

Machine Learning and Logistic Regression Models Incorporating Clinical and CT-Derived Muscle-Fat Parameters for Predicting Geriatric Hip Fracture: A Prospective Study

Dandan Ye

Quanzhou Orthopedic-Traumatological Hospital

Xiaoxiong Lin

Quanzhou Orthopedic-Traumatological Hospital

Zhiwei Xu

Quanzhou Orthopedic-Traumatological Hospital

Jiahong Huang

Quanzhou Orthopedic-Traumatological Hospital

Jinquan Lin

Quanzhou Orthopedic-Traumatological Hospital

Zhenbin Wu

wuzhenbin2026@163.com

Quanzhou Orthopedic-Traumatological Hospital

Article

Keywords: Hip fracture, Skeletal muscle index, Machine learning, Logistic regression, Nomogram

Posted Date: October 11th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10956815/v1>

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

Additional Declarations: No competing interests reported.

Abstract

Hip fractures cause substantial morbidity and mortality in older adults. Sarcopenia is a key modifiable risk factor, yet few models integrate CT-derived muscle-fat metrics with clinical data. This study aimed to develop a risk prediction model for elderly hip fractures, compare multiple machine learning algorithms, and convert the optimal model into an interpretable nomogram. This prospective cohort enrolled patients aged ≥ 60 years undergoing hip CT, including fracture cases and controls. Muscle/fat area, mean attenuation, and skeletal muscle index were measured at femoral head, lesser trochanter, and L3. The cohort was split 7:3 into training and validation sets. Eight classifiers were built using Boruta-selected features, with performance assessed by AUC, calibration, and decision curve analysis (DCA). The logistic regression (LR) achieved the highest AUC (0.779). Multivariable regression retained eight independent predictors: age, elevated glucose, osteoporosis, fat area_{1/2}, muscle mean₂, SMI₁, and SMI₃. The nomogram incorporating these variables showed good discrimination (training AUC = 0.826; validation AUC = 0.817), favorable calibration (Hosmer-Lemeshow $P = 0.647$ and 0.412 , respectively), and clinical net benefit on DCA. LR performed comparably or better than other ML algorithms. The nomogram offers a simple, interpretable tool for individualized hip fracture risk stratification using eight routinely available parameters.

Introduction

Hip fracture carries substantial morbidity, disability, and mortality worldwide(1). In China, the age-standardized incidence of hip fracture in older adults exceeds 110 per 100,000 person-years, with over one million new cases annually(2). Diabetes, sarcopenia, and bariatric surgery have recently been recognized as emerging risk factors for fragility fractures(3). Sarcopenia increases fall propensity through biomechanical pathways and reduces hip bone strength via systemic metabolic effects, synergistically elevating fracture risk in the elderly(4). Most consensus guidelines recommend that sarcopenia assessment encompass three core dimensions: muscle mass, muscle strength, and physical performance(5). Computed tomography (CT)-derived skeletal muscle area (SMA) and skeletal muscle index (SMI, defined as SMA divided by height squared) are validated quantitative markers of muscle mass(6). Therefore, CT-based muscle-fat parameters may serve as valuable risk indicators for hip fracture. Previous prediction models for hip fracture have largely focused on single risk domains, such as serum biomarkers, biochemical indices, or the FRAX tool(7–9). Few studies have simultaneously incorporated CT-derived muscle-fat parameters and multidimensional clinical variables to comprehensively estimate fracture risk.

Machine learning (ML), a branch of artificial intelligence, aims to enable computers to learn from large datasets and generate evaluative frameworks for specific tasks(10). Compared with traditional regression models, ML can efficiently integrate heterogeneous clinical and imaging data to uncover latent predictors. However, ML models often operate as “black boxes”, limiting clinical interpretability. Nomograms, by contrast, can visualize the results of complex models and transform abstract predicted probabilities into intuitive, manually calculable scoring scales, balancing accuracy with practical utility(11). Nomograms have been increasingly recognized as indispensable clinical prediction tools across diverse medical fields, offering individualized risk estimates through interpretable and visually intuitive formats. They have been extensively applied in oncology for predicting cancer risk, treatment response, and prognosis(12), as well as in orthopedics for forecasting postoperative complications(13), mortality(14), and fracture-related outcomes(15). These models have demonstrated good discriminative ability and clinical utility, underscoring the value of translating multivariable prediction models into practical bedside tools.

Given the increasing clinical emphasis on sarcopenia and the identified research gap, the present study integrated demographic characteristics, biochemical tests, bone mineral density, and multi-level CT muscle-fat parameters. We compared multiple ML algorithms with LR to construct a hip fracture prediction model, selected the optimal algorithm, and translated it into a visual nomogram. We sought to develop a user-friendly risk-stratification instrument with favorable discriminative performance to quantify individual fracture probability. Such a model would support early recognition of at-risk elderly patients and facilitate personalized preventive interventions—anti-osteoporosis treatment, nutritional

supplementation, and fall precaution strategies—ultimately reducing hip fracture rates and the risk of persistent long-term adverse events.

Methods

Study Design and Participants

This prospective study consecutively enrolled patients aged ≥ 60 years who underwent hip CT at Quanzhou Orthopedic-Traumatological Hospital between January 2025 and March 2026. Hip fracture (femoral neck or intertrochanteric fracture) was confirmed by radiography or CT. This study was conducted in accordance with the Declaration of Helsinki and followed all relevant guidelines and regulations. The study protocol was approved by the Ethics Committee of Quanzhou Orthopedic-Traumatological Hospital (Approval No. IRB-2025-009). Written informed consent was obtained from all participants. The study was registered on ClinicalTrials.gov (Registration No. MR-35-25-057914).

Inclusion criteria were: 1) age ≥ 60 years; 2) low-energy trauma (fall from standing height or lower); 3) complete clinical and imaging data. Exclusion criteria were: 1) old or pathological hip fracture; 2) pre-existing hip disease/surgery or lower-limb functional impairment; 3) unilateral hip fracture with contralateral injury affecting muscle/fat area or density measurements (e.g., contralateral hip, acetabular, or pelvic fractures); 4) poor-quality images (e.g., insufficient scan range, improper positioning). Ultimately, 202 fracture cases and 146 non-fracture controls were enrolled. The cohort was randomly divided (7:3) into training and validation sets. A patient flowchart is shown in Figure. 1.

Data Collection

Baseline data included: age, gender, body mass index (BMI), hypertension, elevated blood glucose, hyperlipidemia, low hemoglobin, alcohol consumption, and fracture type (femoral neck vs. intertrochanteric). Imaging parameters comprised osteoporosis determined by DXA, and CT-based measurements at the femoral head, lesser trochanter, and L3 vertebral levels: muscle area 1/2/3, fat area 1/2/3, mean muscle attenuation (muscle mean1/2/3), mean fat attenuation (fat mean1/2/3), and the corresponding SMI values (SMI1/2/3). Laboratory and imaging examinations were performed within 48 hours after fracture and before surgery.

