



OPEN ACCESS

EDITED BY

Frank Hassel,
Loretto Hospital, Germany

REVIEWED BY

Babak Saravi,
University of Freiburg Medical
Center, Germany
Sherwan Hamawandi,
Hawler Medical University, Iraq

*CORRESPONDENCE

He-fan He
✉ 15860905262@163.com

[†]These authors share first authorship

RECEIVED 28 July 2024

ACCEPTED 21 August 2025

PUBLISHED 11 September 2025

CITATION

Sun Z-D, Fang Y-M, Lin Y-L, Pei M-Q, Liu C-Y and He H-f (2025) Construction and validation of a perioperative blood transfusion model for patients undergoing total hip arthroplasty with osteonecrosis of the femoral head based on machine learning. *Front. Med.* 12:1471746. doi: 10.3389/fmed.2025.1471746

COPYRIGHT

© 2025 Sun, Fang, Lin, Pei, Liu and He. 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.

Construction and validation of a perioperative blood transfusion model for patients undergoing total hip arthroplasty with osteonecrosis of the femoral head based on machine learning

Zhen-Dong Sun[†], Yu-Ming Fang[†], Yan-Ling Lin, Meng-Qin Pei, Chu-Yun Liu and He-fan He*

Department of Anesthesiology, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, Fujian, China

Background: This study aimed to construct a predictive model utilizing multiple machine learning (ML) models to estimate the likelihood of perioperative blood transfusion in patients with osteonecrosis of the femoral head (ONFH) who underwent total hip arthroplasty (THA).

Methods: Patients diagnosed with ONFH who underwent THA at our institution between October 2018 and October 2023 were included in the study. Feature selection was conducted using Lasso regression and correlation analysis. An unbiased evaluation framework incorporating nested resampling was established to assess four ML models. A nomogram was subsequently developed based on the selected features.

Results: Seven features were identified, namely blood loss, hemoglobin (HGB) levels, weight, body temperature, systolic pressure, and direct bilirubin. Four ML models were constructed based on these features. The area under the curve (AUC) values for Random Forest, Extreme Gradient Boosting, Light Gradient Boosting Machine, and Logistic Regression (LR) were 1.00, 1.00, 1.00, and 0.93 in the internal validation set, and 0.89, 0.90, 0.88, and 0.91 in the external test set, respectively. Furthermore, a nomogram model based on LR was developed using the aforementioned seven features, yielding AUC values of 0.95 and 0.90 for the training and test sets, respectively, thereby surpassing the AUC values of preoperative HGB levels (0.80 and 0.76).

Conclusion: Both the ML models and the nomogram exhibit significant potential for forecasting the likelihood of perioperative blood transfusion in patients with ONFH undergoing THA, which may aid clinicians in improving the accuracy of blood transfusion predictions.

KEYWORDS

osteonecrosis of the femoral head, total hip arthroplasty, blood transfusion, machine learning, risk prediction, nomogram model

Introduction

Osteonecrosis of the femoral head (ONFH) is characterized by the necrosis of osteocytes resulting from ischemia of the femoral head (1). The global incidence is anticipated to reach 20 million within a decade (2), with an estimated 15,000–20,000 annual cases occurring in the United States (3). Currently, a definitive prevention method for the prevention for ONFH remains elusive, as non-surgical interventions apart from total hip arthroplasty (THA), are typically ineffective (4). However, THA has the potential to result in substantial blood loss, necessitating perioperative red blood cells (RBC) transfusions in up to 20% of patients (5, 6). Furthermore, allogeneic blood transfusions can lead to various adverse effects, including increased risks of surgical site infections, deep vein thrombosis, and mortality (7–9). Although some strategies can reduce the need for blood transfusions (10), these approaches may not be suitable for all patients due to associated risks (11).

Consequently, identifying patients who require perioperative blood transfusions is essential for adequate preparation and preventive measures. While physicians typically rely on hemoglobin (HGB) levels to inform transfusion decisions, it is crucial to also consider various perioperative factors such as age (12), sex (13), and laboratory values like alanine aminotransferase. Therefore, screening the variables that significantly influence blood transfusion and integrating them into a cohesive predictive model may enhance accuracy.

Machine learning (ML), a subset of artificial intelligence, employs computational modeling to extract insights from data. It improves the accuracy of predictions and classifications by utilizing large datasets for model training and continuously optimizing performance (14–16). ML has shown promise in clinical predictions across various medical specialties, including perioperative blood transfusion predictions (17, 18). However, there is limited scholarly research on the use of ML to predict perioperative transfusion requirements in patients with ONFH.

Thus, the objective of this research was to develop a predictive model to more accurately predict the likelihood of perioperative blood transfusion in patients with ONFH undergoing THA.

Materials and methods

Ethics statement and patient selection

This study adhered to the principles outlined in the Declaration of Helsinki and its amendments and received approval from the Ethics Committee at our hospital. The requirement for written informed consent was waived, and personal identifiers were anonymized prior to data analysis.

A retrospective analysis was conducted on a cohort of 503 patients with ONFH who underwent THA at our hospital between October 2018 and October 2023. The administration of packed RBC during hospital stays, referred to as perioperative blood transfusion, was determined through consultation between the anesthesiologist and the surgeon. Typically, a HGB concentration below 70 g/L indicates the necessity for a blood transfusion.

Potential risk factors

We collected and evaluated the following features of the participants. General information: gender, age, the weight, height, and body mass index at the time of admission. Characteristics of surgery: operative site, surgery routes (anterior, posterior, and lateral approach), incisional length, intraoperative blood loss volume. Laboratory test results originated from the first post-admission blood draw: alanine aminotransferase [ALT], aspartate aminotransferase [AST], albumin [ALB], direct bilirubin, indirect bilirubin, international normalized ratio [INR], prothrombin time [PT], activated partial thromboplastin time, fibrinogen, HGB levels, and platelet count. Others: vital signs at admission, intraoperative hemostatic drugs utility.

Statistical analysis

Initially, we included 28 variables and had a sample size of 503, yielding a 15 EPV (Events per Predictor Variable), which surpasses the commonly recommended benchmark of 10 EPV (19, 20). Then, we standardized the continuous variables. The feature selection process utilized Lasso regression analysis with 10-fold cross-validation across the complete dataset. Features exhibiting pairwise correlation coefficients ≥ 0.6 (Pearson) were subsequently excluded through collinearity screening.

Following this, we extensively resampled non-blood transfusion samples to achieve a 1:1 ratio with blood transfusion samples. We then subjected four ML models—Random Forest (RF), Extreme Gradient Boosting (XGB), Light Gradient Boosting Machine (LGBM), and Logistic Regression (LR)—to hyperparameter tuning and performance evaluation via a nested sampling approach. In the inner loop, all parameter combinations were explored to identify optimal hyperparameters using random search, followed by model training. A four-fold cross-validation was employed for hyperparameter tuning, while a five-fold cross-validation was applied in the outer loop for model evaluation, ensuring the reliability and robustness of performance metrics. Consequently, the entire dataset was divided into training, internal validation, and external test sets in each fold. The evaluation of each ML model's performance was based on the average level of nested sampling. Subsequently, the performance of each model was assessed using receiver operating characteristic (ROC) curves, with a specific focus on the area under the curve (AUC) values. The DeLong test was utilized to verify statistical differences in AUC values among four ML models. Additionally, we employed various model performance evaluation metrics, including sensitivity, specificity, Brier scores, F-beta coefficient, positive predictive value (PPV), negative predictive value (NPV), Log Loss, and Matthews's correlation coefficient (MCC) to comprehensively assess model performance on both the internal validation set and the external test set.

Furthermore, bootstrap resampling was performed to partition the entire dataset into training and validation sets at a ratio of 503:186. Subsequently, a LR-based nomogram model was developed to predict the probability of perioperative blood transfusion based on the selected features. Model performance was

rigorously evaluated using calibration curves, ROC curves, and decision curve analysis (DCA) in both the training and external validation sets. Statistical analyses were conducted using R software (version 4.4.1), with two-tailed tests applied throughout. Statistical significance was defined as $P < 0.05$.

