Check for updates

OPEN ACCESS

EDITED BY Yu Wang, Beihang University, China

REVIEWED BY Yansong Qi, Inner Mongolia People's Hospital, China Wubing He, Fujian Medical University, China Yudi Sang, Rossum Robot, China

*CORRESPONDENCE Li Li ⊠ 13675648533@139.com Xuezhi Shi ⊠ 2798984003@qq.com

[†]These authors have contributed equally to this work

RECEIVED 30 December 2024 ACCEPTED 16 June 2025 PUBLISHED 26 June 2025

CITATION

Li C, Wang C, Zhang J, Zheng W, Shi J, Li L and Shi X (2025) A risk prediction model for poor joint function recovery after ankle fracture surgery based on interpretable machine learning. *Front. Med.* 12:1553274. doi: 10.3389/fmed.2025.1553274

COPYRIGHT

© 2025 Li, Wang, Zhang, Zheng, Shi, Li and Shi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A risk prediction model for poor joint function recovery after ankle fracture surgery based on interpretable machine learning

Congyang Li^{1†}, Chenggang Wang^{1†}, Jiru Zhang¹, Wenjun Zheng², Jing Shi¹, Li Li³* and Xuezhi Shi⁴*

¹Department of Orthopaedics, Lu'an Hospital of Anhui Medical University, Lu'an, China, ²Wound Stoma Care Clinic, Lu'an Hospital of Anhui Medical University, Anhui, China, ³Department of Science and Education, Lu'an Hospital of Anhui Medical University, Lu'an, China, ⁴Nursing Department, Lu'an Hospital of Anhui Medical University, Lu'an, China

Objective: Currently, there is no individualized prediction model for joint function recovery after ankle fracture surgery. This study aims to develop a prediction model for poor recovery following ankle fracture surgery using various machine learning algorithms to facilitate early identification of high-risk patients.

Methods: A total of 750 patients who underwent ankle fracture surgery at Lu'an Hospital Affiliated to Anhui Medical University between January 2018 and December 2023 were followed up. The collected data were chronologically divided into a training set (599 cases) and a test set (151 cases). Feature variables were selected using the Boruta algorithm, and five machine learning algorithms (logistic regression, random forest, extreme gradient boosting, support vector machine, and lasso-stacking) were employed to construct models. The performance of these models was compared on both the training and test sets to select the best-performing model. The decision basis of the optimal model was further analyzed using Shapley Additive Explanation (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME).

Results: In total, 12 characteristic variables were identified using the Boruta algorithm. Among the five machine learning models, random forest model: AUC (training set: 0.840, test set: 0.779), accuracy (training set: 0.781, test set: 0.742); SVM: AUC (training set: 0.809, test set: 0.768), accuracy (training set: 0.751, test set: 0.728); XGBoost: AUC (training set: 0.734, test set: 0.748), accuracy (training set: 0.668, test set: 0.722); logistic regression: AUC (training set: 0.672, test set: 0.691), accuracy (training set: 0.651, test set: 0.656); lasso-stacking model: AUC (training set: 0.877, test set: 0.791), accuracy (training set: 0.796, test set: 0.762). The PR curve and decision curve of the lasso-stacking model were better than those of other models. The lasso-stacking model had the best performance. SHAP analysis showed that functional exercise compliance, combined ligament injury, and open fracture accounted for the largest proportion of SHAP values and were the most important influencing factors.

Conclusion: Through evaluation and comparison of the developed models, the lasso-stacking model demonstrated the best performance and is more suitable for predicting joint function recovery after ankle surgery. This model can be further validated externally and applied in clinical practice.

KEYWORDS

ankle fracture, poor functional recovery, machine learning, interpretability analysis, prediction

1 Introduction

Ankle fractures, often resulting from external violent factors such as falls, sports injuries, and traffic accidents, are prevalent among middle-aged populations (1-3) and currently rank as the third most common fracture type in orthopedic practice (4, 5). Recent studies have revealed a trend of increasing incidence rates (6-8), accompanied by high treatment costs. In the United States, the average cost of outpatient treatment for ankle fractures is \$9,821, while the average cost of inpatient treatment exceeds \$62,000 (9, 10), posing a significant economic burden on patients. Surgery remains the primary treatment modality for ankle fractures (9, 11). However, due to various factors such as patient-specific differences, fracture severity, surgical complications, and postoperative rehabilitation, patients with ankle fractures face risks of joint weakness, stiffness, and residual pain, which can negatively impact joint function recovery, postoperative quality of life, increase the risk of reoperation, and even lead to disability (12-16). Currently, most clinicians focus on surgical technique improvements, postoperative complication prevention, and early postoperative rehabilitation (17-19), while epidemiological aspects of postoperative joint recovery and factors influencing individualized recovery receive less attention. Understanding these influencing factors, early identification of high-risk populations, and timely intervention are crucial for ensuring normal postoperative joint function recovery in patients with ankle fractures. However, there is currently no research on the individualized prediction of postoperative joint function recovery in patients with ankle fractures.

As a pivotal branch of artificial intelligence, machine learning (ML) has gained increasing traction in medical research due to its strengths in non-linear modeling and high-dimensional data processing, particularly in surgical outcome prediction (20, 21). Unlike conventional statistical methods-such as logistic regression and Cox proportional hazards models, which rely on linear assumptions and manual variable selection-ML algorithms autonomously identify latent non-linear relationships and interaction effects within complex datasets, thereby enhancing predictive accuracy and robustness (22). These advantages have translated into notable successes in orthopedic surgical outcome prediction. For instance, Lex et al. (23) conducted a systematic review of predictive models for hip fracture outcomes, demonstrating that ML algorithms achieved superior discriminative performance for 1-year postoperative mortality prediction (mean area under the receiver operating characteristic curve [AUC]: 0.84) compared to traditional models (AUC: 0.79). Similarly, Cai et al. developed a stacked ensemble-based ML classifier to predict the Japanese Orthopedic Association recovery rate for patients with degenerative cervical myelopathy, reporting exceptional performance (AUC: 0.92, accuracy: 90.2%, sensitivity: 90.1%) that significantly outperformed conventional approaches (AUC: 0.78, accuracy: 79.3%, sensitivity: 65.0%) (24). Despite these advancements, the application of ML in predicting outcomes following ankle fracture surgery remains scarcely investigated. Therefore, in this study, we employ several classic machine learning algorithms to construct prediction models for poor postoperative recovery after ankle fractures. By comparing the performance of these models, we select the optimal one and conduct an interpretability analysis to determine the significant influencing factors.

This study aims to provide clinical guidance for early identification of high-risk populations, timely implementation of preventive measures, facilitation of rapid postoperative recovery, reduction of hospitalization costs, and optimization of healthcare resource allocation for ankle fracture patients. By bridging this research gap, we seek to empower clinicians with data-driven tools to improve patient outcomes and quality of life.

2 Methods

2.1 Study designs

This single-center, prospective observational study was conducted with the approval of the Ethics Committee of Lu'an Hospital of Anhui Medical University (2024LLKS-KY-044). All participants provided written informed consent. Data were collected in January 2018 using a structured questionnaire and a clinical electronic inpatient record system. Participants completed the questionnaire online through Questionnaire Star, and hematological indicators were obtained through the clinical electronic inpatient medical record system. The study was conducted in accordance with the ethical standards of the local IRB and the 1975 Declaration of Helsinki.

2.2 Patients and data collection

2.2.1 Study population

The study population consisted of 750 patients who underwent ankle surgery at the orthopedic department of a tertiary hospital in Lu'an City, Anhui Province, between January 2018 and December 2023.

The inclusion criteria were as follows: (1) age \geq 18 years; (2) radiographic evidence of ankle fracture requiring surgery; (3) no previous ankle surgery; (4) voluntary participation in the study with signed informed consent.

The exclusion criteria were as follows: (1) patients choosing conservative treatment; (2) missing data; (3) presence of functional or organic mental disorders with language communication barriers.