Muscle-Fat Parameter Measurement

CT images were analyzed using Slice-O-matic software (version 5.0) with a threshold-based region-of-interest (ROI) delineation method. For the fracture group, the contralateral hip was measured; for the control group, the right hip was used. At the femoral head and lesser trochanter levels, we outlined the peri-hip muscle cross-sectional area and subcutaneous fat area at the same axial slice, and recorded mean CT attenuation values (in Hounsfield units, HU) for muscle and fat. At the L3 vertebral level, the paravertebral muscle area and subcutaneous fat area were similarly measured. Detailed ROI delineation is illustrated in Figure. 2. All ROI measurements were performed independently by two trained observers who were blinded to the patients' fracture status. Inter-observer reliability was assessed using the intraclass correlation coefficient (ICC), and the mean values of the two measurements were used for subsequent analyses. The ICCs for inter-observer agreement were 0.94 (95% CI: 0.89–0.97) for muscle area, 0.91 (95% CI: 0.85–0.95) for fat area, and 0.88 (95% CI: 0.81–0.93) for mean attenuation, demonstrating good to excellent reliability.

Model Development: Machine Learning and Logistic Regression

The Boruta algorithm was adopted to screen significant predictive variables from clinical baseline and imaging parameters. The dataset was split into training (70%) and validation (30%) sets. In the training set, eight classifiers were constructed to predict hip fracture: logistic regression (LR), random forest (RF), support vector machine (SVM), decision tree (DT), naive

Bayes (NB), boosted tree (BT), k-nearest neighbors (KNN), and multilayer perceptron (MLP). All ML models were evaluated under a ten-fold bootstrap resampling framework, with the area under the curve (AUC) as the primary performance metric.

Because LR demonstrated the best performance, subsequent univariable and multivariable LR analyses were performed to build the final model. The multivariable model was visualized as a nomogram. Decision curve analysis (DCA) was applied to assess clinical net benefit in both training and validation cohorts.

Statistical Analysis

Statistical analyses were performed using SPSS 26.0 and R (version 4.4.1). Categorical variables were expressed as percentages; continuous variables as mean $\pm$ standard deviation. Between-group comparisons used independent-samples t-tests (or Mann-Whitney U tests for unequal variances) for continuous variables, and chi-square or Fisher's exact tests for categorical variables. Univariable logistic regression was first conducted to evaluate each factor's association with hip fracture; significant variables were entered into multivariable logistic regression. Model discrimination was evaluated by ROC curves; calibration plots and the Hosmer-Lemeshow goodness-of-fit test; and clinical utility by DCA. A two-sided $P < 0.05$ was considered statistically significant.

Results

Baseline Characteristics

The fracture group comprised 202 patients (69 intertrochanteric [34.2%], 133 femoral neck [65.8%]) and the control group 146 patients. Baseline characteristics and imaging parameters are summarized in Table 1. Significant between-group differences were observed for age, BMI, elevated blood glucose, osteoporosis, muscle area1, fat area1, fat mean1, muscle area2, fat area2, muscle area3, fat area3, fat mean3, SMI1, SMI2, and SMI3 (all $P < 0.05$). No significant differences were found for gender, hypertension, alcohol consumption, hyperlipidemia, low hemoglobin, muscle mean1/2/3, or fat mean2.

Table 1
Baseline Clinical and Radiological Characteristics: Hip Fracture Group vs Control Group

Characteristic	Hip fracture group	Control group	P value
Age (years)	77.3 ± 10.99	70.0 ± 8.41	< 0.001
Gender			0.275
Male	53 (26.2%)	47 (32.2%)	
Female	149 (73.8%)	99 (67.8%)	
BMI (kg/m ²)	21.5 ± 3.34	23.7 ± 3.69	< 0.001
Hypertension			0.440
Yes	87 (43.1%)	56 (38.4%)	
No	115 (56.9%)	90 (61.6%)	
Elevated blood glucose			< 0.001
Yes	138 (68.3%)	69 (47.3%)	
No	64 (31.7%)	77 (52.7%)	
Alcohol consumption			0.653
Yes	2 (1.0%)	3 (2.1%)	
No	200 (99.0%)	143 (97.9%)	
Hyperlipidemia			0.079
Yes	16 (7.9%)	21 (14.4%)	
No	186 (92.1%)	125 (85.6%)	
Low hemoglobin			0.112
Yes	86 (42.6%)	49 (33.6%)	
No	116 (57.4%)	97 (66.4%)	
Osteoporosis			< 0.001
Yes	181 (89.6%)	99 (67.8%)	
No	21 (10.4%)	47 (32.2%)	
Muscle area1 (cm ²)	62.1 ± 14.02	72.2 ± 14.79	< 0.001
Fat area1 (cm ²)	112.0 ± 43.32	134.7 ± 59.87	< 0.001
Muscle mean1 (HU)	34.9 ± 7.22	33.5 ± 8.16	0.094
Fat mean1 (HU)	-95.0 ± 10.63	-97.9 ± 9.28	0.023
Muscle area2 (cm ²)	80.3 ± 19.61	93.8 ± 19.92	< 0.001
Fat area2 (cm ²)	117.9 ± 43.24	132.6 ± 53.53	0.028
Muscle mean2 (HU)	38.0 ± 6.94	36.1 ± 8.12	0.061
Fat mean2 (HU)	-96.0 ± 9.70	-98.1 ± 8.70	0.072
Data are mean ± standard deviation or n (%). BMI = body mass index; SMI = skeletal muscle index; HU = Hounsfield unit.			

Characteristic	Hip fracture group	Control group	P value
Muscle area3 (cm ²)	31.0 ± 9.35	38.1 ± 10.55	< 0.001
Fat area3 (cm ²)	111.5 ± 56.68	143.5 ± 68.04	< 0.001
Muscle mean3 (HU)	31.8 ± 9.78	32.7 ± 10.08	0.465
Fat mean 3 (HU)	-97.9 ± 13.04	-101.7 ± 10.14	0.011
SMI1 (cm ² /m ²)	25.0 ± 4.89	28.7 ± 5.19	< 0.001
SMI2 (cm ² /m ²)	32.4 ± 6.93	37.2 ± 6.99	< 0.001
SMI3 (cm ² /m ²)	12.4 ± 3.27	15.1 ± 3.71	< 0.001
Data are mean ± standard deviation or n (%). BMI = body mass index; SMI = skeletal muscle index; HU = Hounsfield unit.			

