Journal of Inflammation Research

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

# Machine Learning Strategies for Preoperative PJI Diagnosis: Integrative Analysis of Serum and Synovial Fluid Markers

Bin Chen[1–3,*], Yaji Yang[1–3,*], Haotian Zhou[1–3], Feilong Li[1–3], Yidong Shen[1–3], Qiang Cheng[1–3], Wei Huang[1–3], Leilei Qin[1–3]

[1]Department of Orthopaedic Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, People's Republic of China; [2]Chongqing Municipal Health Commission Key Laboratory of Musculoskeletal Regeneration and Translational Medicine, Chongqing Medical University, Chongqing, People's Republic of China; [3]Orthopaedic Research Laboratory of Chongqing Medical University, Chongqing Medical University, Chongqing, People's Republic of China

*These authors contributed equally to this work

Correspondence: Wei Huang; Leilei Qin, Department of Orthopaedics, The First Affiliated Hospital of Chongqing Medical University, Chongqing, 400016, People's Republic of China, Email huangwei68@263.net; 253505921@qq.com



**Background:** Preoperative diagnosis of periprosthetic joint infection (PJI) is crucial for guiding treatment strategies and improving patient outcomes. This study aims to develop a new diagnostic model for the preoperative diagnosis of PJI based on serum and synovial fluid markers and further validate its effectiveness.

**Methods:** We retrospectively collected data from patients admitted for joint revision surgery between January 2018 and October 2022, selecting serum and synovial fluid markers as variables for the study. The most suitable diagnostic markers were selected using LASSO regression, and eight machine learning (ML) models were constructed based on the selected markers. The diagnostic performance and clinical utility of the ML models were assessed using receiver operating characteristic curves, calibration curves, decision curve analysis, and clinical impact analysis. Finally, the best model was compared to existing diagnostic standards using an external validation cohort.

**Results:** A total of 376 cases were analyzed (263 in the training cohort and 113 in the validation cohort), with 111 cases (29.52%) diagnosed as PJI. The ML models included SE-IL6, SE-CRP, ESR, SF-IL6, PMN%, DD, and ALB. The eXtreme Gradient Boosting model was the optimal model, achieving an area under the curve of 0.998 (95% CI 0.993–1) in the test set, outperforming other models. It also recorded the lowest Brier score of 0.062 and the highest F1 score of 0.985. In the external validation cohort, the accuracy, sensitivity, and specificity of the ML diagnostic model were higher than those of the MSIS 2013 and ICM 2018 diagnostic criteria.

**Conclusion:** Our newly developed ML diagnostic model can assist clinicians in rapidly and accurately diagnosing PJI before surgery and has potential value for timing decisions regarding two-stage revisions. It has high economic value and clinical applicability.

**Keywords:** periprosthetic joint infection, machine learning, diagnostic model, preoperative diagnosis, synovial fluid, inflammatory markers

## Introduction

Periprosthetic joint infection (PJI) is one of the catastrophic complications leading to the failure of joint replacement surgery, with an incidence of approximately 1–3%, and it is showing an upward trend.[1,2] It is well known that early detection of infection is closely linked to PJI prognosis. However, hidden clinical signs, false-negative cultures due to antibiotics or poor sampling make preoperative identification of PJI difficult.[3,4] While improved sampling and specimen handling reduce negative culture rates, pathogen immune evasion can still result in false negatives.[5] Additionally, complex sampling and specimen handling increase the risk of contamination, adding financial strain on patients and

Received: 9 October 2024
Accepted: 19 July 2025
Published: 31 July 2025

treatment burden on doctors. Therefore, establishing rapid and convenient non-microbiological diagnostic strategies is the future goal for PJI diagnosis.

