, Andrew MK, et al. Interventions to prevent or reduce the level of frailty in community-dwelling older adults: a scoping review of the literature and international policies. *Age Ageing*. 2017;46(3):383–392. doi:10.1093/ageing/afw247
11. McCarthy J, Davis A. Diagnosis and management of vertebral compression fractures. *Am Fam Physician*. 2016;94(1):44–50.

12. Kojima G. Frailty as a predictor of fractures among community-dwelling older people: a systematic review and meta-analysis. *Bone*. 2016;90:116–122. doi:10.1016/j.bone.2016.06.009
13. Walters S, Chan S, Goh L, Ong T, Sahota O. The prevalence of frailty in patients admitted to hospital with vertebral fragility fractures. *Curr Rheumatol Rev*. 2016;12(3):244–247.
14. Filippiadis DK, Marcia S, Masala S, Deschamps F, Kelekis A. Percutaneous vertebroplasty and kyphoplasty: current status, new developments and old controversies. *Cardiovasc Intervent Radiol*. 2017;40(12):1815–1823. doi:10.1007/s00270-017-1779-x
15. Liu JT, Liao WJ, Tan WC, et al. Balloon kyphoplasty versus vertebroplasty for treatment of osteoporotic vertebral compression fracture: a prospective, comparative, and randomized clinical study. *Osteoporos Int*. 2010;21(2):359–364. doi:10.1007/s00198-009-0952-8
16. Brower RG. Consequences of bed rest. *Crit Care Med*. 2009;37(10 Suppl):S422–8. doi:10.1097/CCM.0b013e3181b6e30a
17. Rousing R, Andersen MO, Jespersen SM, Thomsen K, Lauritsen J. Percutaneous vertebroplasty compared to conservative treatment in patients with painful acute or subacute osteoporotic vertebral fractures: three-months follow-up in a clinical randomized study. *Spine*. 2009;34(13):1349–1354. doi:10.1097/BRS.0b013e3181a4e628
18. Parry SM, Puthucheary ZA. The impact of extended bed rest on the musculoskeletal system in the critical care environment. *Extrem Physiol Med*. 2015;4:16. doi:10.1186/s13728-015-0036-7
19. Qi Z, Zhao S, Li H, Wen Z, Chen B. A study on vertebral refracture and scoliosis after percutaneous kyphoplasty in patients with osteoporotic vertebral compression fractures. *J Orthop Surg Res*. 2024;19(1):302. doi:10.1186/s13018-024-04779-9
20. Rose LD, Bateman G, Ahmed A. Clinical significance of cement leakage in kyphoplasty and vertebroplasty: a systematic review. *Eur Spine J*. 2024;33(4):1484–1489. doi:10.1007/s00586-023-08026-3
21. Lee YH, Chen PQ, Wu CT. Delayed-onset radiculopathy caused by a retropulsed bone fragment after percutaneous kyphoplasty: report of four cases and literature review. *BMC Musculoskelet Disord*. 2022;23(1):529. doi:10.1186/s12891-022-05472-w
22. Tao W, Biao W, Xingmei C, et al. Predictive factors for adjacent vertebral fractures after percutaneous kyphoplasty in patients with osteoporotic vertebral compression fracture. *Pain Physician*. 2022;25(5):E725–e732.
23. Song J, Li J, Zhao R, Chu X. Developing predictive models for surgical outcomes in patients with degenerative cervical myelopathy: a comparison of statistical and machine learning approaches. *Spine J*. 2024;24(1):57–67. doi:10.1016/j.spinee.2023.07.021
24. Boonen S, Van Meirhaeghe J, Bastian L, et al. Balloon kyphoplasty for the treatment of acute vertebral compression fractures: 2-year results from a randomized trial. *J Bone Miner Res*. 2011;26(7):1627–1637. doi:10.1002/jbmr.364
25. Garfin SR, Buckley RA, Ledlie J; Group ftBKO. Balloon kyphoplasty for symptomatic vertebral body compression fractures results in rapid, significant, and sustained improvements in back pain, function, and quality of life for elderly patients. *Spine*. 2006;31(19):2213–2220. doi:10.1097/01.brs.0000232803.71640.ba
26. LeBoff MS, Greenspan SL, Insogna KL, et al. The clinician's guide to prevention and treatment of osteoporosis. *Osteoporos Int*. 2022;33(10):2049–2102. doi:10.1007/s00198-021-05900-y
27. Ho LYW, Cheung DSK, Kwan RYC, Wong ASW, Lai CKY. Factors associated with frailty transition at different follow-up intervals: a scoping review. *Geriatr Nurs*. 2021;42(2):555–565. doi:10.1016/j.gerinurse.2020.10.005
28. Johnston CB, Dagar M. Osteoporosis in older adults. *Med Clin North Am*. 2020;104(5):873–884. doi:10.1016/j.mcna.2020.06.004