Feature Selection and Performance of Machine Learning Models

The boruta algorithm identified 12 important predictors: age, SMI3, muscle area1, muscle area2, SMI2, SMI1, muscle area3, muscle mean2, osteoporosis, muscle mean1, fat area3, and BMI (Figure. 3). Using these features, all eight ML models were constructed. The LR model achieved the highest overall predictive performance, with an AUC of 0.779. Detailed ROC curves and robustness assessments are presented in Figure. 4.

Nomogram Development (Visualization of the Optimal Model)

Given the optimal predictive performance of LR, we performed univariable and multivariable logistic regression. Stepwise multivariable analysis retained eight independent predictors ($P < 0.05$): age, elevated blood glucose, osteoporosis, fat area1, fat area2, muscle mean2, SMI1, and SMI3 (Table 2). A nomogram was constructed based on these eight variables (Figure. 5).

Table 2
Univariate and Multivariate Logistic Regression Analysis of Hip Fracture Risk Factors

Characteristics	Univariate analysis					Multivariate analysis				
	B	OR	95%CI		P	B	OR	95%CI		P
	lower		upper			lower		upper		
Age	0.077	1.080	1.054	1.106	<0.001	0.046	1.047	1.015	1.081	0.004
Gender	-0.289	0.749	0.469	1.196	0.226					
BMI	-0.174	0.840	0.786	0.897	<0.001					
Hypertension	-0.195	0.822	0.533	1.270	0.378					
Elevated blood glucose	-0.878	0.416	0.268	0.645	<0.001	-0.803	0.448	0.264	0.759	0.003
Alcohol Consumption	0.741	2.098	0.346	12.717	0.420					
Hyperlipidemia	0.669	1.953	0.981	3.889	0.057					
Low Hemoglobin	-0.384	0.681	0.438	1.061	0.089					
Osteoporosis	-1.409	0.244	0.138	0.432	<0.001	-0.966	0.381	0.193	0.752	0.005
Muscle area1	-0.050	0.951	0.935	0.967	<0.001					
Fat area1	-0.009	0.991	0.987	0.996	<0.001	-0.019	0.981	0.964	0.998	0.028
Muscle mean1	0.024	1.024	0.996	1.054	0.095					
Fat mean1	0.029	1.030	1.007	1.053	0.010					
Muscle area2	-0.035	0.966	0.954	0.978	<0.001					
Fat area2	-0.006	0.994	0.989	0.998	0.006	0.019	1.019	1.001	1.038	0.035
Muscle mean2	0.035	1.035	1.006	1.066	0.018	0.106	1.112	1.065	1.160	<0.001
Fat mean2	0.026	1.026	1.002	1.051	0.033					
Muscle area3	-0.073	0.930	0.908	0.953	<0.001					
Fat area3	-0.008	0.992	0.988	0.995	<0.001					
Muscle mean3	-0.009	0.991	0.970	1.013	0.413					
Fat mean3	0.028	1.029	1.009	1.049	0.004					
SMI1	-0.147	0.863	0.822	0.906	<0.001	-0.085	0.919	0.856	0.986	0.019
SMI2	-0.101	0.904	0.874	0.936	<0.001					
SMI3	-0.220	0.803	0.749	0.861	<0.001	-0.134	0.875	0.790	0.968	0.010
OR = odds ratio; CI = confidence interval. Only variables with $P < 0.05$ in univariate analysis are shown in full; multivariate model was built by stepwise selection.										

Nomogram Evaluation

The nomogram showed good discrimination: training AUC: 0.826 (95% CI: 0.782–0.870), validation AUC: 0.817 (95% CI: 0.750–0.884). Calibration curves demonstrated excellent agreement between predicted and observed probabilities, with Hosmer-Lemeshow $P = 0.647$ (training) and $P = 0.412$ (validation). DCA confirmed that the nomogram provided positive net benefit across a wide range of clinically reasonable threshold probabilities, outperforming the “treat-all” and “treat-none” strategies (Figure. 6). A dynamic online version of the nomogram is available at https://sulyang.shinyapps.io/Hip_Fracture/.

Discussion

In this prospective study, we systematically constructed and compared eight prediction models—LR, RF, SVM, DT, NB, BT, KNN, and MLP—for hip fracture in older adults using clinical and multi-level CT muscle-fat parameters. Among these, the LR model demonstrated the highest discriminative performance (AUC: 0.779), outperforming the other machine learning classifiers. A nomogram was subsequently developed based on this optimal LR model (training AUC = 0.826; validation AUC = 0.817), which exhibited favorable calibration and provided sustained net clinical benefit across a wide range of risk thresholds on DCA, confirming its good clinical utility. Our findings align with several recent systematic reviews and meta-analyses in geriatric and surgical prediction modeling. Lex et al.(16) in a meta-analysis of 39 hip fracture prediction studies, concluded that when modeling low-dimensional structured clinical variables (demographics, biochemistry, conventional CT metrics), ML offered no significant advantage over conventional LR; only in high-dimensional settings (radiomics, multi-omics) did nonlinear tree-based algorithms capture complex associations. Similarly, Lu et al.(17) in a 2025 prospective cohort on emergence delirium after general anesthesia in older adults, found that gradient-boosting models (XGBoost, LightGBM) showed high training AUCs but lower test AUCs and greater calibration deviation than LR, attributing this to predominantly linear clinical relationships and amplification of single-center noise by nonlinear algorithms. Wan et al. (18) compared five models for 6-month mortality in advanced cancer hospice patients and found LR superior to ML counterparts. Our results concur with these observations. However, Issa et al.(19) reported that ML slightly outperformed LR for 30-day mortality after hip fracture surgery, suggesting that outcome complexity and data dimensionality may modify relative algorithm performance.

Several factors may explain LR's superior performance over other ML algorithms in our dataset. First, our data were characterized by binary outcomes and predominantly linear predictor–outcome relationships without significant interactions, which aligned with LR assumptions but offered little advantage for tree-based or neural network models. These complex algorithms risked overfitting and amplifying noise rather than capturing meaningful signals(20). Second, with a moderate sample size ($n = 348$), complex ML models typically require larger datasets for reliable learning, whereas LR's parsimony conferred greater stability and robustness(21, 22). Together, these findings suggest that in structured clinical datasets with limited sample size and predominantly linear associations, simpler ML models such as LR may be preferable to more complex counterparts.