Results

Clinical characteristics

In the entire dataset, notable differences in gender, weight, systolic pressure, diastolic pressure, operative site, blood loss, ALT, AST, ALB, indirect bilirubin, NIR, PT, and HGB levels were observed between patients in the transfusion and non-transfusion groups, with P below 0.05 as reported in Table 1. The remaining aggregate data did not exhibit statistically significant variances ($P > 0.05$).

Feature selection

To mitigate overfitting, a two-step feature selection process was implemented. Initially, Lasso Regression with 10-fold cross-validation was employed to identify eight features exhibiting stable associations (Figures 1A, B), namely blood loss, HGB levels, weight, body temperature, systolic pressure, diastolic pressure, and direct bilirubin. Subsequently, redundancy was minimized by calculating pairwise Spearman's correlations among these features and removing one variable from any pair with $r \geq 0.6$, specifically excluding systolic blood pressure (Figure 1C). This strategy integrated regularized regression and correlation-based simplification to emphasize generalizable, non-redundant features. Binary classifications were constructed from original variables using clinically informed thresholds. Ultimately, seven variables—blood loss volume, HGB levels, weight, body temperature, systolic pressure, and direct bilirubin—were incorporated into the ML models.

Development and validation of ML models

Using the seven selected features, four ML models were developed within a nested resampling framework to reduce the risk of overfitting. Results from systematic hyperparameter tuning for each model are summarized in Supplementary Table S1. The ROC curves for the RF, XGB, LGBM, and LR models in the external test set are illustrated in Figure 2, while those for the internal validation set are depicted in Supplementary Figure S1. AUC values for these models were 1.00, 1.00, 1.00, and 0.93 in the internal validation set (Supplementary Figure S1) and 0.89, 0.90, 0.88, and 0.91 in the external test set (Figure 2), respectively.

Furthermore, Tables 2, 3 present several key performance indicators for the models, including sensitivity, specificity, Brier score, F-beta coefficient, PPV, NPV, Log Loss, and MCC, indicating that most models demonstrated strong performance. The tree-based ensemble models (RF, XGB, and LGBM) achieved perfect training metrics; however, they exhibited significant declines in

validation sensitivity and PPV, dropping to 0.59–0.62. LR achieved the highest external test AUC (0.91) with minimal sensitivity loss (0.80 compared to 0.81 in the external test set). Prediction uncertainty increased across all models during validation, with RF demonstrating the least reduction in performance.

Additionally, box plots provide an intuitive visualization of AUC, recall, ACC, and Cross-Entropy (CE) in relation to classification performance evaluation (Figure 3; Supplementary Figure S2). The confusion matrix for the four ML models is presented in Supplementary Figure S3, along with the DCA curves in Supplementary Figure S4. To investigate statistically significant differences in AUC values among the four ML models, we conducted the DeLong test (Supplementary Figure S5), which revealed no significant difference in AUC values between LR and the other ML models in external test set. Interpretation of the LR model was performed to reduce ML model complexity (Figure 4).

Nomogram model construction and validation

Moreover, to provide clinicians with a more effective visual assessment tool, a nomogram model incorporating the seven features was developed based on LR (Figure 5). Baseline data for the features are detailed in Supplementary Table S2. Calibration plots indicated strong alignment between expected and actual outcomes in both the training and validation sets, as evidenced by low Brier scores (0.056 for the training set and 0.087 for the validation set) (Figures 6B, E). To assess the clinical efficacy of the nomogram model, we conducted a direct comparison to evaluate its predictive capacity against the preoperative HGB levels of patients. The AUC values for the nomogram model were found to exceed those of the preoperative HGB levels in the training set (0.95, 95% CI = 0.93–0.97) and in the validation set (0.90, 95% CI = 0.84–0.95), compared to training set values of 0.80 (95% CI = 0.74–0.86) and 0.76 (95% CI = 0.66–0.85) in the validation set (Figures 6A, D). This demonstrates the superior discriminative capacity of the nomogram model. DCA curves further confirmed that the nomogram model offers greater net benefits in both cohorts (Figures 6C, F). Importantly, the nomogram model exhibited enhanced diagnostic performance compared to HGB levels, as indicated by the results of AUC and DCA analyses.

Discussion

A prevalent complication associated with THA is considerable blood loss, frequently necessitating allogeneic blood transfusions. Such transfusions have been correlated with adverse outcomes and heightened healthcare costs (21), prompting clinicians to adopt strategies aimed at minimizing their necessity (22, 23). Predicting the need for perioperative blood transfusions can assist in identifying high-risk patients, improving patient care, and yielding better outcomes, lower morbidity rates, and cost savings. Consequently, anticipating transfusion needs and implementing appropriate preoperative interventions is essential.

Traditionally, physicians have relied on HGB levels to guide blood transfusion decisions (24, 25), however, significant

TABLE 1 Baseline characteristics of patients in the study.