2.2.2 Data collection

Based on literature review and expert consultation, researchers identified 31 potential factors influencing poor prognosis after ankle fracture surgery, including 23 categorical and 8 continuous variables. The operational definitions of the candidate predictors are as follows:

- 1. Age (years): Refers to patients over the age of 18 undergoing ankle fracture surgery. This is a continuous variable.
- 2. Gender: Male or female. This is a categorical variable.
- 3. Education level: Describes the educational background of patients undergoing ankle fracture surgery, including primary school or below, junior high to high school, and university or above. This is a categorical variable.
- 4. Varicose veins: Indicates whether patients undergoing ankle fracture surgery have concomitant varicose veins. This is a categorical variable.
- 5. Heart disease: Denotes whether patients undergoing ankle fracture surgery have comorbid heart disease. This is a categorical variable.
- 6. Cerebrovascular disease: Represents whether patients undergoing ankle fracture surgery have comorbid cerebrovascular disease. This is a categorical variable.

- 7. Diabetes: Signifies whether patients undergoing ankle fracture surgery have comorbid diabetes. This is a categorical variable.
- 8. Hypertension: Indicates whether patients undergoing ankle fracture surgery have comorbid hypertension. This is a categorical variable.
- 9. Smoking: Refers to whether there is a history of smoking recorded in the electronic nursing records of patients undergoing ankle fracture surgery. This is a categorical variable.
- 10. Alcohol consumption: Pertains to whether there is a history of alcohol consumption recorded in the electronic nursing records of patients undergoing ankle fracture surgery. This is a categorical variable.
- 11. Injury mechanism: Describes how the ankle fracture occurred, including car accidents, falls, and other mechanisms. This is a categorical variable.
- 12. Injury site: Denotes the surgical site of patients undergoing ankle fracture surgery, including left, right, or bilateral. This is a categorical variable.
- Body mass index (kg/m2): Represents the body mass index of patients at the time of admission for ankle fracture surgery. This is a continuous variable.
- 14. Combined ligament injury: Indicates whether patients with ankle fractures have concomitant ligament injuries. This is a categorical variable.
- 15. Nerve injury: Signifies whether patients with ankle fractures have concomitant nerve injuries. This is a categorical variable.
- 16. Combined joint dislocation: Represents whether patients with ankle fractures have concomitant joint dislocation. This is a categorical variable.
- 17. Open fracture: Denotes whether the ankle fracture is an open fracture. This is a categorical variable.
- Fracture type: Describes the severity of the ankle fracture, including single ankle fracture, double ankle fracture, and triple ankle fracture. This is a categorical variable.
- 19. Surgical waiting time (days): Represents the time from injury to surgery for patients with ankle fractures. This is a continuous variable.
- 20. Perioperative use of blood-activating and stasis-removing drugs: Indicates whether patients with ankle fractures take blood-activating and stasis-removing drugs during the perioperative period. This is a categorical variable.
- 21. Postoperative hemoglobin (g/L): Signifies the hemoglobin value from the first hematological examination after ankle fracture surgery. This is a continuous variable.
- 22. Postoperative albumin (g/L): Represents the albumin value from the first hematological examination after ankle fracture surgery. This is a continuous variable.
- 23. Postoperative red blood cell count ($\times 10^{12}$ /L): Denotes the red blood cell count from the first hematological examination after ankle fracture surgery. This is a continuous variable.
- 24. Postoperative drainage tube: Indicates whether patients undergoing ankle fracture surgery have a postoperative incision drainage tube. This is a categorical variable.
- 25. Operation time (minutes): Represents the difference between the start and end times recorded in the electronic anesthesia record for ankle fracture surgery. This is a continuous variable.

- 26. Length of hospital stay (days): Describes the duration of the hospital stay for patients undergoing ankle fracture surgery. This is a continuous variable.
- 27. Venous thrombosis: Indicates whether venous thrombosis occurs during hospitalization for ankle fracture patients, as detected by ultrasound examination (25, 26). This is a categorical variable.
- 28. Surgical site infection: Represents whether a surgical site infection occurs after ankle fracture surgery. Surgical site infection is diagnosed based on the Guidelines for the Prevention of Surgical Site Infection from the Centers for Disease Control and Prevention and the Healthcare Infection Control Practices Advisory Committee (27, 28). This is a categorical variable.
- 29. The American Society of Anesthesiologists (ASA) Classification: Describes the ASA classification recorded in the electronic anesthesia record for ankle fracture surgery. ASA assesses patients' physical status and surgical risk before anesthesia, ranging from Class 1 to Class 6, for evaluating surgical risk (29, 30). This is a continuous variable.
- 30. Postoperative pain level after functional exercise: Represents the average pain score recorded using The Numeric Rating Scale (NRS) in the electronic nursing records after each rehabilitation exercise guided by a therapist following ankle fracture surgery. Scores 1–3 indicate mild pain, 4–6 moderate pain, and 7–9 severe pain. This is a continuous variable.
- 31. Postoperative functional exercise compliance (PFEC): This refers to the compliance of patients with ankle fractures in performing rehabilitation exercises after being instructed by a rehabilitation therapist following surgery. Assessment is conducted on the day of the patient's discharge using the Orthopedic Rehabilitation Exercise Compliance Scale. Developed by Chinese scholars Tan et al., the scale consists of three dimensions: compliance related to psychological persistence in exercise, compliance related to active learning and persistence in exercise, and compliance related to physical persistence in exercise. It includes 15 items rated on a 1-5 Likert scale, with a total score of 75 points; a higher score indicates better patient compliance with functional exercises. Scores ≤20 indicate low compliance, scores between 20 and 55 points indicate partial compliance, and scores \geq 50 indicate high compliance. The total Cronbach's α coefficient for this scale is 0.930, with coefficients of 0.920, 0.842, and 0.851 for each respective dimension (31). This is a continuous variable.

2.2.3 Criteria for diagnosing poor recovery

The American Orthopaedic Foot and Ankle Society Ankle-Hindfoot Scale (AOFAS) was used to evaluate the functional recovery of patients' ankles 3 months after surgery. This scale, proposed by KITAOKA et al. in 1994 (32), has been widely applied in many countries (33, 34). It is primarily designed to assess ankle function. The scale ranges from 0 to 100 points, with excellent (90–100 points), good (75–89 points), moderate (50–74 points), and poor (<50 points) categories. A higher score indicates better ankle function. In this study, patients with an AOFAS score below 75 at 3 months post-ankle surgery were diagnosed as having a poor prognosis (poor prognosis label = 1, non-poor prognosis = 0).

2.3 Feature selection and model construction

The flowchart of this study is shown in Figure 1. The included patients were divided into two groups chronologically (80% training set, 20% test set). The training set data was used to build the model, while the test set was used to evaluate the performance of the model.

In this study, the Boruta algorithm was employed for feature variable selection. The core steps of this method are divided into constructing shadow features and random forest voting. Shadow features are copies of the original features, where the values are randomly rearranged to eliminate their correlation with the outcome variable. The Boruta algorithm combines shadow features and original features to construct a new dataset, utilizing a random forest model to determine the importance of each feature in the new dataset (35). Random permutation of attribute values among objects leads to a decrease in classification accuracy, and the Z-score, which is the average accuracy loss divided by its standard deviation, serves as an indicator to measure feature importance for variable selection.

In this study, five machine learning algorithms were employed to construct prediction models, namely: random forest (RF), support vector machine (SVM), eXtreme gradient boosting (XGBoost), logistic regression (LR), and least absolute shrinkage and selection operator stacking (lasso-stacking). Among these, lasso-stacking is a stacking ensemble model based on the first four models, with lasso serving as the meta-model. The first four models underwent five-fold cross-validation, while the lasso-stacking model adopted Bootstrap resampling for cross-validation to ensure the stability and accuracy of the models.