From a clinical translation perspective, LR offers unmatched interpretability compared to “black-box” ML models. RF, SVM, and similar models can output predicted probabilities but often require post-hoc tools such as SHAP to explain variable contributions, which may be inaccessible to frontline clinicians. LR directly provides odds ratios and 95% confidence intervals, intuitively conveying the direction and magnitude of each factor's effect(23). Based on the optimal LR model, we constructed a visual nomogram with an accompanying online dynamic calculator. This tool requires only eight routine clinical and CT-based variables and can be applied in outpatient or emergency screening without specialized software. All inputs are obtained from standard examinations, adding no extra radiation or laboratory burden. Similar nomograms have been widely adopted for predicting lymph node metastasis, recurrence, and other outcomes, demonstrating mature translational logic(24, 25).

Multivariable regression analysis identified eight independent risk factors for low-energy hip fracture—age, elevated blood glucose, osteoporosis, fat area1, fat area2, muscle mean2, SMI1, and SMI3—all of which were incorporated into the final nomogram. Advanced age and osteoporosis are well-established risk factors. Hyperglycemia is also a known contributor. Multiple studies have consistently reported higher fracture risk in patients with type 2 diabetes (26, 27), consistent with our

findings. SMI is a standardized CT-based metric for sarcopenia and is recommended by the Asian Working Group for Sarcopenia(5).

Previous studies often relied solely on single-level vertebral muscle measurements(28), overlooking the local mechanical protective role of peri-hip muscles. Our multivariable analysis addressed this gap by simultaneously assessing SMI at three CT levels (femoral head, lesser trochanter, and L3). We found that SMI at both the femoral head (SMI1) and L3 (SMI3) were independent protective factors against hip fracture, whereas SMI at the lesser trochanter (SMI2) was not retained in the final model. This suggests that peri-hip muscle mass quantified at the femoral head level provides predictive value beyond lumbar-based global skeletal muscle assessment, underscoring the importance of site-specific muscle evaluation in fracture risk stratification. Our findings are consistent with those of Huang et al.(29), who also identified sarcopenia as an independent risk factor for hip fracture in Chinese older adults in a longitudinal cohort.

Subcutaneous fat area at the femoral head and lesser trochanter, along with mean muscle attenuation at the lesser trochanter (muscle mean2), are key imaging markers of muscle fat infiltration. Intramuscular fat deposition reduces contractile strength, induces chronic low-grade inflammation, disrupts osteoblast–osteoclast balance, and accelerates bone loss. Our study showed that larger peri-hip fat area and higher muscle CT attenuation (indicative of more severe fat infiltration) were associated with increased fracture risk, highlighting the need to consider both muscle quantity and quality in fracture risk assessment. Existing literature has demonstrated that subcutaneous fat area at the upper thigh and hip muscle CT density are closely linked to adverse outcomes and fall risk after femoral fractures. However, the predictive value of different muscle groups' cross-sectional area and density remains debated. Our study focused on fracture occurrence and clarified that peri-hip fat infiltration is an independent pathogenic factor. Kim et al.(30)reported that subcutaneous fat area at the upper thigh level positively correlated with 1-year mortality in femoral fracture patients. Another study measured cross-sectional area and density of gluteus maximus, medius, and minimus in hip fracture patients; except for the areas of gluteus medius and minimus, most muscle parameters were significantly higher in survivors than in non-survivors, and gluteus maximus density predicted 1-year mortality(4). Ge et al.(31)examined associations between hip muscle groups and fear of falling in older adults with hip fractures; only gluteus medius/minimus density, but not size, was independently associated with lower fear of falling. The relative contributions of mean muscle/fat attenuation versus cross-sectional area thus warrant further investigation.

This study has several limitations. This study has several limitations. First, as a single-center prospective study, its findings may have limited generalizability and require external validation in larger, multicenter cohorts. Second, we did not adjust for several potential confounders, including physical activity, long-term medication use, dietary protein intake, fall history, serum sarcopenia biomarkers, or functional performance scales. Third, the lack of long-term follow-up precludes assessment of the model's predictive value for re-fracture or mortality. Fourth, manual ROI delineation of CT parameters may introduce inter-observer variability, though automated deep-learning segmentation could mitigate this issue. Despite these limitations, the nomogram and online calculator serve as a practical screening tool for older patients undergoing hip CT, and high-risk individuals may benefit from combined bone-muscle interventions to reduce fracture incidence and related morbidity.

In this study, LR outperformed other machine learning models in predicting hip fracture using clinical and multi-level CT muscle-fat parameters. Our nomogram, based on eight independent predictors, demonstrated good discrimination, calibration, and clinical utility. It offers a visual, individualized risk assessment tool that addresses the lack of sarcopenia imaging markers in prior models, and a web-based calculator enables rapid screening. Future work will focus on multicenter external validation, automated CT segmentation, incorporation of functional and serum markers, and interventional trials to confirm whether model-guided stratification reduces fracture incidence.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to restrictions that protect patient privacy and confidentiality.

Declarations

Competing interests

The authors declare no competing interests.

Funding

This work was funded by the Quanzhou City Science and Technology Program of Fujian Province, China (Grant No. 2025QZNG044).

Author Contribution

Dandan Ye: Conceptualization, funding acquisition, writing-original draft, and data analysis. Xiaoxiong Lin: Data acquisition, analysis, and drafting. Zhiwei Xu and Jiahong Huang: Data acquisition. Jinquan Lin: Data acquisition. Zhenbin Wu: Conceptualization, supervision, and final approval. All authors reviewed and approved the final manuscript. Dandan Ye and Xiaoxiong Lin contributed to the work equally and should be regarded as co-first authors.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to restrictions that protect patient privacy and confidentiality.

References

1. Bell, J. J. et al. The SIMPLER Nutrition Pathway for Fragility Fractures: A Quality Improvement Initiative. *Nutrients* **17**, (2025).
2. Zhang, C. et al. Incidence of and trends in hip fracture among adults in urban China: A nationwide retrospective cohort study. *PLoS Med.* **17**, e1003180 (2020).
3. Hoong, C. W. S., Saul, D. & Khosla, S. Sfeir, Advances in the management of osteoporosis. *Bmj* **390**, e081250 (2025).
4. Wang, L., Yin, L., Yang, M. & Cheng, X. Muscle composition and the imminent mortality risk after hip fracture. *J. Cachexia Sarcopenia Muscle.* **13**, 2752–2753 (2022).
5. Chen, L. K. et al. Asian Working Group for Sarcopenia: 2019 Consensus Update on Sarcopenia Diagnosis and Treatment. *J. Am. Med. Dir. Assoc.* **21**, 300–307e302 (2020).
6. Chianca, V. et al. Sarcopenia: imaging assessment and clinical application. *Abdom. Radiol. (NY).* **47**, 3205–3216 (2022).
7. Asrian, G., Suri, A. & Rajapakse, C. Machine learning-based mortality prediction in hip fracture patients using biomarkers. *J. Orthop. Res.* **42**, 395–403 (2024).
8. Xiao, P. et al. No Significant Direct Causal Association Between TNF Pathway Biomarkers and Hip Fracture Risk: A Study Based on Real-World Data. *J Vis. Exp.*, (2026).
9. Chu, W. et al. HALP score-adjusted FRAXplus tool effectively predicts hip fracture risk in elderly women with type 2 diabetes mellitus. *BMC Endocr. Disord.*, (2026).
10. Zhang, K. et al. Radiomics nomogram for the prediction of microvascular invasion of HCC and patients' benefit from postoperative adjuvant TACE: a multi-center study. *Eur. Radiol.* **33**, 8936–8947 (2023).
11. He, Z. et al. Development and validation of a machine learning-based model for diagnosing perioperative malnutrition in older adults with hip fracture. *Front. Nutr.* **13**, 1751993 (2026).