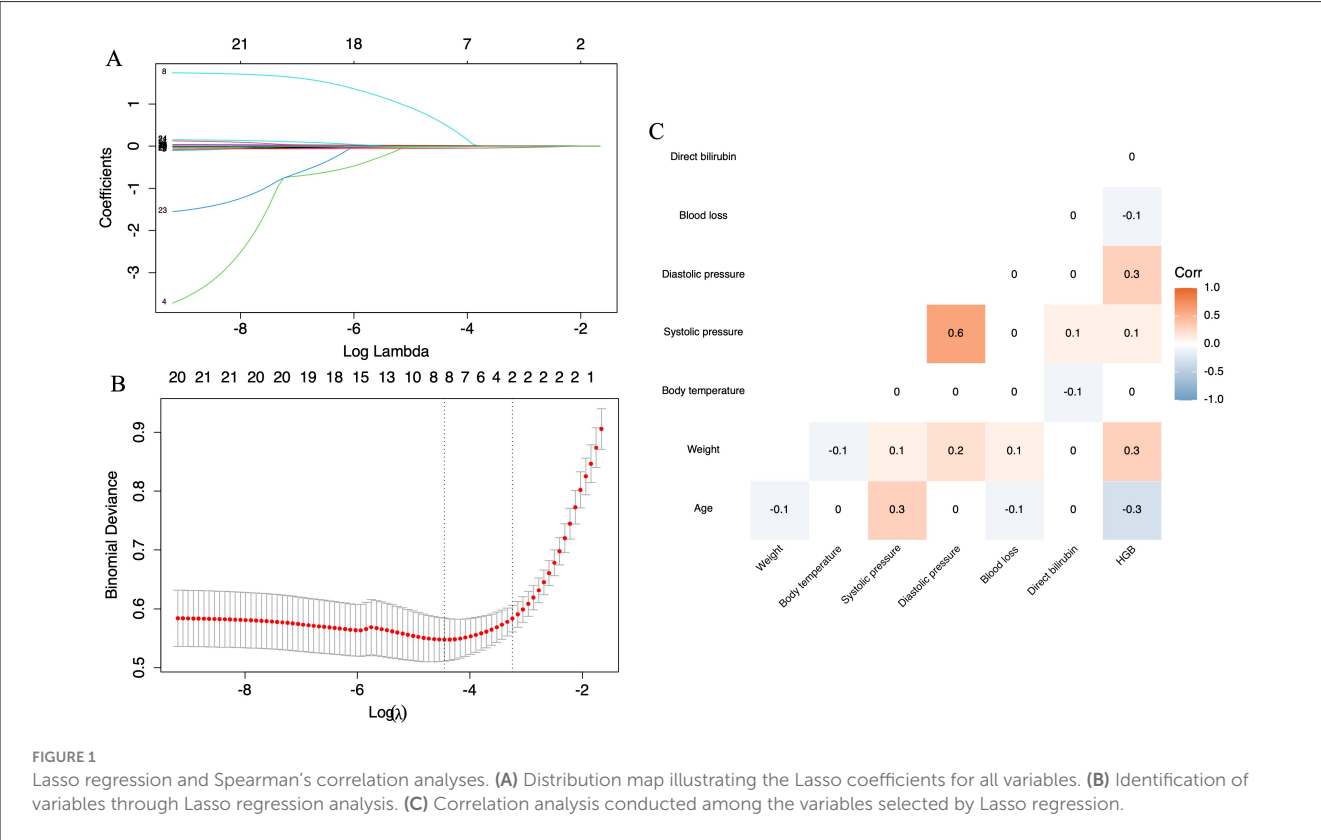
Characteristic	Overall, N = 503	No transfusion, N = 418	Transfusion, N = 85	p-value
Gender, n (%)				0.003
Male	276 (55)	242 (58)	34 (40)	
Female	227 (45)	176 (42)	51 (60)	
Age, Mean (SD)	56.58 (13.34)	56.63 (12.86)	56.33 (15.60)	0.81
Weight, Mean (SD)	63.03 (11.25)	63.57 (10.78)	60.36 (13.06)	0.006
Height, Mean (SD)	1.60 (0.12)	1.60 (0.12)	1.59 (0.12)	0.49
BMI, Mean (SD)	25.09 (5.73)	25.30 (5.83)	24.06 (5.13)	0.11
Hypertension, n (%)				0.38
No	384 (76)	316 (76)	68 (80)	
Yes	119 (24)	102 (24)	17 (20)	
Diabetes, n (%)				0.79
No	465 (92)	387 (93)	78 (92)	
Yes	38 (7.6)	31 (7.4)	7 (8.2)	
Body temperature, Mean (SD)	36.47 (0.19)	36.47 (0.19)	36.50 (0.20)	0.14
Respiratory rate, Mean (SD)	19.67 (1.12)	19.66 (1.03)	19.74 (1.51)	0.80
Pulse, Mean (SD)	81.18 (10.47)	81.32 (10.46)	80.48 (10.56)	0.60
Systolic pressure, Mean (SD)	131.10 (18.65)	132.03 (18.45)	126.56 (19.06)	0.020
Diastolic pressure, Mean (SD)	82.63 (10.91)	83.28 (10.91)	79.40 (10.40)	0.005
Operative site, n (%)				<0.001
Left	242 (48)	211 (50)	31 (36)	
Right	244 (49)	199 (48)	45 (53)	
Both	17 (3.4)	8 (1.9)	9 (11)	
Surgery routes, n (%)				0.60
Anterior	31 (6.2)	27 (6.5)	4 (4.7)	
Posterior	304 (60)	255 (61)	49 (58)	
External	168 (33)	136 (33)	32 (38)	
Incisional length, n (%)				0.060
<12	236 (47)	204 (49)	32 (38)	
≥12	267 (53)	214 (51)	53 (62)	
Blood loss, Mean (SD)	324.65 (292.19)	257.89 (132.38)	652.94 (540.36)	<0.001
Hemostatic drugs, n (%)				0.28
None	68 (14)	54 (13)	14 (16)	
Tranexamic acid	294 (58)	240 (57)	54 (64)	
Aminocaproic acid	134 (27)	118 (28)	16 (19)	
Aminomethylbenzoic acid	7 (1.4)	6 (1.4)	1 (1.2)	
ALT, Mean (SD)	21.94 (19.93)	22.20 (15.78)	20.64 (33.71)	<0.001
AST, Mean (SD)	22.95 (23.46)	23.21 (24.38)	21.70 (18.31)	0.007
Albumin, Mean (SD)	42.90 (4.85)	43.23 (4.77)	41.29 (4.90)	<0.001
Direct bilirubin, Mean (SD)	4.69 (10.38)	4.28 (1.93)	6.70 (24.90)	0.47
Indirect bilirubin, Mean (SD)	5.93 (3.04)	6.08 (3.07)	5.18 (2.73)	0.007
NIR, Mean (SD)	0.95 (0.07)	0.95 (0.07)	0.98 (0.09)	0.002

(Continued)

TABLE 1 (Continued)

Characteristic	Overall, N = 503	No transfusion, N = 418	Transfusion, N = 85	p-value
PT, Mean (SD)	11.83 (1.18)	11.77 (1.16)	12.17 (1.27)	0.008
APTT, Mean (SD)	28.41 (3.52)	28.35 (3.28)	28.70 (4.53)	0.19
FIB, Mean (SD)	3.64 (1.34)	3.62 (1.34)	3.74 (1.36)	0.58
HGB, Mean (SD)	133.87 (19.69)	137.26 (17.87)	117.19 (19.84)	<0.001
PLT, Mean (SD)	262.79 (78.09)	259.46 (71.72)	279.15 (102.88)	0.087

ALT, Alanine Aminotransferase; AST, Aspartate Aminotransferase; NIR, International Normalized Ratio; PT, Prothrombin Time; APTT, Activated Partial Thromboplastin Time; FIB, Fibrinogen; HGB, Hemoglobin; PLT, Platelet Count.



discrepancies in the transfusion thresholds have been noted (26, 27). In this study, we identified seven key features—blood loss volume, HGB levels, body weight, body temperature, systolic pressure, and direct bilirubin—as predictive factors for the necessity of blood transfusions during the perioperative period.

The associations between the characteristic variables examined in this study and perioperative transfusions have been documented extensively in the literature. Previous studies have demonstrated a positive correlation between blood loss and the likelihood of requiring a blood transfusion (28, 29). Our findings align with these observations. Intraoperative blood loss may elevate the demand for blood transfusions, potentially due to a complex physiological cascade that includes reductions in blood volume, oxygenation, coagulation function, metabolic acidosis, impaired immune response, and other contributing factors.

HGB levels play a critical role in determining the necessity of perioperative blood transfusions (30, 31). One study revealed a

five-fold increase in the likelihood of requiring a blood transfusion when preoperative HGB levels fall below 120 g/L (32), and another study arrived at a similar conclusion (33). This phenomenon can be attributed to the diminished ability of patients with lower preoperative HGB levels to effectively compensate for intraoperative blood loss (34).

Our results indicate that individuals with lower body weight are more likely to require a blood transfusion than those with higher body weight. Several studies have corroborated these findings (35, 36). Patients with lower body weight possess a reduced RBC volume (37, 38), which may hinder their ability to compensate for blood loss. Consequently, lighter individuals may find it more challenging to recover from significant blood loss compared to those with higher body weight (37). Additionally, studies have indicated that factors such as age, systolic pressure, and direct bilirubin significantly influence perioperative blood transfusions (9). Maintaining normal body temperature is crucial,