29. Genant HK, Wu CY, van Kuijk C, Nevitt MC. Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res*. 1993;8(9):1137–1148. doi:10.1002/jbmr.5650080915
30. Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res*. 2016;18(12):e323. doi:10.2196/jmir.5870
31. Zhou D, Liu X, Wang X, et al. A prognostic nomogram based on LASSO Cox regression in patients with alpha-fetoprotein-negative hepatocellular carcinoma following non-surgical therapy. *BMC Cancer*. 2021;21(1):246. doi:10.1186/s12885-021-07916-3
32. Gao L, Cao Y, Cao X, et al. Machine learning-based algorithms to predict severe psychological distress among cancer patients with spinal metastatic disease. *Spine J*. 2023;23(9):1255–1269. doi:10.1016/j.spinee.2023.05.009
33. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925–1931. doi:10.1093/eurheartj/ehu207
34. Karhade AV, Shah AA, Bono CM, et al. Development of machine learning algorithms for prediction of mortality in spinal epidural abscess. *Spine J*. 2019;19(12):1950–1959. doi:10.1016/j.spinee.2019.06.024
35. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565–574. doi:10.1177/0272989x06295361
36. Wang K, Tian J, Zheng C, et al. Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. *Comput Biol Med*. 2021;137:104813. doi:10.1016/j.compbiomed.2021.104813
37. Zhang A, Lin Y, Kong M, et al. A nomogram for predicting the risk of new vertebral compression fracture after percutaneous kyphoplasty. *Eur J Med Res*. 2023;28(1):280. doi:10.1186/s40001-023-01235-y
38. DeVries Z, Hoda M, Rivers CS, et al. Development of an unsupervised machine learning algorithm for the prognostication of walking ability in spinal cord injury patients. *Spine J*. 2020;20(2):213–224. doi:10.1016/j.spinee.2019.09.007
39. Ma Y, Lu Q, Yuan F, Chen H. Comparison of the effectiveness of different machine learning algorithms in predicting new fractures after PKP for osteoporotic vertebral compression fractures. *J Orthop Surg Res*. 2023;18(1):62. doi:10.1186/s13018-023-03551-9
40. Coyle PC, Sions JM, Velasco T, Hicks GE. Older adults with chronic low back pain: a clinical population vulnerable to frailty? *J Frailty Aging*. 2015;4(4):188–190. doi:10.14283/jfa.2015.75
41. Gerhart JI, Burns JW, Post KM, et al. Relationships between sleep quality and pain-related factors for people with chronic low back pain: tests of reciprocal and time of day effects. *Ann Behav Med*. 2017;51(3):365–375. doi:10.1007/s12160-016-9860-2
42. Kolar P, Sulc J, Kyncl M, et al. Postural function of the diaphragm in persons with and without chronic low back pain. *J Orthop Sports Phys Ther*. 2012;42(4):352–362. doi:10.2519/jospt.2012.3830
43. Kim HJ, Jun B, Lee HW, Kim SH. Influence of frailty status on the health-related quality of life in older patients with chronic low back pain: a retrospective observational study. *Qual Life Res*. 2024;33(7):1905–1913. doi:10.1007/s11136-024-03658-4
44. Compston J, Genant H. Epidemiology and diagnosis of postmenopausal osteoporosis. In: Rizzoli R, editor. *Atlas of Postmenopausal Osteoporosis*. Springer Healthcare Ltd.; 2010:33–60.

45. Kado DM, Miller-Martinez D, Lui LY, et al. Hyperkyphosis, kyphosis progression, and risk of non-spine fractures in older community dwelling women: the study of osteoporotic fractures (SOF). *J Bone Miner Res*. 2014;29(10):2210–2216. doi:10.1002/jbmr.2251
46. Kim HJ, Park S, Park SH, et al. Prevalence of frailty in patients with osteoporotic vertebral compression fracture and its association with numbers of fractures. *Yonsei Med J*. 2018;59(2):317–324. doi:10.3349/ymj.2018.59.2.317
47. Calvani R, Martone AM, Marzetti E, et al. Pre-hospital dietary intake correlates with muscle mass at the time of fracture in older Hip-fractured patients. *Front Aging Neurosci*. 2014;6:269. doi:10.3389/fnagi.2014.00269
48. Chen X, Mao G, Leng SX. Frailty syndrome: an overview. *Clin Interv Aging*. 2014;9:433–441. doi:10.2147/cia.S45300
49. Greco EA, Pietschmann P, Migliaccio S. Osteoporosis and sarcopenia increase frailty syndrome in the elderly. *Front Endocrinol*. 2019;10:255. doi:10.3389/fendo.2019.00255
50. Dawson A, Dennison E. Measuring the musculoskeletal aging phenotype. *Maturitas*. 2016;93:13–17. doi:10.1016/j.maturitas.2016.04.014