12. Wang, H. et al. Application and progress of nomograms in gastric cancer. *Front. Med. (Lausanne)*. **12**, 1510742 (2025).
13. Hao, Z. et al. Development and validation of a nomogram model for predicting postoperative nonunion in femoral shaft fractures. *BMC Musculoskelet. Disord.* **26**, 874 (2025).
14. Gong, Y. et al. Development and temporal external validation of a novel nomogram for predicting one-year mortality in the older adult with hip fracture. *Front. Med. (Lausanne)*. **12**, 1532196 (2025).
15. Wu, W. et al. A novel indicator for predicting functional recovery in elderly hip fracture patients. *Front. Med. (Lausanne)*. **12**, 1538038 (2025).
16. Lex, J. R. et al. Artificial Intelligence for Hip Fracture Detection and Outcome Prediction: A Systematic Review and Meta-analysis. *JAMA Netw. Open.* **6**, e233391 (2023).
17. Lu, Y. et al. Comparison of machine learning and logistic regression models for predicting emergence delirium in elderly patients: A prospective study. *Int. J. Med. Inf.* **199**, 105888 (2025).
18. Cheng, W. et al. Machine learning models to predict 6-month mortality risk in home-based hospice patients with advanced cancer. *Asia Pac. J. Oncol. Nurs.* **12**, 100679 (2025).
19. Issa, Z. et al. Are machine learning models superior to logistic regression models to predict 30-day mortality post-hip fracture surgery? *JBMR Plus.* **10**, zia079 (2026).
20. Barreñada, L., Dhiman, P., Timmerman, D., Boulesteix, A. L. & Van Calster, B. Understanding overfitting in random forest for probability estimation: a visualization and simulation study. *Diagn. Progn Res.* **8**, 14 (2024).
21. Tsegaye, B. et al. Larger sample sizes are needed when developing a clinical prediction model using machine learning in oncology: methodological systematic review. *J. Clin. Epidemiol.* **180**, 111675 (2025).
22. Silvey, S. & Liu, J. Sample Size Requirements for Popular Classification Algorithms in Tabular Clinical Data: Empirical Study. *J. Med. Internet Res.* **26**, e60231 (2024).
23. Luthra, A. K. et al. A Comparison of Machine Learning Methods and Conventional Logistic Regression for the Prediction of In-Hospital Mortality in Acute Biliary Pancreatitis. *Pancreas* **51**, 1292–1299 (2022).
24. Wang, J., Geng, Q., Wang, M., Yang, M. & Zhang, Y. Dynamic nomogram for predicting axillary lymph node metastasis in breast cancer based on MRI analysis and the platelet-to-lymphocyte ratio (PLR). *Cancer Imaging*, (2026).
25. Zhang, Z. et al. A nomogram predicting postoperative recurrence risk in SF1/TPIT nonfunctioning pituitary neuroendocrine tumors: integration of tumor heterogeneity quantification on T2-weighted imaging. *J Neurosurg*, 1–10 (2026).
26. Zou, X. et al. Type 2 Diabetes Is Associated With Increased Complications and Mortality After Hip Fracture in Older Adults Aged 60 Years or Older. *Diabetes Obes. Metab.* (2026).
27. Zoulakis, M., Johansson, L., Litsne, H., Axelsson, K. & Lorentzon, M. Type 2 Diabetes and Fracture Risk in Older Women. *JAMA Netw. Open.* **7**, e2425106 (2024).
28. Li, E. L. et al. Based on CT scans at the 12th thoracic spine level, assessing the impact of skeletal muscle and adipose tissue index on one-year postoperative mortality in elderly hip fracture patients: a propensity score-matched multicenter retrospective study. *BMC Musculoskelet. Disord.* **26**, 21 (2025).
29. Huang, S., Chang, L. & Cai, Z. Sarcopenia is associated with increased hip fracture risk among middle-aged and elderly Chinese adults in longitudinal analysis of CHARLS data. *Sci. Rep.* **15**, 24598 (2025).
30. Kim, Y. J. et al. Subcutaneous fat area at the upper thigh level is a useful prognostic marker in the elderly with femur fracture. *J. Cachexia Sarcopenia Muscle.* **12**, 2238–2246 (2021).
31. Ge, Y. et al. Gluteal muscle density, but not size, is independently associated with fall risk in older adults with hip fractures: A cross-sectional study using opportunistic CT imaging. *Maturitas* **211**, 108975 (2026).