The hyperparameter optimization methods for the random forest, SVM, XGBoost, and lasso-stacking algorithms were implemented as follows:

- 1. Random forest: Within the tidymodels framework, hyperparameter tuning was performed using the randomForest engine through a 5-fold cross-validation grid search. The key hyperparameters optimized included: mtry (number of features considered at each split): Initially set as sqrt(p), with optimal values selected between 2 and 10; min_n (minimum node size): Optimized within the range of 20 to 50; trees (number of trees in the forest): Evaluated across 200–500 candidate values. The grid search for all hyperparameters was executed using the tidymodels framework, with AUC (area under the ROC curve) as the evaluation metric. The optimal parameter combination was identified as mtry = 33, min_n = 39, and trees = 235, achieving an ROC_AUC of 0.692.
- 2. XGBoost: Hyperparameter optimization was performed using the xgboost engine within the tidymodels framework, employing a 5-fold cross-validated grid search. This approach enabled systematic evaluation of model performance across diverse hyperparameter combinations to identify optimal settings. The following hyperparameters were prioritized for tuning: mtry (number of features evaluated at each node split), trees (total number of decision trees in the ensemble), min_n (minimum sample size required for terminal leaf nodes),



Study flow chart. LR: RF, random forest; logistic regression; SVM, support vector machine; XGBoost, eXtreme gradient boosting; L-S, Lasso-Stacking; AUC, area under the receiver operating characteristic curve; DCA, decision curve analysis; SHAP, Shapley additive explanations; LIME, Local Interpretable Model-Agnostic Explanations.

tree_depth (maximum permissible depth of individual trees), learn_rate (learning rate governing stepwise contributions of trees to final predictions), loss_reduction (minimum loss reduction required for node splitting), sample_size (proportion of training samples utilized per iteration), and stop_iter (number of iterations without improvement for early stopping). All grid search procedures were executed under the tidymodels framework, with AUC (area under the ROC curve) serving as the primary evaluation metric. With trees fixed at 1,000 and stop_iter set to 25, the optimal parameter combination was identified as mtry = 3, min_n = 9, tree_depth = 2, learn_ rate = 0.00564, loss_reduction = 0.0351, and sample_ size = 0.871, achieving an ROC_AUC of 0.639.

- 3. SVM: The model was implemented using the svm_rbf() function from the tidymodels framework with the kernlab engine. A radial basis function (RBF) kernel was employed, with two critical hyperparameters—cost (C) and rbf_sigma (γ)—optimized via a 5-fold cross-validated grid search over predefined parameter spaces. The cost parameter governed the penalty for misclassifications, while rbf_sigma determined the width of the RBF kernel, thereby influencing model flexibility. Hyperparameter tuning was guided by maximization of the area under the ROC curve (AUC) to ensure an optimal balance between model complexity and generalization capability. The optimal parameter combination was identified as cost (C) = 5.15 and rbf_sigma (γ) = 0.00905.
- 4. Lasso-stacking: A stacked ensemble model was constructed within the tidymodels framework, incorporating random forest, support vector machine, XGBoost, and logistic regression as base learners, with a lasso-regularized logistic regression model serving as the meta-learner. Model validation was conducted using bootstrap resampling-based cross-validation, with ROC_AUC adopted as the performance metric. Hyperparameter tuning focused on the lasso penalty parameter (λ), which governs feature sparsity in the meta-learner. The final model selected $\lambda = 0.03316$, yielding a sparse ensemble where only a subset of base learners contributed to predictions. The stacked ensemble demonstrated a mean cross-validated ROC AUC of 0.703.

2.4 Model assessment

In this study, the performance of the models was evaluated separately on the training and test sets to determine the best model. Initially, the receiver operating characteristic (ROC) curve was plotted, and the area under the ROC curve (AUC) was calculated to quantify its discriminatory performance. The calibration of the model was assessed by plotting a calibration curve and computing the Brier score. Subsequently, the precision–recall (P–R) curve was utilized to further evaluate the model's discriminatory ability by plotting the relationship between positive predictive value (PPV) and true positive rate (TPR) for all thresholds. Additionally, the clinical decision curve (DCA) was employed to assess the clinical net benefit of each model. The DeLong test was used to evaluate the robustness of the model. Finally, additional metrics such as accuracy, Kappa, precision, specificity, sensitivity, and F1 score were used to evaluate the predictive capabilities of the models (36).

2.5 Model interpretability analysis

In the optimal model, the contribution and significance of each feature variable to the outcome were determined based on the Shapley Additive Explanation (SHAP) values. Furthermore, Local Interpretable Model-Agnostic Explanations (LIME) were used to provide further interpretation of the model (37–40).

2.6 Quality control

To reduce bias in the data collection process, which may affect the research results, all objective data were obtained from electronic medical records and electronic nursing records and entered into the database through a double-check process. Unified training was conducted before the questionnaire survey, standardized sentences were used during the survey, and questionnaire scores were also entered into the database using a double cross-check. The database is maintained by a designated person, and once data are crosschecked and entered, they cannot be changed. In the process of model construction, to reduce the bias caused by time division, we tested the robustness of the model using random stratification of the data.

2.7 Statistical analysis

Descriptive and differential statistical analyses were conducted using SPSS 26.0 software, while model construction was performed using R 4.3.2 software. Measurement data conforming to a normal distribution were described using mean \pm standard deviation, and comparisons between groups were made using the t-test. Count data were described using frequencies and rates, and comparisons between groups were performed using the chi-square test or Fisher's exact probability method. The R package "Boruta" was used for Boruta analysis, while "tidymodels" and "stacks" were employed for model training. Stacking utilized bootstrap resampling for hyperparameter tuning, while grid search methods were used for hyperparameter tuning in the remaining models. The "fastshap," "shapviz," and "lime" packages were utilized to complete the interpretability analysis of SHAP and LIME. A *p*-value less than 0.05 was considered statistically significant.

3 Results

3.1 Baseline characteristics

In this study, a total of 750 patients undergoing ankle surgery were included, with a mean age of (49.93 \pm 15.84) years. Among them, 414 were men and 336 were women. Poor postoperative recovery occurred in 248 patients, with an incidence rate of 33.1%. The data were divided into the training set and the testing set in chronological order at a ratio of 8:2. A comparison of differences between the two datasets revealed no significant differences among the variables, indicating comparability between the two datasets. Detailed baseline patient characteristics and the results of the difference comparison are presented in Table 1.

TABLE 1 Comparative analysis of data variability between the train set and the test set.

| Clinical characteristics | Total (<i>n</i> = 750) | Training set (n = 599) | Testing set (n = 151) | χ²/t/Fisher | p |
|-----------------------------|-------------------------|---------------------------|--------------------------|-------------|------|
| Age (years) | 49.93 ± 15.84 | 49.78 ± 15.67 | 50.54 ± 16.54 | -0.53 | 0.60 |
| Gender (%) | | | | 2.34 | 0.13 |
| Male | 414 (55.2) | 339 (56.6) | 75 (49.7) | | |
| Female | 336 (44.8) | 260 (43.4) | 76 (50.3) | | |
| Education Level (%) | | | | 4.45 | 0.11 |
| Primary school and below | 279 (37.2) | 217 (36.2) | 62 (41.1) | | |
| Middle school to high | 356 (47.5) | 282 (47.1) | 74 (49.0) | | |
| school | | | | | |
| University and above | 115 (15.3) | 100 (16.7) | 15 (9.9) | | |
| VV (%) | | | | 0.65 | 0.42 |
| Yes | 78 (10.4) | 65 (10.9) | 13 (8.6) | | |
| No | 672(89.6) | 534 (89.1) | 138 (91.4) | | |
| HD (%) | | | | 0.17 | 0.68 |
| Yes | 116 (15.5) | 91 (15.2) | 25 (16.6) | | |
| No | 634 (84.5) | 508 (84.8) | 126 (83.4) | | |
| CD (%) | | | | 0.17 | 0.68 |
| Yes | 184 (24.5) | 145 (24.2) | 39 (25.8) | | |
| No | 566 (75.5) | 454 (75.8) | 112 (74.2) | | |
| Diabetes (%) | | | | 0.73 | 0.39 |
| Yes | 174 (23.2) | 135 (22.5) | 39 (25.8) | | |
| No | 576 (76.8) | 464 (77.5) | 112 (74.2) | | |
| Hypertension (%) | | | | 5.27 | 0.02 |
| Yes | 237 (31.6) | 201 (33.6) | 36 (23.8) | | |
| No | 513 (68.4) | 398 (66.4) | 115 (76.2) | | |
| IM (%) | | | | 1.73 | 0.42 |
| Fall | 474 (63.2) | 376 (62.8) | 98 (64.9) | | |
| Traffic accident injury | 141 (18.8) | 118 (19.7) | 23 (15.2) | | |
| Others | 135(18.0) | 105(17.5) | 30(19.9) | | |
| Smoking (%) | | | | 0.27 | 0.61 |
| Yes | 186 (24.8) | 151 (25.2) | 35 (23.2) | | |
| No | 564 (75.2) | 448 (74.8) | 116 (76.8) | | |
| Alcohol consumption (%) | | | | 0.03 | 0.86 |
| Yes | 203 (27.1) | 163 (27.2) | 40 (26.5) | | |
| No | 547 (72.9) | 436 (72.8) | 111 (73.5) | | |
| IS (%) | | | | 0.46 | 0.79 |
| Left | 386 (51.5) | 312 (52.1) | 74 (49.0) | | |
| Right | 335 (44.7) | 264 (44.1) | 71 (47.0) | | |
| Both | 29 (3.9) | 23 (3.8) | 6 (4.0) | | |
| BMI (kg/m²) | 24.18 ± 3.39 | 24.20 ± 3.38 | 24.11 ± 3.42 | 0.29 | 0.78 |
| CLI (%) | | | | 1.03 | 0.31 |
| Yes | 275 (36.7) | 225 (37.6) | 50 (33.1) | | |
| No | 475 (63.3) | 374 (62.4) | 101 (66.9) | | |
| FT (%) | | | | 3.10 | 0.21 |
| Unilateral ankle fracture | 597(79.6) | 473 (79.0) | 124 (82.1) | | |
| Bilateral ankle fracture | 38(5.1) | 28 (4.7) | 10 (6.6) | | |
| Trimalleolar ankle fracture | 115 (15.3) | 98 (16.3) | 17 (11.3) | | |