Systemic inflammatory markers like C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR) are important in diagnosing infections, and have been incorporated into the diagnostic criteria for PJI by the Musculoskeletal Infection Society (MSIS) and the International Consensus Meeting (ICM). However, their specificity for PJI diagnosis remains limited, as they are often affected by non-infectious inflammatory conditions.[6] Moreover, in chronic PJI or infections caused by low-virulence organisms, elevations in ESR and CRP are often mild or absent. Our previous studies have also shown that these markers cannot be used as stand-alone indicators and must be interpreted in conjunction with other diagnostic parameters.[4,7–9] In contrast, synovial fluid markers have become increasingly important in clinical diagnosis due to their site-specificity and high diagnostic efficiency.[10] However, aspiration of joint fluid may sometimes fail due to dry tap, making PJI assessment difficult.[11] With the increase in new markers and relevant clinical data, diagnosing PJI has become more complicated, especially in smaller hospitals or clinics, where testing for certain diagnostic markers remains limited by technology and resources. Identifying key, user-friendly diagnostic markers from numerous complex indicators and clinical data, and applying them to PJI diagnosis, remains a major research focus.[12] Although the Musculo-Skeletal Infection Society (MSIS) and International Consensus Meetings (ICM) criteria are widely adopted, their diagnostic accuracy is still limited. Although the MSIS and ICM criteria are widely adopted, their diagnostic accuracy is still limited. For instance, the MSIS 2011 criteria demonstrated a specificity of up to 99.5%, but with a sensitivity of only 79.3%; the sensitivity of the ICM 2013 criteria was similarly limited at 86.9%. While the ICM 2018 criteria have shown improved sensitivity and specificity (both exceeding 95% in some studies), their reliance on intraoperative cultures may lead to delays in treatment and uncertainty in preoperative diagnosis.[13] Although advanced imaging techniques such as MRI have been explored, they remain costly and are prone to metal artifact interference. Even with metal artifact reduction sequences (MARS), the image quality of MRI is often suboptimal.[14] Moreover, MRI is time-consuming, associated with multiple contraindications, and expensive. Similarly, FDG-PET using fluorodeoxyglucose as a tracer shows highly variable results across studies, has a high cost, and carries risks of contrast allergy and other limitations.[15] In addition to serological and synovial fluid inflammatory markers, baseline patient data (such as body mass index [BMI], albumin [ALB], and blood cell count) are also closely related to infection status, but their diagnostic specificity and sensitivity for PJI are limited when used alone, which has led to their underutilization in clinical decision-making.[16,17] Thus, utilizing real clinical data with advanced algorithms to identify optimal marker combinations, analyze their weight in PJI diagnosis, and quantify the probability of infection presents a more effective strategy for future PJI diagnosis and second-stage revision timing.

Machine learning (ML), as a form of artificial intelligence, can identify relationships between data, generates algorithms, and fine-tunes parameters to enhance the predictive accuracy of models, ultimately assisting in decision-making for future data.[18,19] Some studies have developed ML-based models for PJI prediction and diagnosis, showing its potential in addressing the challenges of PJI diagnosis. However, these models are difficult to apply for early preoperative prediction, as they include some uncommon tests and rely on intraoperative and postoperative data.[20–22] It is widely recognized that accurately differentiating PJI from aseptic cases preoperatively would greatly reduce patient suffering and prevent the waste of medical resources and costs. Thus, we aim to focus on developing a diagnostic model based on preoperative indicators to distinguish PJI from aseptic failure, providing better support for clinicians in making preoperative surgical decisions.

This study aims to retrospectively collect preoperative serum, synovial fluid markers, and baseline data from PJI patients based on MSIS, ICM, and European Bone and Joint infection society (EBJIS) criteria and data accessibility principles. Key variables will be identified, and various ML models will be employed to develop a novel preoperative PJI diagnostic model. Model interpretability and clinical applicability will be improved using Shapley Additive Explanations (SHAP), with continued evaluation of its clinical utility and accuracy in a prospective clinical cohort. We hope that this accessible, low-cost, and widely applicable preoperative diagnostic strategy to assist clinicians in rapidly and accurately diagnosing PJI and selecting the optimal timing for second-stage revision, enabling timely and appropriate treatment for patients.

# Materials and Methods

## Collection and Extraction of Patient Data

We retrospectively reviewed clinical data from our institution's database for patients admitted for joint revision surgery between January 2018 and October 2022. Patients requiring revision due to infection or aseptic failure were included (with ethical approval from the hospital's committee [No.20187101]) and randomly split into training and test cohort in a 7:3 ratio. We excluded other conditions that could interfere with the diagnosis of PJI. (Figure 1) The study was carried out following the STROBE guidelines and the STARD requirements, adhering to the latest version of the Declaration of Helsinki. Since the data used for analysis had identifiable information removed, patient consent for clinical data use was not required. Patients meeting the MSIS 2013 criteria were diagnosed with PJI, while aseptic failure refers to revision surgeries performed for reasons other than infection, such as wear, loosening, dislocation, instability, adverse local tissue reactions, or other causes.