Figures

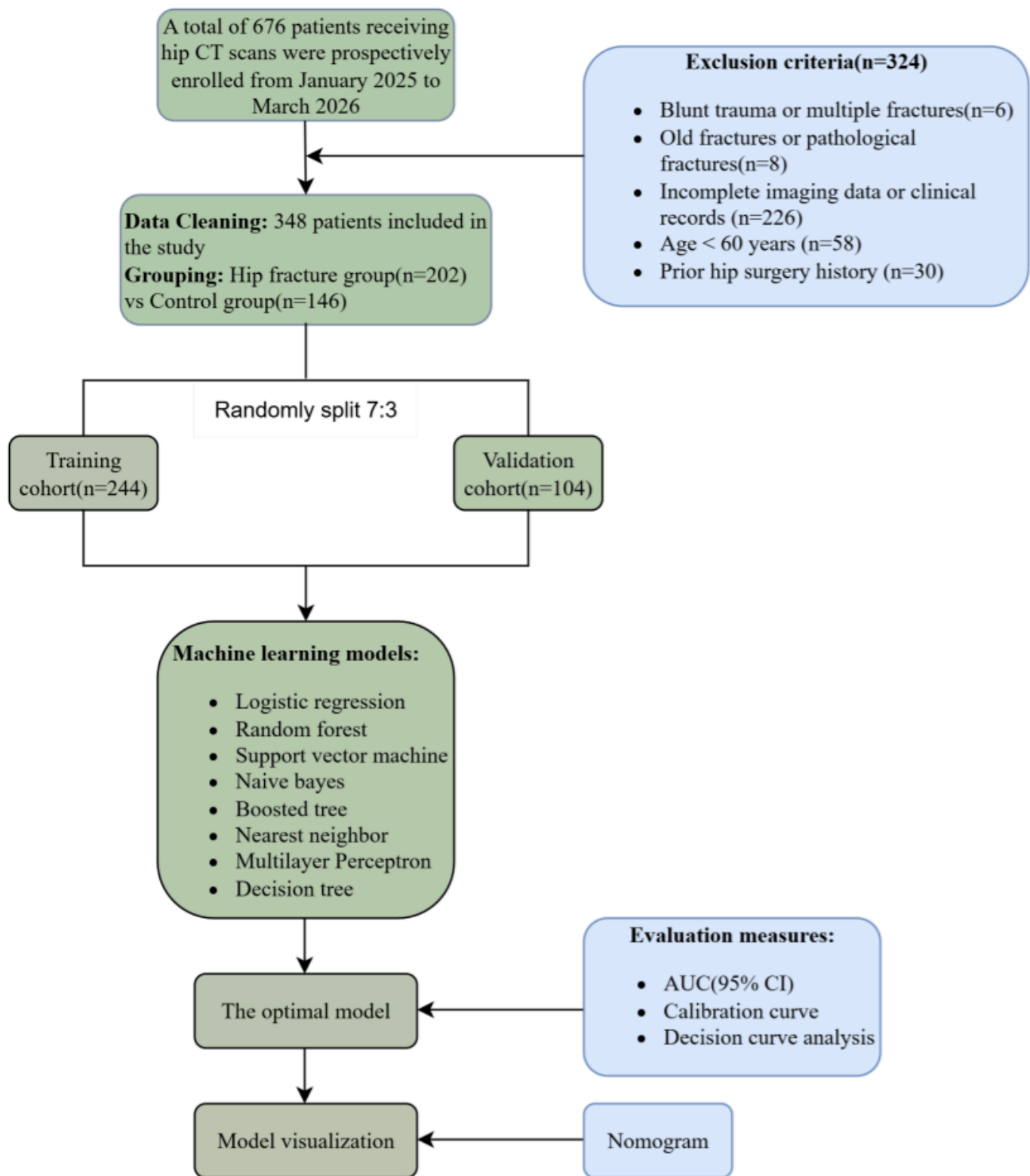


Figure 1

Patient flowchart and study design. AUC, area under the curve; CI, confidence interval.

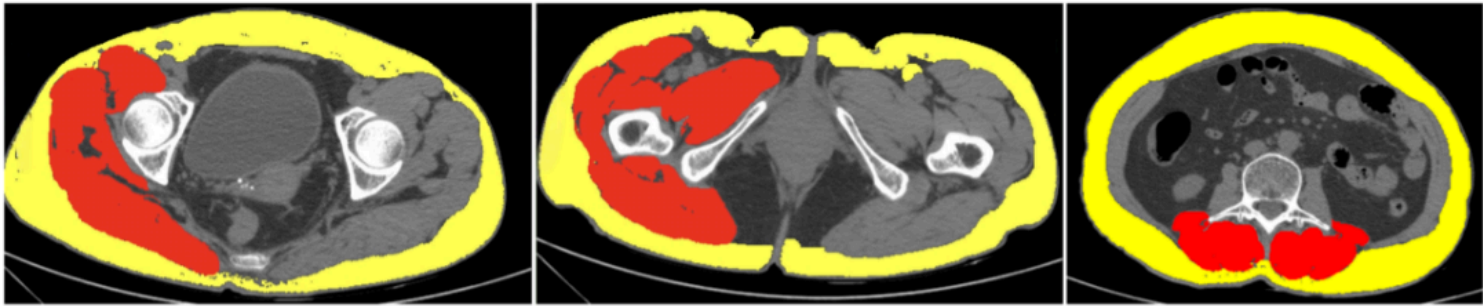


Figure 2

Axial CT images at each level with delineated regions (red: skeletal muscle; yellow: subcutaneous fat).