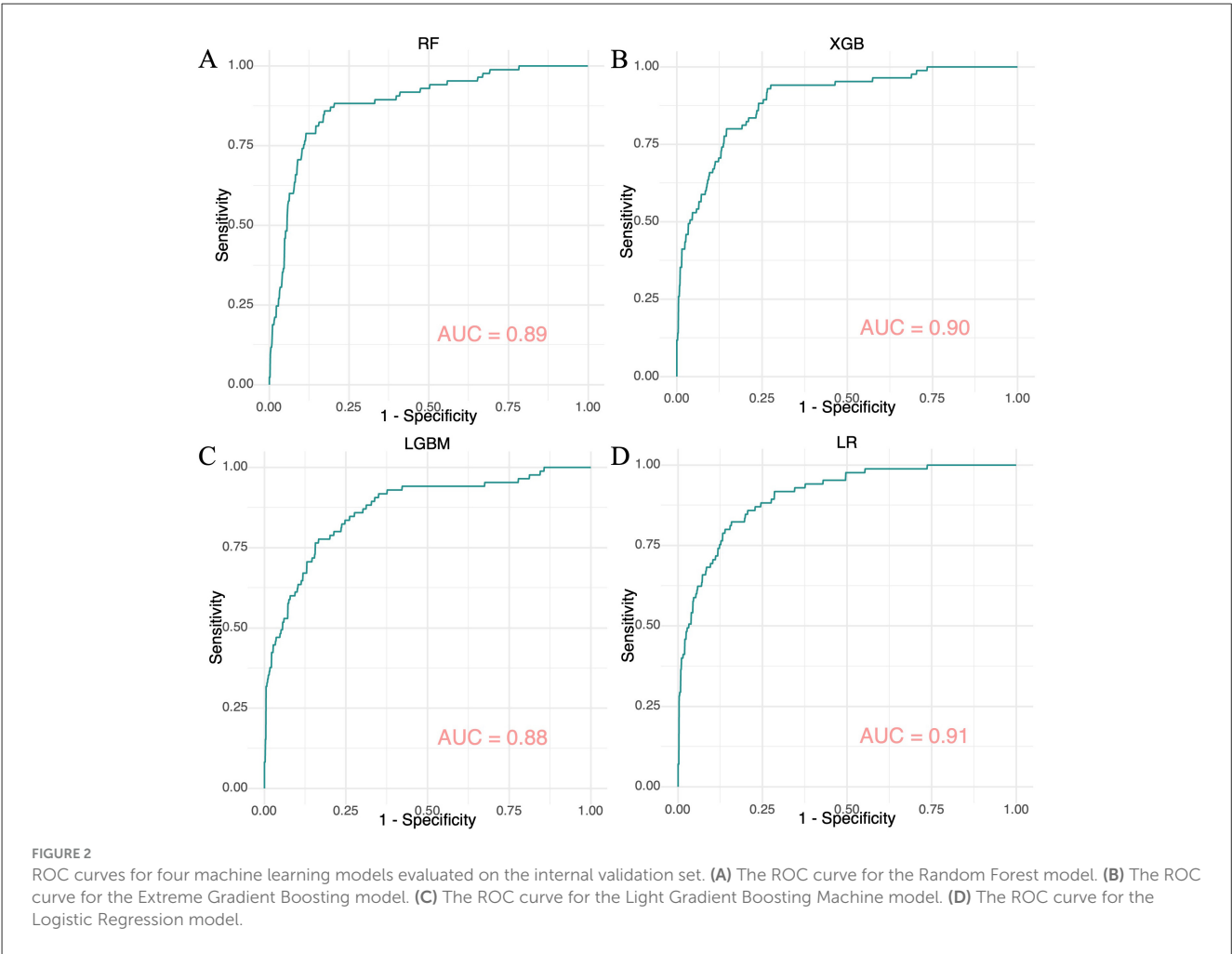


TABLE 2 Performance evaluation of machine learning models within the internal validation set.

Learner	AUC	Sensitivity	Specificity	Brier	F-beta	PPV	NPV	Log loss	MCC
RF	1.00	1.00	0.99	0.05	0.98	0.97	1.00	0.14	0.98
Xgb	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.01	1.00
Lgb	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.00	1.00
LR	0.93	0.81	0.85	0.23	0.64	0.53	0.96	0.38	0.56

as hypothermia can exacerbate blood loss and increase the need for transfusions. Research on THA demonstrates that stable body temperatures can mitigate blood loss and reduce transfusion rates (39).

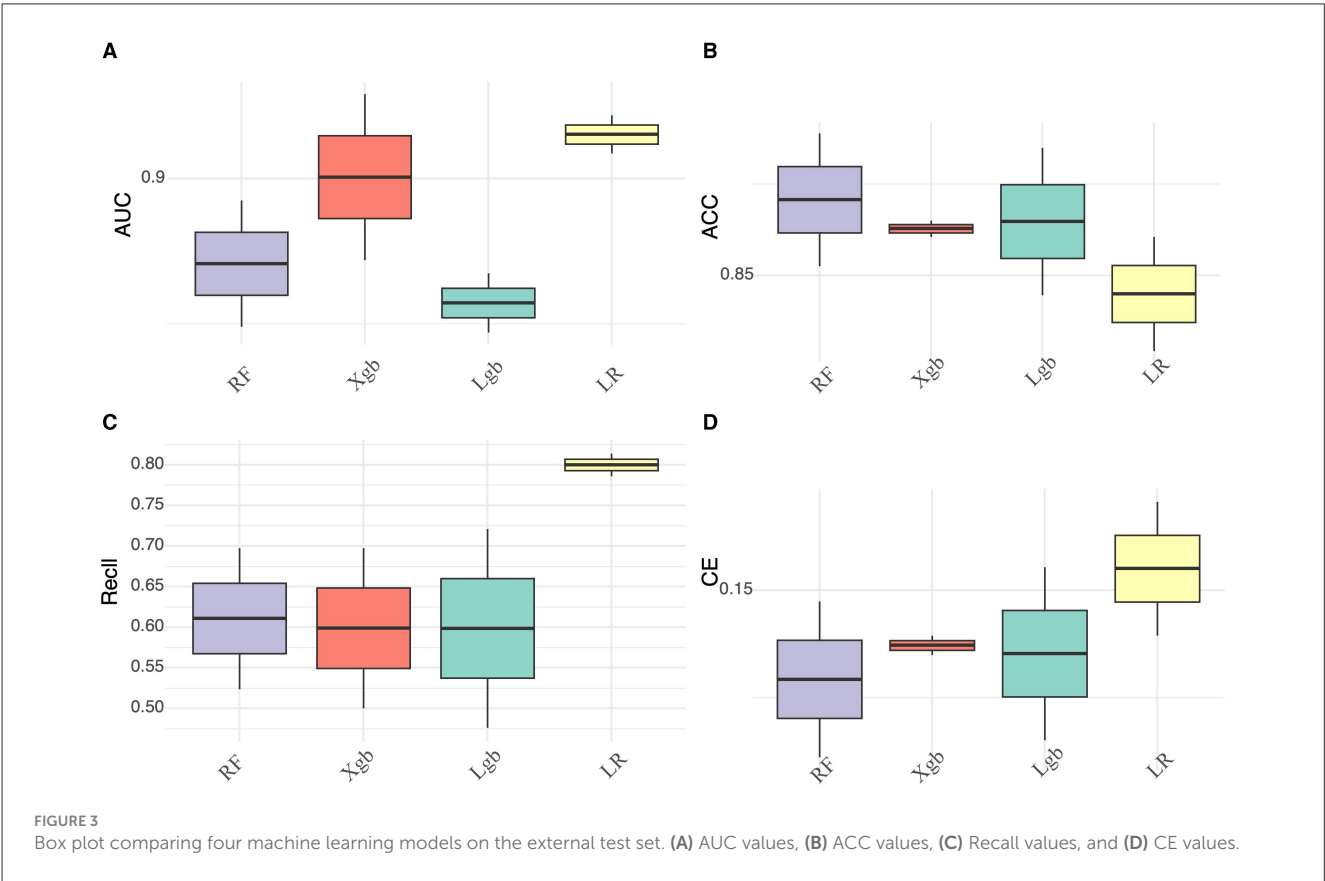
ML has capabilities for effectively managing non-linear and complex data structures and addressing issues associated with high-dimensional data and missing values. Concurrently, it exhibits a robust capacity for pattern recognition and utilization. Various studies have confirmed the utility of ML in clinical diagnosis and prognosis (40, 41). Recently, there has been growing interest in applying ML to the field of blood transfusion (42, 43). Currently, no models exist to predict the likelihood of perioperative blood transfusions in patients with ONFH undergoing THA. While some predictive models perform well, they face clinical application

challenges due to suboptimal performance (44). Moreover, there is a notable gap in evaluating predictive performance specific to ONFH cases.