As previously mentioned, this study aims to develop a PJI diagnostic model based on preoperative indicators to assist clinicians in preoperative management and decision-making, while ensuring the model's clinical usability. Following the principle of easy accessibility of preoperative indicators, and referencing MSIS,[23] ICM,[24,25] EBJIS standards,[26] as well as prior literature, we selected the following preoperative indicators for further screening: Serum (SE) markers: neutrophil percentage (N%), lymphocyte percentage (L%), platelets (PLT), ESR, SE-CRP, D-dimer (DD), serum

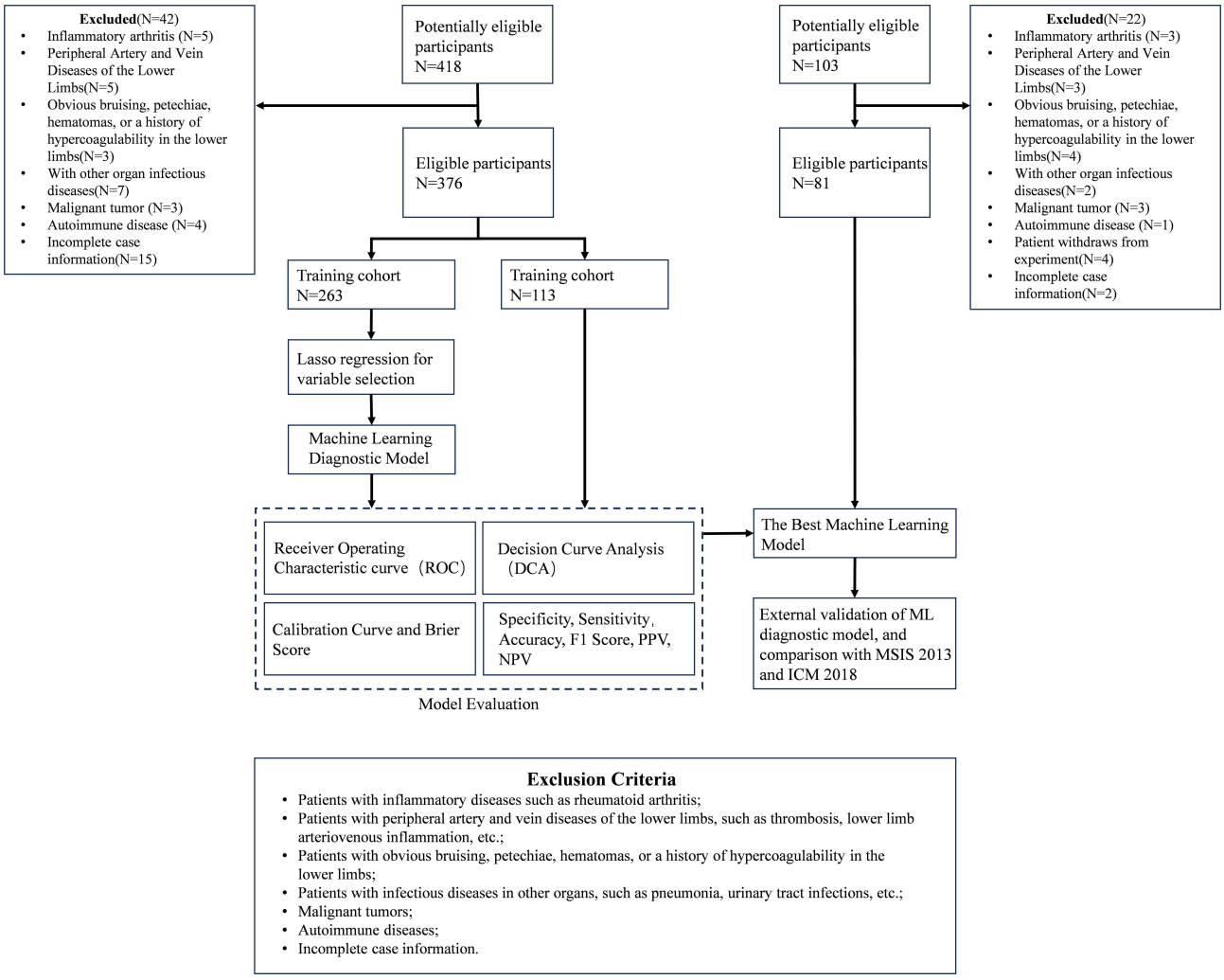


**Figure 1** Flowchart of the study.

Interleukin 6 (SE-IL6), and albumin (ALB). Synovial fluid (SF) markers: polymorphonuclear cell percentage (PMN%) and SF-IL6. We also gathered demographic data from patients, including sex, age, weight, and BMI.

## Model Development and Statistical Analysis

In the training cohort of this study, we applied Least Absolute Shrinkage and Selection Operator (Lasso) regression to select variables. Cross-validation was used to determine the relationship between various λ values and binomial deviance. As λ increased, unimportant variables with coefficients reduced to zero were excluded, allowing us to identify key variables and assess their importance. To maximize the diagnostic performance, we built eight models using the selected variables, including: Support Vector Machine (SVM), Neural Network (NNET), Random Forest (RF), eXtreme Gradient Boosting (XGB), K-Nearest Neighbors (KNN), Logistic Regression (LR), Gradient Boosting Machine (GBM), Classification and Regression Trees (CART).

In the test cohort, the diagnostic performance of each model was assessed by plotting ROC curves and analyzing the AUC to evaluate the model's ability to distinguish between binary samples. An AUC closer to 1 indicates stronger discriminatory power. Calibration curves were generated to evaluate whether the predicted probabilities accurately reflect the actual probability of events, with curves approaching the y=x line indicating higher prediction accuracy. The decision curve analysis (DCA) can help assess the utility of the model in practical decision-making and is used to evaluate its clinical applicability. The F1 score is the harmonic mean of precision (the proportion of true positives among predicted positives) and recall (the proportion of true positives among all actual positives). A higher F1 score indicates a more accurate model. Accuracy, sensitivity, and specificity were also included in our model evaluation.