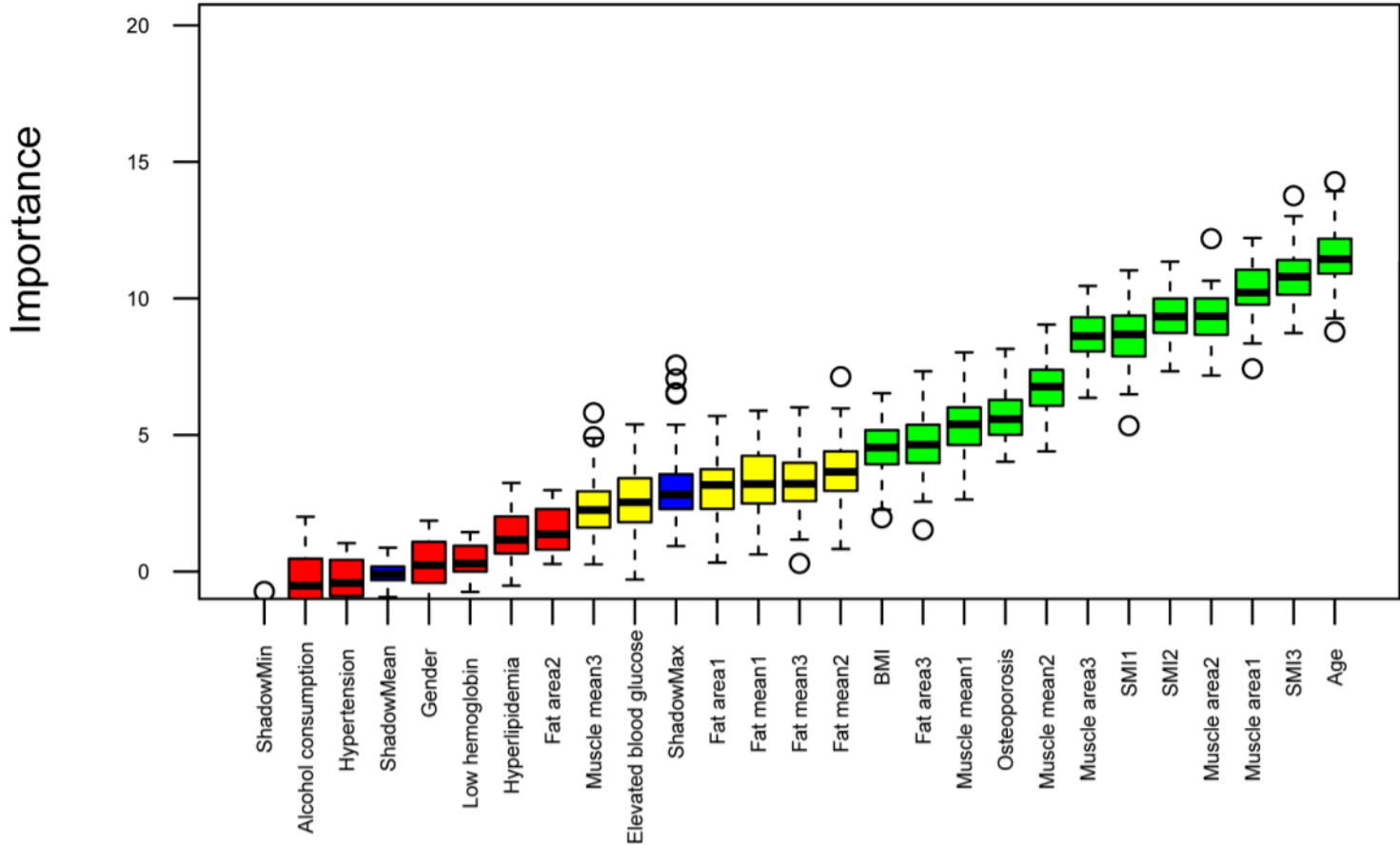


Figure 3

The boruta algorithm results for important feature selection.

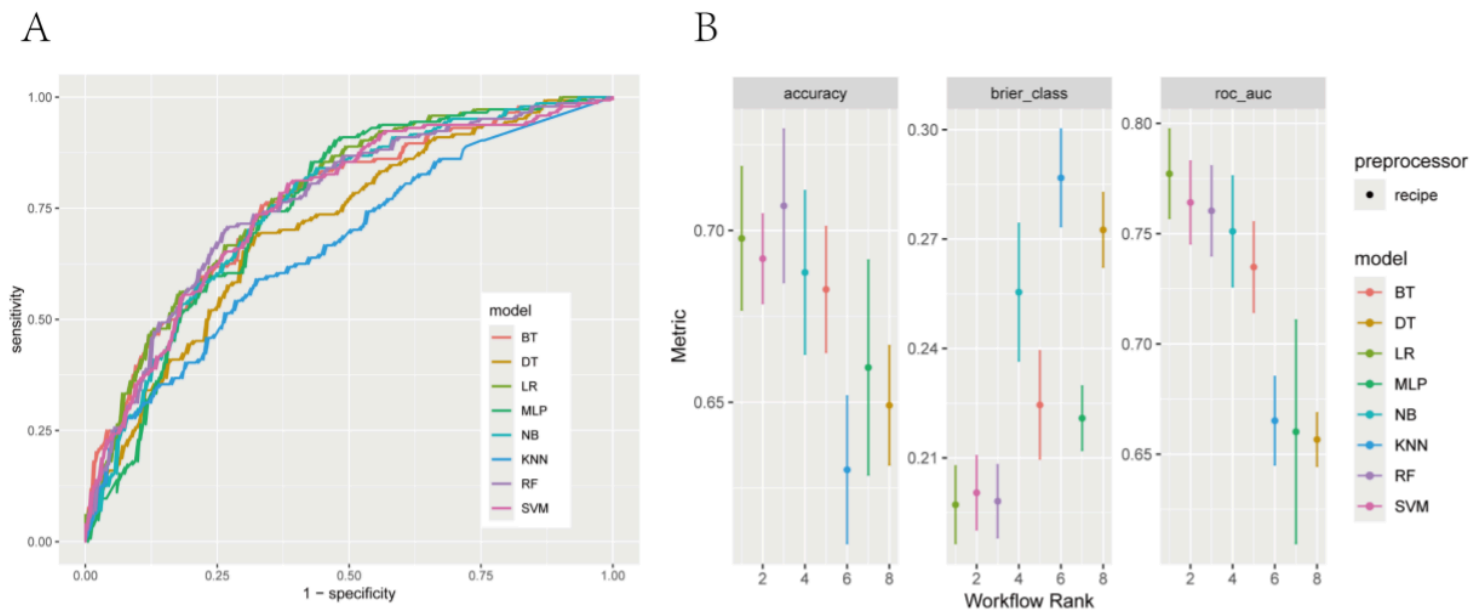


Figure 4

Receiver operating characteristic (ROC) curves, predictive performance, and robustness assessment of different machine learning classifiers.