(Continued)

TABLE 1 (Continued)

| Clinical characteristics | Total (<i>n</i> = 750) | Training set (n = 599) | Testing set (<i>n</i> = 151) | $\chi^2/t/Fisher$ | p |
|-----------------------------|-------------------------|---------------------------|----------------------------------|-------------------|------|
| NI (%) | | | | 1.23 | 0.27 |
| Yes | 145 (19.3) | 111 (18.5) | 34 (22.5) | | |
| No | 605 (80.7) | 488 (81.5) | 117 (77.5) | | |
| OF (%) | | | | 1.12 | 0.29 |
| Yes | 146 (19.5) | 112 (18.7) | 34 (22.5) | | |
| No | 604 (80.5) | 487 (81.3) | 117 (77.5) | | |
| CJD (%) | | | | 0.26 | 0.61 |
| Yes | 140 (18.7) | 114 (19.0) | 26 (17.2) | | |
| No | 610 (81.3) | 485 (81.0) | 125 (82.8) | | |
| SWT (days) | 7.24 ± 3.40 | 7.26 ± 3.38 | 7.17 ± 3.50 | 0.30 | 0.76 |
| PUBSD (%) | | | | 0.44 | 0.51 |
| Yes | 634 (84.5) | 509 (85.0) | 125 (82.8) | | |
| No | 116 (15.5) | 90 (15.0) | 26 (17.2) | | |
| PHb (g/L) | 118.46 ± 17.71 | 118.33 ± 17.40 | 119.01 ± 18.92 | -0.42 | 0.76 |
| PAlb (g/L) | 41.39 ± 4.64 | 41.38 ± 4.63 | 41.41 ± 4.66 | -0.07 | 0.95 |
| PRBC (×10 ¹² /L) | 4.03 ± 0.61 | 4.03 ± 0.60 | 4.03 ± 0.63 | 0.10 | 0.92 |
| OT (min) | 108.51 ± 45.31 | 108.99 ± 46.12 | 106.63 ± 42.03 | 0.57 | 0.57 |
| PDT (%) | | | | 0.003 | 0.96 |
| Yes | 242 (32.2) | 193 (32.2) | 49 (32.4) | | |
| No | 508 (67.7) | 406 (67.8) | 102 (67.6) | | |
| ASA (%) | | | | | 0.41 |
| 1 | 671 (89.5) | 538 (89.8) | 133 (88.1) | | |
| 2 | 71 (9.5) | 56 (9.3) | 15 (9.9) | | |
| 3 | 8 (1.1) | 5 (0.9) | 3 (2.0) | | |
| LHS (days) | 15.10 ± 6.30 | 15.12 ± 6.27 | 15.06 ± 6.44 | 0.10 | 0.92 |
| VT(%) | | | | 2.115 | 0.15 |
| Yes | 54 (7.2) | 39 (6.5) | 15 (9.9) | | |
| No | 696 (92.8) | 560 (93.5) | 136 (90.1) | | |
| SSI (%) | | | | 1.63 | 0.20 |
| Yes | 35 (4.7) | 25 (4.2) | 10 (6.6) | | |
| No | 715 (95.3) | 574 (95.8) | 141 (93.4) | | |
| PPLAFE (%) | | | | 0.63 | 0.74 |
| Low | 367 (48.9) | 294 (49.1) | 73 (48.3) | | |
| Medium | 325 (43.3) | 261 (43.6) | 64 (42.4) | | |
| High | 58 (7.7) | 44 (7.3) | 14 (9.3) | | |
| PFEC (%) | | | | 1.58 | 0.46 |
| Low | 234 (31.2) | 183 (20.6) | 51 (33.8) | | |
| Medium | 327 (43.6) | 268 (44.7) | 59 (39.1) | | |
| High | 189 (25.2) | 148 (24.7) | 41 (27.1) | | |
| PR (%) | | | | <0.001 | 0.99 |
| Yes | 248 (33.1) | 198 (33.1) | 50 (33.1) | | |
| No | 502 (66.9) | 401 (66.9) | 101 (66.9) | | |

VV, Varicose Veins; HD, Heart Disease; CD, Cerebrovascular Disease; IM, Injury Mechanism; IS, Injury Site; BMI, Body Mass Index; CLI, Combined Ligament Injury; NI, Nerve Injury; CJD, Combined Joint Dislocation; OF, Open Fracture; FT, Fracture Type; SWT, Surgical Waiting Time; PUBSD, Perioperative Use of Blood-activating and Stasis-removing Drugs; PHb, Postoperative Hemoglobin; PAlb, Postoperative Albumin; PRBC, Postoperative Red Blood Cell Count; PDT, Postoperative Drainage Tube; OT, Operation Time; LHS, Length of Hospital Stay; VT, Venous Thrombosis; SSI, Surgical Site Infection; ASA, The American Society of Anesthesiologists; PPLAFE, Postoperative Pain Level After Functional Exercise; PFEC, Postoperative Functional Exercise Compliance.



Level_Primary school and below; Education Level_Medium: Education Level_Middle school to high school; Education Level_High: University and above; PPLAFE, Postoperative Pain Level After Functional Exercise; IS, Injury Site; BMI, Body Mass Index; CLI, Combined Ligament Injury; NI, Nerve Injury; CJD, Combined Joint Dislocation; OF, Open Fracture; FT, Fracture Type; SWT, Surgical Waiting Time; PHb, Postoperative Hemoglobin; OT, Operation Time; SSI, Surgical Site Infection; ASA, The American Society of Anesthesiologists; PFEC, Postoperative Functional Exercise Compliance.

3.2 Feature selection

As shown in Figure 2, using the Boruta algorithm, 12 potential predictors were selected based on the importance of their Z-scores. The selected predictors are as follows: postoperative albumin, operation time, injury mechanism, injury site, combined ligament injury, fracture severity, combined nerve injury, open fracture, combined joint dislocation, postoperative infection, pain level after postoperative rehabilitation exercise, and compliance with functional exercise.