Although ML models perform well in predictive tasks, their internal decision-making processes are often opaque and difficult to interpret, leading clinicians to potentially distrust the model's output and limiting its clinical adoption. We applied SHAP to visualize the model's output by calculating the contribution of each included variable to the diagnostic prediction and providing visual explanations for individual cases. In SHAP analysis, the raw output value of the model for each individual sample is represented as f(x), which corresponds to the log-odds in a binary classification model. A positive f(x) indicates that the model tends to classify the sample as "PJI", while a negative f(x) suggests a tendency toward "aseptic failure". By applying the sigmoid function (P = 1 / (1 + e^{-f(x)})), f(x) can be transformed into a probability value, allowing comparison with a diagnostic threshold for binary classification decisions.

## Prospective Validation of the Model

We prospectively included patients who underwent joint revision surgery between January 2022 and August 2024 for external validation of the diagnostic model (Clinical Trial Registration No. ChiCTR1800020440, Ethics Approval No. 2021–257) (Figure 1). Currently, there is no gold standard for PJI, but patients in whom bacteria are detected (preoperatively, intraoperatively) are generally considered to have PJI. Therefore, in this section, we used positive bacterial cultures from preoperative and intraoperative specimens as the diagnostic standard for PJI. Bacterial detection methods included routine cultures (tissue or synovial fluid), special cultures (eg, tuberculosis, fungi), pathological examination, polymerase chain reaction, gene chip, nucleic acid probe technology, and next-generation sequencing. All patients provided informed consent.

The ML models developed in this study were compared with the widely used MSIS 2013 and ICM 2018 criteria. The diagnostic performance of the ML models was externally validated using metrics such as sensitivity, specificity, and accuracy.

## Statistical Analysis

Continuous variables with a normal distribution are presented as mean ± standard deviation (SD), and statistical significance between two groups is determined by parametric *t*-tests. Otherwise, variables are described using interquartile range (IQR) and compared using the Mann–Whitney *U*-test. Categorical variables are presented as the number and percentage of patients in each category, and statistical significance is determined using the $\chi^2$-test or Fisher's exact test. A *P value* <0.05 is considered statistically significant. All statistical analyses and plots were performed using R 4.2.1.