In this study, we developed four ML models through hyperparameter tuning for perioperative blood transfusions in patients with ONFH undergoing THA. RF mitigates the risk of overfitting by integrating multiple decision trees and demonstrates exceptional robustness in the presence of high-dimensional data and noise (45). Furthermore, it facilitates the assessment of feature importance (46). XGB employs second-order Taylor expansion and incorporates regularization terms, resulting in enhanced prediction accuracy and robust generalization (47). The LGBM utilizes histogram algorithms and prioritizes the growth of leaf nodes, achieving remarkably fast training speeds and low memory

TABLE 3 Performance evaluation of machine learning models within the external test set.

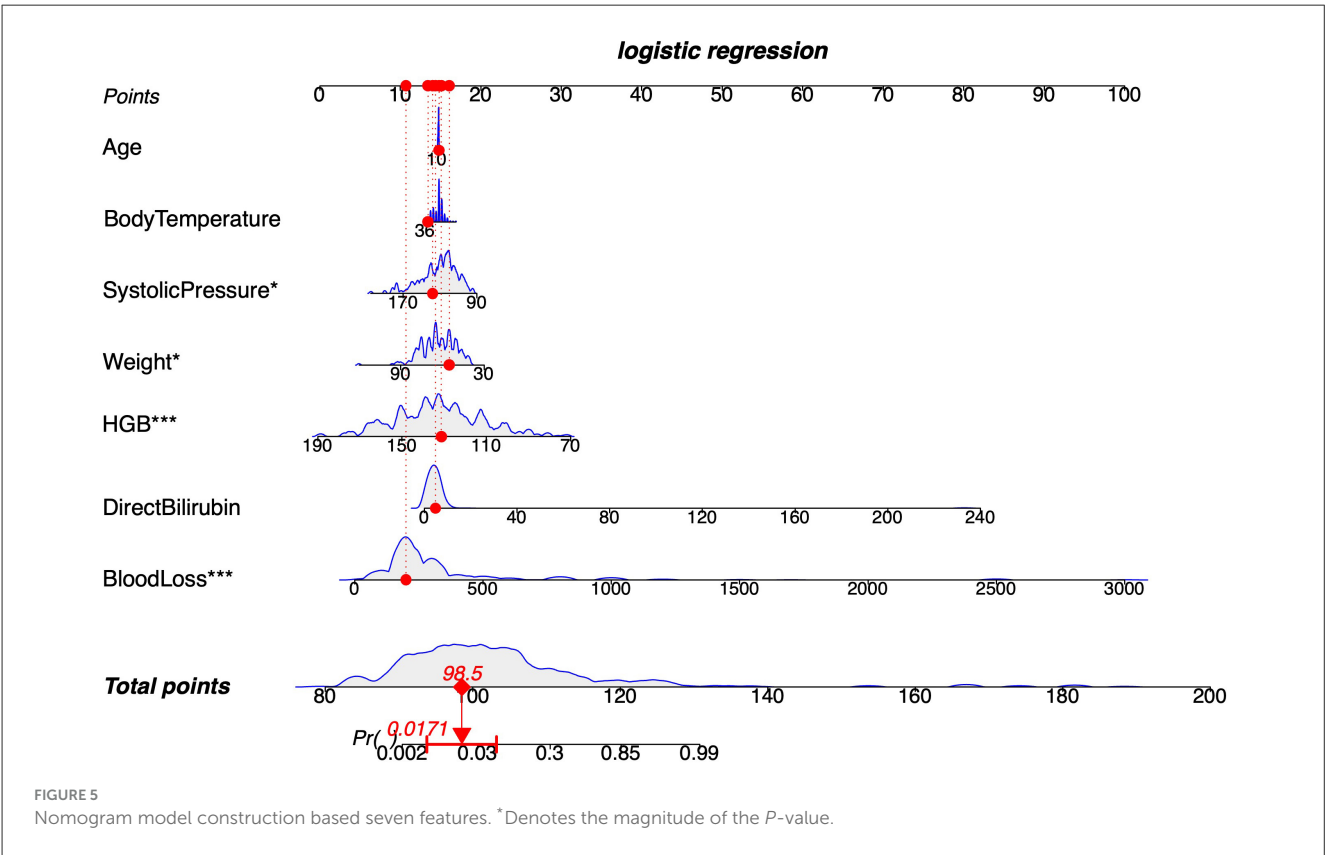
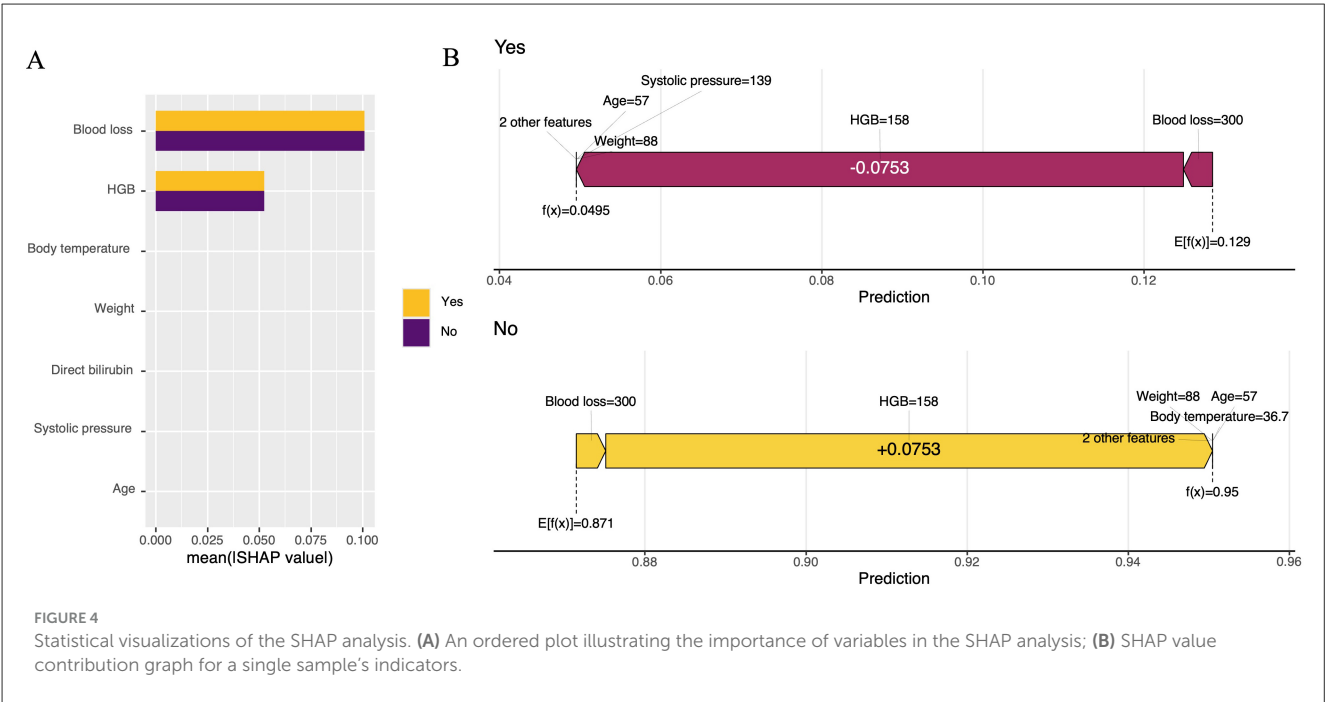
Learner	AUC	Sensitivity	Specificity	Brier	F-beta	PPV	NPV	Log loss	MCC
RF	0.89	0.61	0.92	0.19	0.61	0.62	0.92	0.33	0.54
Xgb	0.90	0.60	0.92	0.21	0.59	0.59	0.92	0.39	0.51
Lgb	0.88	0.60	0.92	0.24	0.59	0.59	0.92	0.68	0.51
LR	0.91	0.80	0.85	0.25	0.64	0.53	0.95	0.40	0.56



consumption, which renders it suitable for ultra-large-scale datasets (47). LR offers strong interpretability of probability outputs, allows for the adjustment of decision thresholds, and is characterized by its simplicity, computational efficiency, and speed, making it suitable for large-scale structured data (48). LR was deemed the most clinically suitable due to its balanced performance across various metrics. The RF, XGB, and LGBM models exhibited excellent results but had low Brier scores and Log Loss, indicating overfitting to the internal validation set. Their performance significantly declined on the external test set, particularly regarding sensitivity and PPV, highlighting their limited generalizability. Additionally, the AUC values of the four models did not display any statistical differences in DeLong test conducted on the external test set. LR demonstrated clinical validity through four primary features in our study: first, it possessed higher validation sensitivity, ensuring better identification of true transfusion candidates, which is critical in surgical settings. Second, it balanced specificity and NPV, maintaining strong negative classification without missing positive cases. Third, it exhibited stable performance from internal