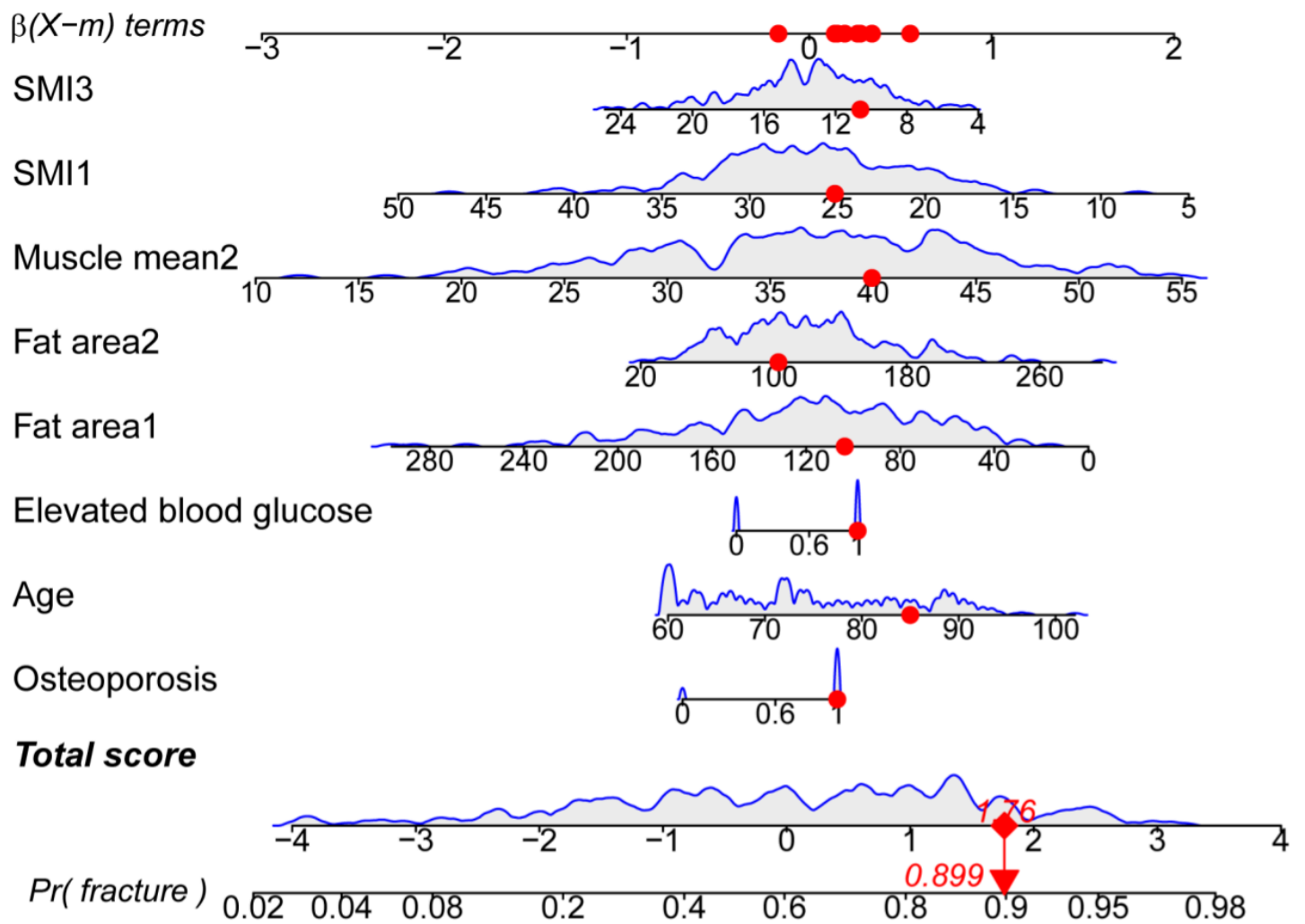


Figure 5

Nomogram model. The nomogram was constructed based on LR model to predict the risk of hip fracture. The dynamic nomogram can be accessed at https://sulyang.shinyapps.io/Hip_Fracture/. This model comprises SMI3, SMI1, muscle mean2, fat area2, fat area1, elevated blood glucose, age, osteoporosis.

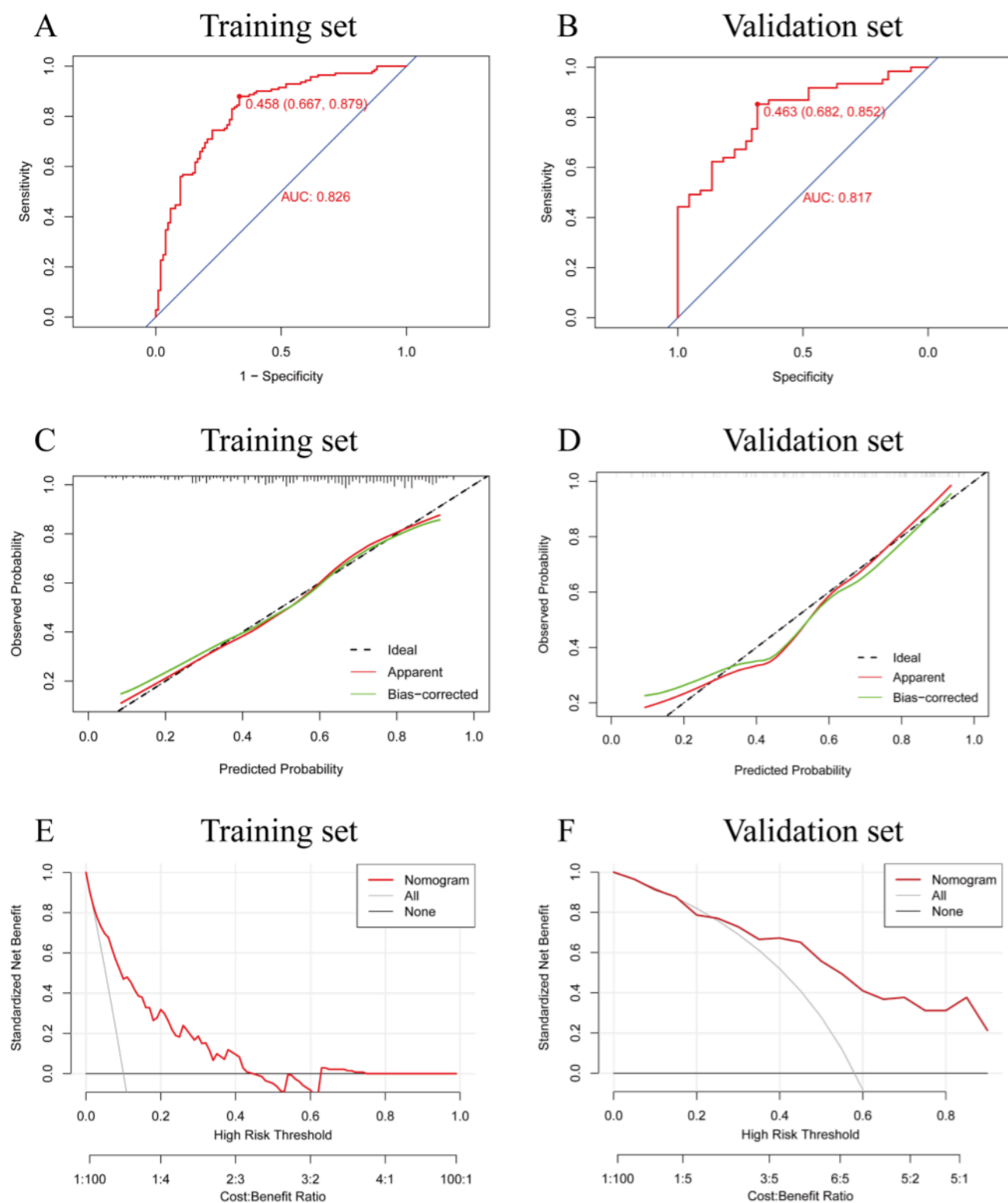


Figure 6

Validation of the nomogram for predicting hip fracture risk.

Panels A-B are ROC curves of the nomogram in the training set and validation set, with AUC values of 0.826 and 0.817 respectively, indicating good discriminative performance. Panels C-D display calibration curves for the two cohorts; the bias-corrected curves (green) closely overlap the ideal 45° diagonal dashed line, demonstrating favorable consistency between predicted and observed hip fracture probabilities. Panels E-F present DCA. The red curve of the nomogram

maintains a higher standardized net benefit than the two extreme strategies (“All” treat all patients, “None” treat no patients) across a wide range of clinically relevant risk thresholds, confirming the model’s valuable clinical decision-making utility.