# Results

## Demographic Data

In this study, we excluded a total of 42 patients (Figure 1). A final total of 376 cases were included, of which 111 (29.52%) were diagnosed with PJI. Results showed no significant differences in demographic characteristics (sex, age, weight, BMI) between the PJI and aseptic failure groups. Serum markers (SF-IL6, SE-CRP, ESR, PLT, N%, DD) and synovial fluid markers (SF-IL6, PMN%) were significantly higher in the PJI group compared to the aseptic failure group, while ALB and L% also exhibited significant differences (Table 1). The study population (N=376) was randomly divided into a training cohort and a test cohort in a 7:3 ratio for the construction and validation of the prediction model. Table 2 shows that there were no statistically significant differences in patient characteristics between the training and test cohort.

## Key Variable Selection and Diagnostic Performance Comparison of Multiple Models

In the process of variable selection using Lasso regression, the Lambda.1st parameter was applied to increase the penalty, ensuring that the selected variables had higher clinical significance. The final key variables selected included SE-IL6, SE-CRP, ESR, SF-IL6, PMN%, DD, and ALB, with their coefficients reflecting their importance and relevance in PJI diagnosis (Figure 2A and B; Supplementary Table 1).

The seven selected variables were applied to the ML model for PJI diagnosis and validated on the test cohort. By plotting ROC curves and analyzing AUC for eight machine learning models—XGB, NNET, RF, SVM, KNN, LR, GBM, and CART—XGB achieved a higher AUC of 0.998 (95% CI 0.993–1) in the test cohort, demonstrating superior discrimination for PJI compared to the other models, although they also showed good discriminatory ability (Figure 3A and B). To further evaluate the clinical utility of the model at different thresholds in a clinical setting, we constructed and analyzed calibration curves, Brier scores, and DCA. The calibration curve of the XGB model was closer to the "ideal calibration line" (diagonal) compared to other models, indicating that XGB's predicted probabilities were more reliable. The Brier scores of the different models further supported this result: XGB had the lowest Brier score of 0.062, meaning its predictions were closest to the actual outcomes (Figure 3C). The DCA curves showed that XGB demonstrated superior clinical applicability (Figure 3D). Additionally, we evaluated different ML models using metrics such as F1 score, sensitivity, and specificity. XGB achieved the highest F1 score of 0.985, accurately diagnosing the disease (high precision) and identifying most patients (high recall), thereby

**Table 1** Comparison of Demographic and Clinical Characteristics Between PJI and Aseptic Failure Patients

| Variables | Total(N=376) | Aseptic Failure(N=265) | PJI(N=111) | *P value* |
|---|---|---|---|---|
| Sex, N (%) | | | | 0.828 |
| Male | 176 (46.8%) | 125 (47.2%) | 51 (45.9%) | |
| Female | 200 (53.2%) | 140 (52.8%) | 60 (54.1%) | |
| Age (y), Median (Q1, Q3) | 66.00(59.00, 74.00) | 66.00(58.00, 73.00) | 67.00(59.50, 74.50) | 0.555 |
| Weight (kg), Median (Q1, Q3) | 60.00(54.38, 70.00) | 60.00(55.00, 70.00) | 60.00(53.50, 69.00) | 0.747 |
| BMI (kg/m2), Median (Q1, Q3) | 23.77(20.86, 25.81) | 23.63(20.80, 25.92) | 23.88(21.12, 25.37) | 0.725 |
| SE-IL6 (ng/mL), Median (Q1, Q3) | 4.82(4.16, 6.42) | 4.35(4.06, 5.15) | 8.06(6.29, 12.58) | <0.001 |
| SE-CRP (mg/mL), Median (Q1, Q3) | 5.76(3.92, 17.23) | 4.49(3.43, 6.52) | 26.20(12.95, 87.45) | <0.001 |
| ESR (mm/h), Median (Q1, Q3) | 32.00(16.00, 62.11) | 27.00(15.00, 52.00) | 57.13(23.00, 89.00) | <0.001 |
| SFIL-6 (ng/mL), Median (Q1, Q3) | 5.00(2.38, 11.06) | 3.33(2.05, 7.98) | 12.35(7.53, 20.32) | <0.001 |
| PMN%, Median (Q1, Q3) | 55.33(32.82, 71.41) | 52.40(12.00, 60.76) | 74.62(55.31, 84.44) | <0.001 |
| PLT ($10^9$/L), Median (Q1, Q3) | 172.00(124.00, 234.29) | 147.87(120.00, 216.34) | 209.00(168.50, 255.02) | <0.001 |
| N%, Median (Q1, Q3) | 61.60(55.89, 68.62) | 59.90(54.50, 66.60) | 65.19(58.85, 72.60) | <0.001 |
| L%, Median (Q1, Q3) | 24.95(17.90, 30.02) | 26.00(18.80, 30.78) | 22.70(15.77, 28.40) | 0.006 |
| DD (mg/mL), Median (Q1, Q3) | 0.68(0.33, 1.37) | 0.50(0.27, 0.83) | 1.74(1.19, 2.12) | <0.001 |
| ALB (g/L), Median (Q1, Q3) | 40.00(36.00, 44.00) | 42.00(38.00, 45.00) | 34.00(31.00, 38.00) | <0.001 |