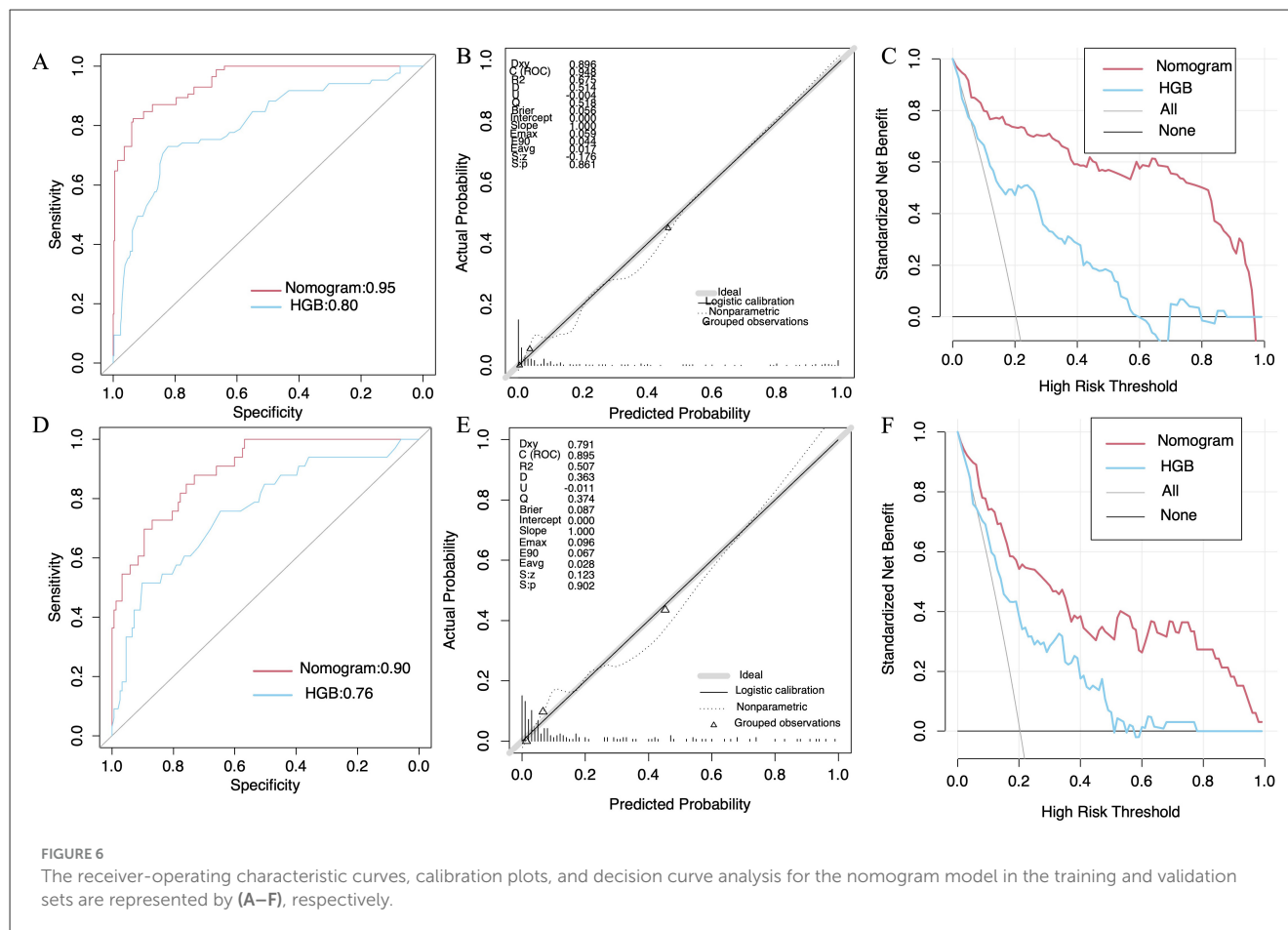
validation to external test, showing the smallest AUC decrease, indicating reliable generalization. Finally, the DCA illustrated that the LR model yields a significant net benefit and maintains consistency between the internal validation and external test sets. Ensemble methods displayed slightly better calibration and Log Loss; however, their reduced sensitivity renders them clinically less viable. The higher MCC of LR indicates its superior classification accuracy in clinical settings.

The SHapley Additive exPlanations (SHAP) is a method for interpreting ML models that is grounded in the Shapley value from game theory (49). By quantifying the marginal contribution of each feature to the model's predictions, SHAP offers both global and local interpretability (50). The SHAP analysis identified critical features: HGB levels (positive correlation), body weight (negative correlation), and age (strong global influence) (9). High HGB levels suggest compensatory hemoconcentration during blood loss, while greater body weight indicates a better physiological reserve (37). Advanced age is strongly correlated with risk due to diminished hemodynamic



adaptability. Blood loss and hypothermia exhibited synergistic effects, correlating with clinical observations of worsened coagulopathy HGB (29). These patterns confirm the model's alignment with established pathophysiology, underscoring its clinical relevance.

We also constructed a nomogram based on LR to predict transfusion risk. The nomogram demonstrated robust discriminatory ability, with consistent AUC values in both training and validation sets. Direct comparisons revealed that the predictive performance of this nomogram significantly



outperformed HGB levels. Furthermore, the DeLong test indicated a statistically significant difference in AUC values between the nomogram and both LR and HGB levels. This finding demonstrates that the predictive capabilities of LR and the nomogram exceed those of HGB levels. Moreover, the DCA curves indicate that the nomogram model provides a substantially higher net benefit compared to HGB levels. This visual tool supports clinicians in risk assessment, allowing for improved perioperative blood management and surgical planning. By identifying high-risk patients, it may help reduce transfusion-related complications through targeted interventions. The nomogram's combination of clarity and accuracy makes it an invaluable tool for enhancing perioperative care in neurovascular surgeries.

This model aids in identifying patients requiring perioperative blood transfusions and optimizing blood management strategies, such as autologous blood dilution, intraoperative hypotension management, and postoperative recovery. For successful integration into clinical practice, it is vital to clearly define the model's objectives, emphasizing the enhancement of predictive accuracy to prevent unnecessary transfusions. The clinical implementation plan consists of four key components: (1) Electronic Medical Record (EMR) integration via HL7 interfaces to automatically extract data from laboratories and operative notes; (2) Real-time prediction modules capable of issuing prioritized alerts during surgical scheduling based on risk; (3)

Multidisciplinary protocols to monitor transfusion deviations and response times at three pilot sites; and (4) Oversight by a blood management committee that recalibrates models biannually using federated learning across institutions. This framework ensures clinical utility, regulatory compliance, and auditability through performance dashboards.