3.3 Model construction, evaluation, and comparison

In both the training and testing sets, all models achieved accuracy and AUC values above 0.60, with Brier scores less than 0.25. The DeLong test *p*-values for AUC between the training and testing sets were all greater than 0.05 (Tables 2–4). Calibration curves indicated that all models demonstrated good calibration (Figure 3). The ROC curves for the training and testing sets are shown in Figure 4. Among the models, the L-S model exhibited the

TABLE 2 Performance of five machine learning-based models for predicting poor joint recovery after ankle fracture in the training set.

| Model | Delong test p-value (vs. LR model) | DeLong test <i>p</i> - value (vs. L-S model) | AUC(95%CI) | Accuracy | Карра | Precision | Specificity | Sensitivity | F1 score | Brier score |
|--------------------|--|--|----------------------|----------|-------|-----------|-------------|-------------|-------------|----------------|
| RF | < 0.001 | < 0.001 | 0.840(0.807 ~ 0.874) | 0.781 | 0.521 | 0.650 | 0.805 | 0.732 | 0.689 | 0.169 |
| LR | - | < 0.001 | 0.672(0.626 ~ 0.719) | 0.651 | 0.282 | 0.480 | 0.643 | 0.667 | 0.558 | 0.203 |
| SVM | <0.001 | < 0.001 | 0.809(0.770 ~ 0.848) | 0.751 | 0.481 | 0.593 | 0.733 | 0.788 | 0.677 | 0.169 |
| XGBoost | < 0.001 | < 0.001 | 0.734(0.692 ~ 0.776) | 0.668 | 0.326 | 0.498 | 0.641 | 0.722 | 0.590 | 0.198 |
| Lasso- stacking | <0.001 | - | 0.877(0.847 ~ 0.906) | 0.796 | 0.566 | 0.654 | 0.788 | 0.813 | 0.725 | 0.18 |

AUC, area under the curve; 95%CI, confidence interval.

TABLE 3 Performance of five machine learning-based models for predicting poor joint recovery after ankle fracture in the testing set.

| Model | DeLong test <i>p-</i> value (vs. LR model) | DeLong test <i>p-</i> value (vs. L-S model) | AUC(95%CI) | Accuracy | Карра | Precision | Specificity | Sensitivity | F1 score | Brier score |
|--------------------|--|---|----------------------|----------|-------|-----------|-------------|-------------|-------------|----------------|
| RF | 0.005 | 0.257 | 0.779(0.698 ~ 0.860) | 0.742 | 0.448 | 0.590 | 0.752 | 0.720 | 0.649 | 0.190 |
| LR | - | 0.003 | 0.691(0.603 ~ 0.779) | 0.656 | 0.307 | 0.486 | 0.624 | 0.720 | 0.581 | 0.199 |
| SVM | 0.053 | 0.314 | 0.768(0.686 ~ 0.851) | 0.728 | 0.446 | 0.563 | 0.693 | 0.800 | 0.661 | 0.183 |
| XGBoost | 0.031 | 0.153 | 0.748(0.663 ~ 0.832) | 0.722 | 0.430 | 0.557 | 0.693 | 0.780 | 0.650 | 0.198 |
| Lasso- stacking | 0.003 | - | 0.791(0.711 ~ 0.871) | 0.762 | 0.502 | 0.606 | 0.743 | 0.800 | 0.690 | 0.193 |

AUC, area under the curve; 95%CI, confidence interval.

TABLE 4 Results of AUC DeLong test between training set and testing set of five machine learning.

| Model | Training set | | Test | ting set | z | p | |
|----------------|--------------|--------------------|-------|---------------|--------|-------|--|
| | AUC | 95%CI | AUC | 95%CI | | | |
| RF | 0.840 | $0.807 \sim 0.874$ | 0.779 | 0.698 ~ 0.860 | 1.373 | 0.171 | |
| LR | 0.672 | 0.626 ~ 0.719 | 0.691 | 0.603 ~ 0.779 | -0.377 | 0.707 | |
| SVM | 0.809 | $0.770 \sim 0.848$ | 0.768 | 0.686 ~ 0.851 | 0.881 | 0.379 | |
| XGBoost | 0.734 | $0.692 \sim 0.776$ | 0.748 | 0.663 ~ 0.832 | -0.287 | 0.775 | |
| Lasso-stacking | 0.877 | 0.847 ~ 0.906 | 0.791 | 0.711 ~ 0.871 | 1.968 | 0.051 | |

AUC, area under the curve; 95%CI, confidence interval.

highest AUC (training set: 0.877, testing set: 0.791) and accuracy (training set: 0.796, testing set: 0.762), indicating strong and robust discriminatory ability. The LR model is often used as a traditional baseline model; therefore, the DeLong test was employed to compare the AUC of other models with both the L-S model and the LR model. In the training set, the Delong test *p*-values (both vs. the LR model and vs. the lasso-stacking model) were below 0.05. In the testing set, the DeLong test p-values compared to the LR model were less than 0.05 except for the SVM model, while p-values compared to the lasso-stacking model were greater than 0.05 except for the LR model. The Lasso-Stacking model also performed better or comparably to other models in terms of accuracy, Kappa, precision, specificity, sensitivity, and F1 score in both the training and testing sets. Additionally, the L-S model achieved the highest area under the PR curve among the five models (training set: 0.808, testing set: 0.634; Figure 5). Decision curve analysis revealed that the SVM model and the lasso-stacking model provided the highest clinical net benefit (Figure 6). In addition, the data were randomly grouped according to the ratio of 3:7 (dataset 1 and dataset 2), and the performance of the model in the two datasets was consistent with that observed in the training set and test set. Detailed results are presented in Supplementary Table S1S1 and S2 in the Appendix. DeLong test *p*-values > 0.05 indicate that each model demonstrated more robustness. Furthermore, the data were stratified based on demographic characteristics and fracture types. The performance of each model across different demographic characteristics is detailed in Supplementary Table S4 in the Appendix, and the performance of each model across different fracture types is detailed in Supplementary Table S5 in the Appendix. Accuracy and ROC curves of each model across different demographic





characteristics and across different fracture types are compared in Appendix Supplementary Figures S1–S4. The comparison showed that the lasso-stacking model consistently exhibited the best performance. In conclusion, the lasso-stacking model emerged as the optimal model and appears most suitable for predicting poor

3.4 Model interpretation

Through SHAP analysis, clinical practitioners can gain insights into the decision-making basis of the lasso-stacking model. Figure 7A presents a SHAP summary plot where feature variables are sorted based on SHAP importance, from highest to lowest.

recovery after ankle fracture surgery.





Figure 7B is a bar chart where feature variables are arranged according to the mean absolute SHAP value, also in descending order. The SHAP analysis reveals that the variables, in terms of importance from greatest to least, are as follows: functional exercise compliance, concomitant ligament injury, open fracture, nerve injury, fracture severity, injury mechanism, concomitant joint dislocation, postoperative albumin level, pain level after postoperative rehabilitation exercise, operation duration, injury site,

and postoperative infection. Figure 7C illustrates the SHAP plots for categorical variables, while Figure 7D showcases the SHAP plots for continuous variables. Figure 7E demonstrates the contribution of various features to the model's prediction for a single sample of an ankle fracture patient who did not experience poor recovery after surgery. Meanwhile, Figure 8 utilizes the LIME algorithm to provide additional explanations for individual prediction results, complementing the interpretability analysis offered by SHAP.



FIGURE 7

SHAP plots. (A) SHAP summary plot shows feature importance for each predictor of the Lasso-Stacking model in descending order. The upper predictors are more important to the model's predictive outcome. For each patient's Lasso-Stacking model, a point should be created for each feature attribute value. The distance of a point from the baseline SHAP value of zero indicates the strength of its effect on the model output. The points are coloured according to the value of the feature, with yellow representing high feature values and red representing low feature values. (B) Bar chart of mean absolute SHAP for each predictor of the Lasso-Stacking model in descending order (C) SHAP chart for each categorical variable. For each patient's Lasso-Stacking model, a point should be created for each feature attribute value. The distance of a point from the baseline SHAP value of zero indicates the strength of its effect on the model output. Yellow for positive results, purple for negative results (D) SHAP chart for each continuous variable. For each patient's Lasso-Stacking model, a point should be created for each feature attribute value. The distance of a point from the baseline SHAP value of zero indicates the strength of its effect on the model output. Yellow for positive results, purple for negative results (D) SHAP chart for each continuous variable. For each patient's Lasso-Stacking model, a point should be created for each feature attribute value. The distance of a point from the baseline SHAP value of zero indicates the strength of its effect on the model output. Yellow for positive results, purple for negative results (D) SHAP chart for each continuous variable. For each patient's Lasso-Stacking model, a point should be created for each feature attribute value. The distance of a point from the baseline SHAP value of zero indicates the strength of its effect on the model output. Yellow for positive results, purple for negative results (E) The force plots provide personalized feature attributions using one examples. I