**Note**: Statistically significant (*P-value* ≤ 0.05).

**Abbreviations**: PJI, periprosthetic joint infections; BMI, body mass index; N%, neutrophil percentage; L%, lymphocyte percentage; PLT, platelets; ESR, erythrocyte sedimentation rate; SE-CRP, Serum C-reactive protein; DD, D-dimer; SE-IL6, serum interleukin 6; ALB, albumin; PMN%, polymorphonuclear cell percentage; SF-IL6, synovial fluid interleukin 6.


https://doi.org/10.2147/JIR.S499903

**Table 2** Comparison of Demographic and Clinical Characteristics Between Train and Test Cohorts

| Variables | Total (N=376) | Trian (N=263) | Test (N=113) | *P value* |
|---|---|---|---|---|
| Sex, N (%) | | | | 0.301 |
| Male | 176 (46.8%) | 121 (46.0%) | 55 (48.7%) | |
| Female | 200 (53.2%) | 142 (54.0%) | 58 (51.3%) | |
| Age (y), Median (Q1, Q3) | 66.00(59.00, 74.00) | 66.00(57.50, 73.00) | 66.00(60.00, 74.00) | 0.300 |
| Weight (kg), Median (Q1, Q3) | 60.00(54.38, 70.00) | 60.00(55.00, 70.00) | 60.00(53.00, 69.00) | 0.306 |
| BMI (kg/m$^2$), Median (Q1, Q3) | 23.77(20.86, 25.81) | 23.83(21.25, 25.98) | 23.39(20.43, 25.01) | 0.141 |
| SE-IL6 (ng/mL), Median (Q1, Q3) | 4.82(4.16, 6.42) | 4.69(4.16, 6.41) | 5.07(4.15, 6.42) | 0.449 |
| SE-CRP (mg/mL), Median (Q1, Q3) | 5.76(3.92, 17.23) | 5.98(3.98, 16.80) | 5.73(3.87, 17.70) | 0.904 |
| ESR (mm/h), Median (Q1, Q3) | 32.00(16.00, 62.11) | 35.00(16.50, 61.50) | 29.00(16.00, 64.00) | 0.371 |
| SF-IL6 (ng/mL), Median (Q1, Q3) | 5.00(2.38, 11.06) | 5.00(2.74, 10.96) | 5.01(2.21, 11.44) | 0.633 |
| PMN%, Median (Q1, Q3) | 55.33(32.82, 71.41) | 55.54(13.51, 70.32) | 54.70(39.81, 74.52) | 0.458 |
| PLT ($10^9$/L), Median (Q1, Q3) | 172.00(124.00, 234.29) | 174.00(124.00, 232.03) | 171.00(128.08, 235.00) | 0.961 |
| N%, Median (Q1, Q3) | 61.60(55.89, 68.62) | 60.90(55.91, 69.15) | 62.02(55.60, 67.30) | 0.840 |
| L%, Median (Q1, Q3) | 24.95(17.90, 30.02) | 25.83(18.30, 30.64) | 22.88(15.80, 28.94) | 0.078 |
| DD (mg/mL), Median (Q1, Q3) | 0.68(0.33, 1.37) | 0.66(0.32, 1.37) | 0.69(0.35, 1.40) | 0.995 |
| ALB (g/L), Median (Q1, Q3) | 40.00(36.00, 44.00) | 40.00(35.00, 44.00) | 40.00(37.00, 44.00) | 0.528 |