This study has certain limitations, including single-center data sourcing and the potential for selection bias. The retrospective single-center design inherently restricts the external validation of these findings. Moreover, retrospective studies inevitably involve missing data. Although rigorous internal controls were employed, the absence of multi-institutional data may limit the generalizability of results to broader populations. Furthermore, the retrospective nature of this study necessitates careful consideration of potential confounding factors. But factors such as prior surgical interventions, comorbidities beyond hypertension and diabetes, and smoking were not considered in the present study. We plan to incorporate these variables into our model in future research. Additionally, the retrospective design prevents precise estimations of clinical decision thresholds; as such analyses require outcome data from actual implementation scenarios. Future work will encompass cost-benefit analysis with stakeholder engagement to operationalize these thresholds. Research directions include multicenter validation, prospective studies, exploration of new data sources and features, and continuous iteration and optimization of the model to enhance its generalizability, accuracy, and clinical

applicability. Furthermore, our models have only undergone internal validation and require external validation to confirm their effectiveness.

In conclusion, this study developed four ML models and a nomogram model to effectively predict the likelihood of blood transfusion in patients with ONFH undergoing THA. The model's capability to identify patients with a low probability of requiring a transfusion could diminish unnecessary repeat testing, such as complete blood counts and additional preoperative laboratory tests. Additionally, it can assist clinicians in implementing strategies to reduce bleeding and prepare for transfusions in high-risk patients. Consequently, this model could serve as a valuable tool for clinicians in preoperative preparation and in reducing unnecessary medical procedures.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Ethics Committee of the Second Affiliated Hospital of Fujian Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. The Ethics Committee/Institutional Review Board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because this study was a retrospective study using the medical records of the patients and had no direct impact on the patients.

Author contributions

Z-DS: Methodology, Writing – original draft. Y-MF: Writing – original draft. Y-LL: Formal analysis, Visualization, Writing –

original draft. M-QP: Validation, Visualization, Writing – original draft. C-YL: Investigation, Supervision, Writing – original draft. H-fH: Project administration, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by grants from the Innovation of Science and Technology, Fujian Province (2023Y9233) and the Medical Innovation Science and Technology Project of Fujian Province (2024CXB006).

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.

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.2025.1471746/full#supplementary-material>

References

- Shah KN, Racine J, Jones LC, Aaron RK. Pathophysiology and risk factors for osteonecrosis. *Curr Rev Musculoskelet Med.* (2015) 8:201–9. doi: 10.1007/s12178-015-9277-8
- Mont MA, Zywił MG, Marker DR, McGrath MS, Delanois RE. The natural history of untreated asymptomatic osteonecrosis of the femoral head: a systematic literature review. *J Bone Joint Surg Am.* (2010) 92:2165–70. doi: 10.2106/JBJS.I.00575
- Lieberman JR, Engstrom SM, Meneghini RM, SooHoo NF. Which factors influence preservation of the osteonecrotic femoral head? *Clin Orthop Relat Res.* (2012) 470:525–34. doi: 10.1007/s11999-011-2050-4
- Mont MA, Salem HS, Piuze NS, Goodman SB, Jones LC. Nontraumatic osteonecrosis of the femoral head: where do we stand today?: a 5-year update. *J Bone Joint Surg Am.* (2020) 102:1084–99. doi: 10.2106/JBJS.19.01271
- Menendez ME, Lu N, Huybrechts KF, Ring D, Barnes CL, Ladha K, et al. Variation in use of blood transfusion in primary total hip and knee arthroplasties. *J Arthroplasty.* (2016) 31:2757–63.e2. doi: 10.1016/j.arth.2016.05.022
- Saleh A, Small T, Chandran Pillai AL, Schiltz NK, Klika AK, Barsoum WK. Allogenic blood transfusion following total hip arthroplasty: results from the nationwide inpatient sample, 2000 to 2009. *J Bone Joint Surg Am.* (2014) 96:e155. doi: 10.2106/JBJS.M.00825
- Everhart JS, Sojka JH, Mayerson JL, Glassman AH, Scharschmidt TJ. Perioperative allogeneic red blood-cell transfusion associated with surgical site infection after total hip and knee arthroplasty. *J Bone Joint Surg Am.* (2018) 100:288–94. doi: 10.2106/JBJS.17.00237
- Friedman R, Homering M, Holberg G, Berkowitz SD. Allogeneic blood transfusions and postoperative infections after total hip or knee arthroplasty. *J Bone Joint Surg Am.* (2014) 96:272–8. doi: 10.2106/JBJS.L.01268
- Hart A, Khalil JA, Carli A, Huk O, Zukor D, Antoniou J. Blood transfusion in primary total hip and knee arthroplasty: Incidence, risk factors, and thirty-day complication rates. *J Bone Joint Surg Am.* (2014) 96:1945–51. doi: 10.2106/JBJS.N.00077
- Xu S, Chen JY, Zheng Q, Lo NN, Chia SL, Tay KJD, et al. The safest and most efficacious route of tranexamic acid administration in total joint arthroplasty: a systematic review and network meta-analysis. *Thromb Res.* (2019) 176:61–6. doi: 10.1016/j.thromres.2019.02.006