4 Discussion

In this study, the ankle function of patients 3 months after ankle fracture surgery was evaluated using the AOFAS scale. It was found

that 33.1% of patients scored less than 75 on the AOFAS, indicating poor joint function recovery after surgery. Despite continuous improvements in surgical techniques and rehabilitation exercises for ankle fractures in recent years, the incidence of poor joint recovery



LIME algorithm explains individual prediction results plot. Parsed as an example of an ankle fracture. The picture shows the predicted expected probability of 80% poor postoperative recovery, estimated by the Lasso-Stacking model. This probability was determined by the predictive model. The length of each feature bar is proportional to the weight of that feature in the prediction. Longer bars represent features that contribute more to the predicted outcome. IM, Injury Mechanism; IS, Injury Site; CLI, Combined Ligament Injury; NI, Nerve Injury; CJD, Combined Joint Dislocation; OF, Open Fracture; FT, Fracture Type; PHb, Postoperative Hemoglobin; OT, Operation Time; SSI, Surgical Site Infection; PPLAFE, Postoperative Pain Level After Functional Exercise; PFEC, Postoperative Functional Exercise Compliance; TAI, Traffic accident injury; UAF, Unilateral ankle fracture; BAF, Bilateral ankle fracture; TAF, Trimalleolar ankle fracture.

after surgery remains high. Therefore, early identification of high-risk groups and prompt implementation of relevant intervention measures for these groups are key to promoting good joint function recovery.

In this study, we constructed risk prediction models for poor functional recovery after ankle fracture surgery using five common machine learning algorithms: random forest, XGBoost, SVM, logistic regression, and lasso-stacking. After evaluating each model using metrics such as AUC, we found that all models were capable of predicting the occurrence of poor joint recovery after ankle fracture surgery to some extent, but there were significant differences in their predictive performance. The choice of machine learning algorithm mainly depends on the distribution of feature variables and model fitness (41). Logistic regression, as the simplest and most basic machine learning algorithm, performs well in predicting linear relationships between its characteristic variables and outcomes but is less effective for predicting non-linear relationships. In this study, it demonstrated the worst predictive performance. Random forest, a type of bagging integration algorithm, combines multiple decision trees. It constructs a random forest by integrating multiple decision trees and makes predictions based on the voting results of the random forest (42). In this study, the performance of the random forest is relatively stable. Compared to other models, the random forest model excels particularly when analyzing higherdimensional data (43), which was well demonstrated in this study. Its overall predictive performance ranked second only to the lasso-stacking model. XGBoost may have advantages in predicting higher-dimensional data, but it may perform poorly in lower-dimensional datasets, with only 12 feature variables in this case (40). This limitation likely prevented it from fully leveraging its algorithmic strengths in this study. SVM operates by solving a hyperplane that can correctly classify the training data and maximize the geometric margin, making it effective in handling binary classification problems (44). This is well reflected in our study, where its overall predictive performance is slightly inferior to the lassostacking and random forest algorithms. Finally, the lasso-stacking model exhibited the best predictive performance in this study. The stacking model used in this study integrates the four different types of base models mentioned above, resulting in improved performance of the fused model. Under the fusion framework, the previously output results are input into a second-layer learner to obtain a better-performing prediction model. The lasso-stacking model in this study effectively learns from the advantages of the first four models and performs better in predictive performance than the first four models. Therefore, it is more suitable for predicting the recovery of patients' joints after ankle fracture surgery. Current evaluation of post-ankle fracture joint function predominantly relies on the American Orthopaedic Foot & Ankle Society (AOFAS) score. While this instrument effectively assesses postoperative functional recovery, it lacks prospective predictive capability. To address this limitation, our study developed and compared five machine learning models, ultimately selecting the lasso-stacking ensemble as the optimal solution. This model leverages preoperative and perioperative clinical indicators from hospitalized patients to predict 3-month postoperative functional outcomes, providing actionable insights for clinicians to tailor individualized rehabilitation protocols. The lasso-stacking algorithm demonstrated superior discriminative performance, achieving area under the ROC curve (AUC) values of 0.877 (training set) and 0.791 (test set). Model robustness was rigorously validated through stratified analyses incorporating random data partitioning, demographic stratification, and fracture-type subgrouping, confirming consistent predictive stability across heterogeneous patient cohorts.

Through interpretive analysis of the optimal model, this study reveals that postoperative functional exercise compliance is the most critical factor influencing joint function recovery for patients after ankle fracture surgery. Numerous studies have shown (45–47) that early rehabilitation exercise is essential for the recovery of joint function after ankle fracture surgery. However, most clinicians currently focus on improving rehabilitation methods while paying little attention to patient compliance with these exercises. This has led to poor joint function recovery in some patients despite receiving feasible rehabilitation training guidance. Therefore, in clinical practice, besides instructing patients on correct rehabilitation methods, it is also crucial to urge them to perform effective exercises, thus reducing the risk of poor postoperative joint function recovery.

Additionally, fracture-related factors such as associated ligament damage, nerve damage, joint dislocation, injury mechanism, open

fractures, injury location, and fracture severity also significantly impact joint function recovery after ankle fracture surgery. Among these, associated ligament damage is the most critical, possibly due to the instability it causes in joint function and the increased risk of ligament rerupture with early and extensive functional exercises (11, 48). This can lead to a fear of postoperative exercises among some patients, affecting their recovery. Hence, such patients require closer follow-up and individualized exercise plans.

Furthermore, postoperative albumin levels and operation time can also affect joint function recovery. According to SHAP, low postoperative albumin levels and prolonged surgical time increase the risk of poor joint function recovery. Patients with low postoperative albumin often have poorer nutritional status and may experience weakness, affecting their rehabilitation efforts (49, 50). In addition, albumin also plays roles in anti-oxidation, inflammation regulation, and immune response modulation (51, 52). In the state of hypoalbuminemia, the human body is more prone to trigger systemic inflammatory responses and immune suppression, resulting in elevated levels of C-reactive protein (CRP) and interleukin-6 (IL-6) in the body. This can cause tissue edema and stiffness around the joint, increase the risk of incision or joint cavity infections, and delay the recovery process. In addition, albumin plays a key role in maintaining plasma colloid osmotic pressure (52). A reduction in albumin levels can lead to interstitial fluid retention, resulting in postoperative limb swelling, stiffness, and pain, which in turn may inhibit patients' willingness to perform functional exercises. Albumin deficiency can also affect collagen synthesis, cell proliferation, and matrix reconstruction (53), which will lead to the decline of the regeneration ability of fibroblasts, muscle cells, and chondrocytes, thus affecting wound healing and tissue repair and directly hindering the recovery of joint function. Therefore, it is an important factor that should receive more attention from clinical workers. Longer surgical procedures can lead to increased intraoperative bleeding and a higher risk of postoperative malnutrition. In addition, long-term surgery leads to soft tissue exposure, stretching, and electric coagulation for a long time, which aggravates local inflammation and scar formation, and affects the functional recovery of muscles and ligaments. Furthermore, prolonged stretch or compression can lead to muscle ischemia-reperfusion injury and impair local tissue activity. Prolonged use of anesthetics (especially neuromuscular blocking agents) may affect postoperative nerve activation and muscle tone recovery and delay the progress of functional exercise. Finally, prolonged operation time may increase the risk of infection and complications. Excessive pain during postoperative rehabilitation exercises can also cause fear among patients, disrupting their rehabilitation plans and affecting joint recovery (12, 54).

Finally, while it is widely known that postoperative infection can impact recovery and limb function, its influence was found to be relatively minor in this study. This may be due to advancing surgical techniques in recent years, which have reduced the incidence of postoperative infections and subsequent poor recovery, thereby diminishing the significance of this factor.

In this study, the development of a risk prediction model for postoperative poor joint function recovery in patients with ankle fractures is innovative. The constructed model demonstrated strong performance and can be preliminarily applied to support early and accurate prevention and control in clinical practice. In addition, interpretability analysis of the model identified important factors influencing poor joint function recovery in patients with ankle fracture after surgery, which provides some insights for clinical workers to intervene early. Finally, this study may serve as a foundation for future research on causal relationships and intervention strategies. However, there are certain limitations in this study. First, as a single-center study, it only validated internal data without external validation of the model, which may have an impact on the generalizability and robustness of the findings. Second, some data were incomplete, such as the degree of limb swelling and rehabilitation exercise methods, which were not included in the analysis. Previous studies have shown (15, 18) that rehabilitation exercises involving removable ankle support and different exercise duration can influence the recovery of joint function after ankle fracture surgery. However, this type of data is not optimized at present, and the lack of this metric may have an impact on the performance of the model. Finally, this study did not consider the changes in patients' characteristics after discharge. Therefore, in the future, it is necessary to strengthen regional cooperation, further optimizing data collection and processing, continuously refining the model for different population data, and improving its accuracy. After multi-dimensional verification of the model, a visual early warning platform can be constructed to enable clinical workers to quickly identify high-risk patients and implement accurate prevention and control in the early stage.

5 Conclusion

Finally, this study developed five machine learning models and found that the lasso-stacking model showed the best performance, making it the most suitable for predicting high-risk populations with poor joint function recovery after ankle fracture surgery. Explanatory analysis of this model helped clarify its decision-making basis. In the future, the model should be verified and improved through multicenter external validation to support its application in clinical practice.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material; further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving humans were approved by Medical Ethics Committee of Lu 'an People's Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

CL: Conceptualization, Data curation, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. CW: Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. JZ: Conceptualization, Data curation, Supervision, Validation, Writing – original draft, Writing – review & editing. WZ: Investigation, Project administration, Software, Supervision, Validation, Writing – review & editing. JS: Funding acquisition, Investigation, Project administration, Resources, Writing – original draft. LL: Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing – review & editing. XS: Methodology, Project administration, Resources, Software, Supervision, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Jensen SL, Andresen BK, Mencke S, Nielsen PT. Epidemiology of ankle fractures: a prospective population-based study of 212 cases in Aalborg, Denmark. *Acta Orthop Scand.* (1998) 69:48–50. doi: 10.3109/17453679809002356

2. Mandi DM, Nickles WA, Mandracchia VJ, Halligan JB, Toney PA. Ankle fractures. *Clin Podiatr Med Surg.* (2006) 23:375–422. doi: 10.1016/j.cpm.2006.02.001

3. Scheer RC, Newman JM, Zhou JJ, Oommen AJ, Naziri Q, Shah NV, et al. Ankle fracture epidemiology in the United States: patient-related trends and mechanisms of injury. *J Foot Ankle Surg.* (2020) 59:479–83. doi: 10.1053/j.jfas.2019.09.016 4

4. Wu AM, Bisignano C, James SL, Abady GG, Abedi A, Abu-Gharbieh E, et al. Global, regional, and national burden of bone fractures in 204 countries and territories, 1990–2019: a systematic analysis from the global burden of disease study 2019. *Lancet Healthy Longevity*. (2021) 2:e580–92. doi: 10.1016/S2666-7568(21)00172-0

5. Jennison T, Brinsden M. Fracture admission trends in England over a ten-year period. *Ann R College Surg Engl.* (2019) 101:208–14. doi: 10.1308/rcsann.2019.0002

6. Zhu Z, Zhang T, Shen Y, Shan PF. The burden of fracture in China from 1990 to 2019. Arch Osteoporos. (2023) 19:1. doi: 10.1007/s11657-023-01353-4

7. Hemmann P, Friederich M, Körner D, Klopfer T, Bahrs C. Changing epidemiology of lower extremity fractures in adults over a 15-year period-a National Hospital Discharge Registry study. *BMC Musculoskelet Disord.* (2021) 22:456. doi: 10.1186/s12891-021-04291-9

8. Kannus P, Palvanen M, Niemi S, Parkkari J, Jrvinen M. Increasing number and incidence of low-trauma ankle fractures in elderly people: Finnish statistics during 1970–2000 and projections for the future. *Bone.* (2002) 31:430–3. doi: 10.1016/s8756-3282(02)00832-3

9. Vanderkarr MF, Ruppenkamp JW, Vanderkarr M, Parikh A, Holy CE, Putnam M. Incidence, costs and post-operative complications following ankle fracture–a US claims database analysis. *BMC Musculoskelet Disord.* (2022) 23:1129. doi: 10.1186/s12891-022-06095-x

10. Belatti DA, Phisitkul P. Economic burden of foot and ankle surgery in the US Medicare population. *Foot Ankle Int.* (2014) 35:334–40. doi: 10.1177/1071100713519777

11. Carter TH, Duckworth AD, White TO. Medial malleolar fractures: current treatment concepts. *Bone Joint J.* (2019) 101-B:512-21. doi: 10.1302/0301-620x.101B5.BJJ-2019-0070

12. Dean DM, Ho BS, Lin A, Fuchs D, Ochenjele G, Merk B, et al. Predictors of patient-reported function and pain outcomes in operative ankle fractures. *Foot Ankle Int.* (2017) 38:496–501. doi: 10.1177/1071100716688176

13. Chong HH, Hau MYT, Mishra P, Rai P, Mangwani J. Patient outcomes following ankle fracture fixation. *Foot Ankle Int.* (2021) 42:1162–70. doi: 10.1177/10711007211003073

14. Mococain P, Bejarano-Pineda L, Glisson R, Kadakia RJ, Akoh CC, Chen J, et al. Biomechanical effect on joint stability of including deltoid ligament repair in an ankle fracture soft tissue injury model with deltoid and syndesmotic disruption. *Foot Ankle Int.* (2020) 41:1158–64. doi: 10.1177/1071100720929007

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2025.1553274/ full#supplementary-material

15. Lewis SR, Pritchard MW, Parker R, Searle HKC, Beckenkamp PR, Keene DJ, et al. Rehabilitation for ankle fractures in adults. *Cochrane Database Syst Rev.* (2024) 9:CD005595. doi: 10.1002/14651858.CD005595.pub4

16. Jansen H, Jordan M, Frey S, Hölscher-Doht S, Meffert R, Heintel T. Active controlled motion in early rehabilitation improves outcome after ankle fractures: a randomized controlled trial. *Clin Rehabil.* (2018) 32:312–8. doi: 10.1177/0269215517724192

17. Bartoníček J, Rammelt S, Tuček M. Posterior malleolar fractures: changing concepts and recent developments. *Foot Ankle Clin.* (2017) 22:125-45. doi: 10.1016/j.fcl.2016.09.009

18. Chen B, Ye Z, Wu J, Wang G, Yu T. The effect of early weight-bearing and later weight-bearing rehabilitation interventions on outcomes after ankle fracture surgery: a systematic review and meta-analysis of randomised controlled trials. *J Foot Ankle Res.* (2024) 17:e12011. doi: 10.1002/jfa2.12011

19. Keene DJ, Williamson E, Bruce J, Willett K, Lamb SE. Early ankle movement versus immobilization in the postoperative management of ankle fracture in adults: a systematic review and meta-analysis. *J Orthop Sports Phys Ther*. (2014) 44:690–C7. doi: 10.2519/jospt.2014.5294

20. Maki S, Furuya T, Inoue M, Shiga Y, Inage K, Eguchi Y, et al. Machine learning and deep learning in spinal injury: a narrative review of algorithms in diagnosis and prognosis. *J Clin Med.* (2024) 13:705. doi: 10.3390/jcm13030705

21. Liu F, Liu C, Tang X, Gong D, Zhu J, Zhang X. Predictive value of machine learning models in postoperative mortality of older adults patients with hip fracture: a systematic review and meta-analysis. *Arch Gerontol Geriatr.* (2023) 115:105120. doi: 10.1016/j.archger.2023.105120

22. Kuo RYL, Harrison CJ, Jones BE, Geoghegan L, Furniss D. Perspectives: a surgeon's guide to machine learning. *Int J Surg.* (2021) 94:106133. doi: 10.1016/j.ijsu.2021.106133