11. Yalamanchili S, Talei B, Azizzadeh B, Auersvald A, Frankel AS. Wound healing complications with tranexamic acid: not the silver bullet after all. *Aesthet Surg J*. (2023) 43:1409–15. doi: 10.1093/asj/sjad177
12. Ricchetti ET, Abboud JA, Kuntz AF, Ramsey ML, Glaser DL, Williams GR, et al. Total shoulder arthroplasty in older patients: increased perioperative morbidity? *Clin Orthop Relat Res*. (2011) 469:1042–9. doi: 10.1007/s11999-010-1582-3
13. Burchard R, Daginnus A, Soost C, Schmitt J, Graw JA. Gender differences in blood transfusion strategy for patients with hip fractures - a retrospective analysis. *Int J Med Sci*. (2020) 17:620–5. doi: 10.7150/ijms.33954
14. Johnson PM, Lin DJ, Zbontar J, Zitnick CL, Sriram A, Muckley M, et al. Deep learning reconstruction enables prospectively accelerated clinical knee MRI. *Radiology*. (2023) 307:e220425. doi: 10.1148/radiol.220425
15. Aromolaran O, Aromolaran D, Isewon I, Oyelade J. Machine learning approach to gene essentiality prediction: a review. *Brief Bioinform*. (2021) 22. doi: 10.1093/bib/bbab128
16. Garriga R, Mas J, Abraha S, Nolan J, Harrison O, Tadros G, et al. Machine learning model to predict mental health crises from electronic health records. *Nat Med*. (2022) 28:1240–8. doi: 10.1038/s41591-022-01811-5
17. Zapf MAC, Fabbri DV, Andrews J, Li G, Freundlich RE, Al-Droubi S, et al. Development of a machine learning model to predict intraoperative transfusion and guide type and screen ordering. *J Clin Anesth*. (2023) 91:111272. doi: 10.1016/j.jclineane.2023.111272
18. Lou SS, Liu H, Lu C, Wildes TS, Hall BL, Kannampallil T. Personalized surgical transfusion risk prediction using machine learning to guide preoperative type and screen orders. *Anesthesiology*. (2022) 137:55–66. doi: 10.1097/ALN.0000000000004139
19. Wynants L, Bouwmeester W, Moons KG, Moerbeek M, Timmerman D, Van Huffel S, et al. A simulation study of sample size demonstrated the importance of the number of events per variable to develop prediction models in clustered data. *J Clin Epidemiol*. (2015) 68:1406–14. doi: 10.1016/j.jclinepi.2015.02.002
20. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol*. (2007) 165:710–8. doi: 10.1093/aje/kwk052
21. Sassoon A, D'Apuzzo M, Sems S, Cass J, Mabry T. Total hip arthroplasty for femoral neck fracture: comparing in-hospital mortality, complications, and disposition to an elective patient population. *J Arthroplasty*. (2013) 28:1659–62. doi: 10.1016/j.arth.2013.01.027
22. Niemi TT, Pitkänen M, Syrjäla M, Rosenberg PH. Comparison of hypotensive epidural anaesthesia and spinal anaesthesia on blood loss and coagulation during and after total hip arthroplasty. *Acta Anaesthesiol Scand*. (2000) 44:457–64. doi: 10.1034/j.1399-6576.2000.440418.x
23. Lenet T, McIsaac DI, Hallett JH, Jerath A, Lalu MM, Nicholls SG, et al. Intraoperative blood management strategies for patients undergoing noncardiac surgery: the Ottawa intraoperative transfusion consensus. *JAMA Netw Open*. (2023) 6:e2349559. doi: 10.1001/jamanetworkopen.2023.49559
24. Roubinian NH, Murphy EL, Swain BE, Gardner MN, Liu V, Escobar GJ, et al. Predicting red blood cell transfusion in hospitalized patients: role of hemoglobin level, comorbidities, and illness severity. *BMC Health Serv Res*. (2014) 14:213. doi: 10.1186/1472-6963-14-213
25. Kim Y, Bagante F, Gani F, Ejaz A, Xu L, Wasey JO, et al. Nomogram to predict perioperative blood transfusion for hepatopancreaticobiliary and colorectal surgery. *Br J Surg*. (2016) 103:1173–83. doi: 10.1002/bjs.10164
26. Schneider C, Boddy AP, Fukuta J, Groom WD, Streets CG. Predicting blood transfusion in patients undergoing minimally invasive oesophagectomy. *Int J Surg*. (2014) 12:1342–7. doi: 10.1016/j.ijsu.2014.10.016
27. Janny S, Eurin M, Dokmak S, Toussaint A, Farges O, Paugam-Burtz C. Assessment of the external validity of a predictive score for blood transfusion in liver surgery. *HPB*. (2015) 17:357–61. doi: 10.1111/hpb.12376
28. Bian T, Zhang L, Man S, Li H, Dou Y, Zhou Y. Predisposing factors for allogeneic blood transfusion in patients with ankylosing spondylitis undergoing primary unilateral total hip arthroplasty: a retrospective study. *J Orthop Surg Res*. (2023) 18:9. doi: 10.1186/s13018-022-03464-z
29. Wang Q, Huang Q, Pei F. Incidence and risk factors for blood transfusion in simultaneous bilateral total hip arthroplasty. *Jt Dis Relat Surg*. (2021) 32:590–7. doi: 10.52312/jdrs.2021.333
30. Carling MS, Jeppsson A, Eriksson BI, Brisby H. Transfusions and blood loss in total hip and knee arthroplasty: a prospective observational study. *J Orthop Surg Res*. (2015) 10:48. doi: 10.1186/s13018-015-0188-6
31. Dillon MF, Collins D, Rice J, Murphy PG, Nicholson P, Mac Elwaine J. Preoperative characteristics identify patients with hip fractures at risk of transfusion. *Clin Orthop Relat Res*. (2005) 439:201–6. doi: 10.1097/01.blo.0000173253.59827.7b
32. Adunsky A, Lichtenstein A, Mizrahi E, Arad M, Heim M. Blood transfusion requirements in elderly hip fracture patients. *Arch Gerontol Geriatr*. (2003) 36:75–81. doi: 10.1016/S0167-4943(02)00059-6
33. Bian FC, Cheng XK, An YS. Preoperative risk factors for postoperative blood transfusion after hip fracture surgery: establishment of a nomogram. *J Orthop Surg Res*. (2021) 16:406. doi: 10.1186/s13018-021-02557-5
34. Liu B, Pan J, Zong H, Wang Z. Establishment and verification of a perioperative blood transfusion model after posterior lumbar interbody fusion: a retrospective study based on data from a local hospital. *Front Surg*. (2021) 8:695274. doi: 10.3389/fsurg.2021.695274
35. Yoshihara H, Yoneoka D. Predictors of allogeneic blood transfusion in total hip and knee arthroplasty in the United States, 2000–2009. *J Arthroplasty*. (2014) 29:1736–40. doi: 10.1016/j.arth.2014.04.026
36. Barr PJ, Donnelly M, Cardwell C, Alam SS, Morris K, Parker M, et al. Drivers of transfusion decision making and quality of the evidence in orthopedic surgery: a systematic review of the literature. *Transfus Med Rev*. (2011) 25:304–16. doi: 10.1016/j.tmr.2011.04.003
37. Khanna MP, Hébert PC, Fergusson DA. Review of the clinical practice literature on patient characteristics associated with perioperative allogeneic red blood cell transfusion. *Transfus Med Rev*. (2003) 17:110–9. doi: 10.1053/tmr.2003.50008
38. Alghamdi AA, Davis A, Brister S, Corey P, Logan A. Development and validation of Transfusion Risk Understanding Scoring Tool (TRUST) to stratify cardiac surgery patients according to their blood transfusion needs. *Transfusion*. (2006) 46:1120–9. doi: 10.1111/j.1537-2995.2006.00860.x
39. Pan P, Song K, Yao Y, Jiang T, Jiang Q. The impact of intraoperative hypothermia on blood loss and allogeneic blood transfusion in total knee and hip arthroplasty: a retrospective study. *Biomed Res Int*. (2020) 2020:1096743. doi: 10.1155/2020/1096743
40. Zhou CM, Wang Y, Yang JJ, Zhu Y. Predicting postoperative gastric cancer prognosis based on inflammatory factors and machine learning technology. *BMC Med Inform Decis Mak*. (2023) 23:53. doi: 10.1186/s12911-023-02150-2
41. Foustier G, Rodrigues EM, Giamporcaro GM, Falcone M. A machine learning approach to predict response to immunotherapy in type 1 diabetes. *Cell Mol Immunol*. (2021) 18:515–7. doi: 10.1038/s41423-020-00594-4
42. Lang FF, Liu LY, Wang SW. Predictive modeling of perioperative blood transfusion in lumbar posterior interbody fusion using machine learning. *Front Physiol*. (2023) 14:1306453. doi: 10.3389/fphys.2023.1306453
43. Cavazos DR, Sayeed Z, Court T, Chen C, Little BE, Darwiche HF. Predicting factors for blood transfusion in primary total knee arthroplasty using a machine learning method. *J Am Acad Orthop Surg*. (2023) 31:e845–58. doi: 10.5435/JAAOS-D-23-00063
44. Zang H, Hu A, Xu X, Ren H, Xu L. Development of machine learning models to predict perioperative blood transfusion in hip surgery. *BMC Med Inform Decis Mak*. (2024) 24:158. doi: 10.1186/s12911-024-02555-7
45. O'Brien R, Ishwaran H. A random forests quantile classifier for class imbalanced data. *Pattern Recognit*. (2019) 90:232–49. doi: 10.1016/j.patcog.2019.01.036
46. Touw WG, Bayjanov JR, Overmars L, Backus L, Boekhorst J, Wels M, et al. Data mining in the life sciences with random forest: a walk in the park or lost in the jungle? *Brief Bioinform*. (2013) 14:315–26. doi: 10.1093/bib/bbs034
47. Zhao Z, Xu Y, Zhao Y. SXGBsite: prediction of protein-ligand binding sites using sequence information and extreme gradient boosting. *Genes*. (2019) 10:965. doi: 10.3390/genes10120965
48. Kavelaars X, Mulder J, Kaptein M. Bayesian multilevel multivariate logistic regression for superiority decision-making under observable treatment heterogeneity. *BMC Med Res Methodol*. (2023) 23:220. doi: 10.1186/s12874-023-02034-z
49. Tempel F, Ihlen EAF, Adde L, Strümke I. Explaining human activity recognition with SHAP: validating insights with perturbation and quantitative measures. *Comput Biol Med*. (2025) 188:109838. doi: 10.1016/j.combiomed.2025.109838
50. Chandra SS, Kumar R, Arjunasamy A, Galagali S, Tantri A, Naganna SR. Predicting the compressive strength of polymer-infused bricks: a machine learning approach with SHAP interpretability. *Sci Rep*. (2025) 15:8090. doi: 10.1038/s41598-025-89606-9