23. Lex JR, Di Michele J, Koucheki R, Pincus D, Whyne C, Ravi B. Artificial intelligence for hip fracture detection and outcome prediction: a systematic review and meta-analysis. *JAMA Netw Open*. (2023) 6:e233391. doi: 10.1001/jamanetworkopen.2023.3391

24. Cai Z, Sun Q, Li C, Xu J, Jiang B. Machine-learning-based prediction by stacking ensemble strategy for surgical outcomes in patients with degenerative cervical myelopathy. *J Orthop Surg Res.* (2024) 19:539. doi: 10.1186/s13018-024-05004-3

25. Barrosse-Antle ME, Patel KH, Kramer JA, Baston CM. Point-of-care ultrasound for bedside diagnosis of lower extremity DVT. *Chest.* (2021) 160:1853–63. doi: 10.1016/j.chest.2021.07.010

26. Needleman L, Cronan JJ, Lilly MP, Merli GJ, Adhikari S, Hertzberg BS, et al. Ultrasound for lower extremity deep venous thrombosis: multidisciplinary recommendations from the Society of Radiologists in ultrasound consensus conference. *Circulation*. (2018) 137:1505–15. doi: 10.1161/CIRCULATIONAHA.117.030687

27. O'Hara LM, Thom KA, Preas MA. Update to the Centers for Disease Control and Prevention and the healthcare infection control practices advisory committee guideline for the prevention of surgical site infection (2017): a summary, review, and strategies for implementation. *Am J Infect Control.* (2018) 46:602–9. doi: 10.1016/j.ajic.2018.01.018

29. Horvath B, Kloesel B, Todd MM, Cole DJ, Prielipp RC. The evolution, current value, and future of the American Society of Anesthesiologists Physical Status Classification System. *Anesthesiology*. (2021) 135:904–19. doi: 10.1097/ALN.00000000003947

30. Li G, Walco JP, Mueller DA, Wanderer JP, Freundlich RE. Reliability of the ASA physical status classification system in predicting surgical morbidity: a retrospective analysis. *J Med Syst.* (2021) 45:83–8. doi: 10.1007/s10916-021-01758-z

31. Tan Y, He H, Yang X, Li X, Mi J. Development and reliability and validity test of the compliance scale of functional exercise for orthopedic patients. *Chin Nurs Manag.* (2019) 19:1626–31. doi: 10.3969/j.issn.1672-1756.2019.11.007

32. Kitaoka HB, Alexander IJ, Adelaar RS, Nunley JA, Myerson MS, Sanders M. Clinical rating systems for the ankle-hindfoot, midfoot, hallux, and lesser toes. *Foot Ankle Int.* (1994) 15:349–53. doi: 10.1177/107110079401500701

33. de Boer AS, Tjioe RJC, Van der Sijde F, et al. The American Orthopaedic Foot and Ankle Society ankle-Hindfoot scale; translation and validation of the Dutch language version for ankle fractures. *BMJ Open*. (2017) 7:e017040. doi: 10.1136/bmjopen-2017-017040

34. Erichsen J, Froberg L, Viberg B, Damborg F, Jensen C. Danish language version of the American orthopedic foot and ankle society ankle-Hindfoot scale (AOFAS-AHS) in patients with ankle-related fractures. *J Foot Ankle Surg.* (2020) 59:657–63. doi: 10.1053/j.jfas.2019.08.027

35. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. J Stat Softw. (2010) 36:1–13. doi: 10.18637/jss.v036.i11

36. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med.* (2012) 22:276-82. doi: 10.11613/BM.2012.031

37. Loveleen G, Mohan B, Shikhar BS, Nz J, Shorfuzzaman M, Masud M. Explanationdriven HCI model to examine the mini-mental state for Alzheimer's disease. *ACM Trans Multimed Comput Commun Appl.* (2023) 20:1–16. doi: 10.1145/3527174

38. Bhandari M, Shahi TB, Siku B, Neupane A. Explanatory classification of CXR images into COVID-19, pneumonia and tuberculosis using deep learning and XAI. *Comput Biol Med.* (2022) 150:106156. doi: 10.1016/j.compbiomed.2022.106156

39. Patel AN, Murugan R, Srivastava G, Maddikunta PKR, Yenduri G, Gadekallu TR, et al. An explainable transfer learning framework for multi-classification of lung diseases in chest X-rays. *Alex Eng J.* (2024) 98:328–43. doi: 10.1016/j.aej.2024.04.072

40. Huang D, Gong L, Wei C, Wang X, Liang Z. An explainable machine learningbased model to predict intensive care unit admission among patients with communityacquired pneumonia and connective tissue disease. *Respir Res.* (2024) 25:246. doi: 10.1186/s12931-024-02874-3

41. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak*. (2019) 19:281–16. doi: 10.1186/s12911-019-1004-8

42. Shi L, Wei H, Zhang T, Li Z, Chi X, Liu D, et al. A potent weighted risk model for evaluating the occurrence and severity of diabetic foot ulcers. *Diabetol Metab Syndr.* (2021) 13:92–11. doi: 10.1186/s13098-021-00711-x

43. Zhang Y, Zhang Z, Wei L, Wei S. Construction and validation of nomograms combined with novel machine learning algorithms to predict early death of patients with metastatic colorectal cancer. *Front Public Health.* (2022) 10:1008137. doi: 10.3389/fpubh.2022.1008137

44. Shia WC, Chen DR. Classification of malignant tumors in breast ultrasound using a pretrained deep residual network model and support vector machine. *Comput Med Imaging Graph*. (2021) 87:101829. doi: 10.1016/j.compmedimag.2020.101829

45. Matthews PA, Scammell BE, Coughlin TA, Nightingale J, Ollivere BJ. Early motion and directed exercise (EMADE) following ankle fracture fixation: a pragmatic randomized controlled trial. *Bone Joint J.* (2024) 106-B:949–56. doi: 10.1302/0301-620X.106B9.BJJ-2023-1433.R1

46. Zhao K, Dong S, Wang W. When is the optimum time for the initiation of early rehabilitative exercise on the postoperative functional recovery of peri-ankle fractures? A network meta-analysis. *Front Surg.* (2022) 9:911471. doi: 10.3389/fsurg.2022.911471

47. Altuwairqi A. Comparative analysis of rehabilitation strategies following ankle fracture surgery: a systematic review. *Cureus.* (2024) 16:e64315. doi: 10.7759/cureus.64315

48. Lee S, Lin J, Hamid KS, Bohl DD. Deltoid ligament rupture in ankle fracture: diagnosis and management. *J Am Acad Orthop Surg.* (2019) 27:e648. doi: 10.5435/JAAOS-D-18-00198

49. Ljungqvist O, Weimann A, Sandini M, Baldini G, Gianotti L. Contemporary perioperative nutritional care. *Annu Rev Nutr.* (2024) 44:231–55. doi: 10.1146/annurev-nutr-062222-021228

50. Martindale RG. Novel nutrition strategies to enhance recovery after surgery. J Parenter Enter Nutr. (2023) 47:476–81. doi: 10.1002/jpen.2485. Epub 2023 Mar 20

51. Pompili E, Zaccherini G, Baldassarre M, Iannone G, Caraceni P. Albumin administration in internal medicine: a journey between effectiveness and futility. *Eur J Intern Med.* (2023) 117:28–37. doi: 10.1016/j.ejim.2023.07.003

52. Belinskaia DA, Voronina PA, Shmurak VI, Jenkins RO, Goncharov NV. Serum albumin in health and disease: esterase, antioxidant, transporting and signaling properties. *Int J Mol Sci.* (2021) 22:10318. doi: 10.3390/ijms221910318

53. Kuten Pella O, Hornyák I, Horváthy D, Fodor E, Nehrer S, Lacza Z. Albumin as a biomaterial and therapeutic agent in regenerative medicine. *Int J Mol Sci.* (2022) 23:10557. doi: 10.3390/ijms231810557

54. Rethman KK, Mansfield CJ, Moeller J, de Oliveira Silva D, Stephens JA, di Stasi S, et al. Kinesiophobia is associated with poor function and modifiable through interventions in people with patellofemoral pain: a systematic review with individual participant data correlation meta-analysis. *Phys Ther.* (2023) 103:pzad074. doi: 10.1093/ptj/pzad074