**Note**: Statistically significant (*P-value* ≤ 0.05).

**Abbreviations**: PJI, periprosthetic joint infections; BMI, body mass index; N%, neutrophil percentage; L%, lymphocyte percentage; PLT, platelets; ESR, erythrocyte sedimentation rate; SE-CRP, Serum C-reactive protein; DD, D-dimer; SE-IL6, serum interleukin 6; ALB, albumin; PMN%, polymorphonuclear cell percentage; SF-IL6, synovial fluid interleukin 6.

reducing misdiagnoses and missed diagnoses. XGB's sensitivity of 0.971, specificity of 1.000, and accuracy of 0.991 were also superior to those of other models (Table 3). Overall analysis indicates that XGB can be regarded as the best-performing model in our study.

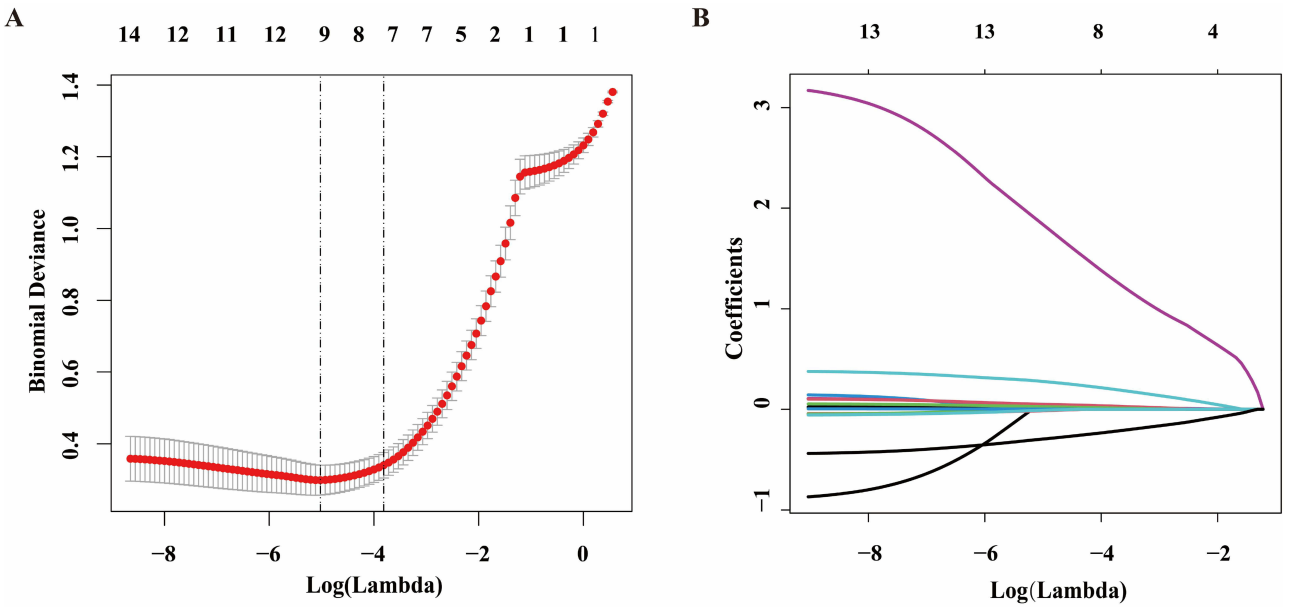


**Figure 2** Demographics and variable characteristics selection using LASSO regression. (**A**) Binomial bias versus log (Lambda) in the LASSO regression model. The dashed line on the left indicates Lambda.min (the minimum value of Lambda) and the dashed line on the right indicates Lambda.1st (the maximum value within one standard error of Lambda.min). The numbers at the top of the figure indicate the number of variables included in the model at each Lambda value. (**B**) LASSO coefficient trajectories for selected variables under log (Lambda). Unimportant coefficients shrink to zero as Lambda increases. Y-axis denotes the coefficient of the variable and x-axis denotes log (Lambda).

**Abbreviations**: Lasso, Least Absolute Shrinkage and Selection Operator.

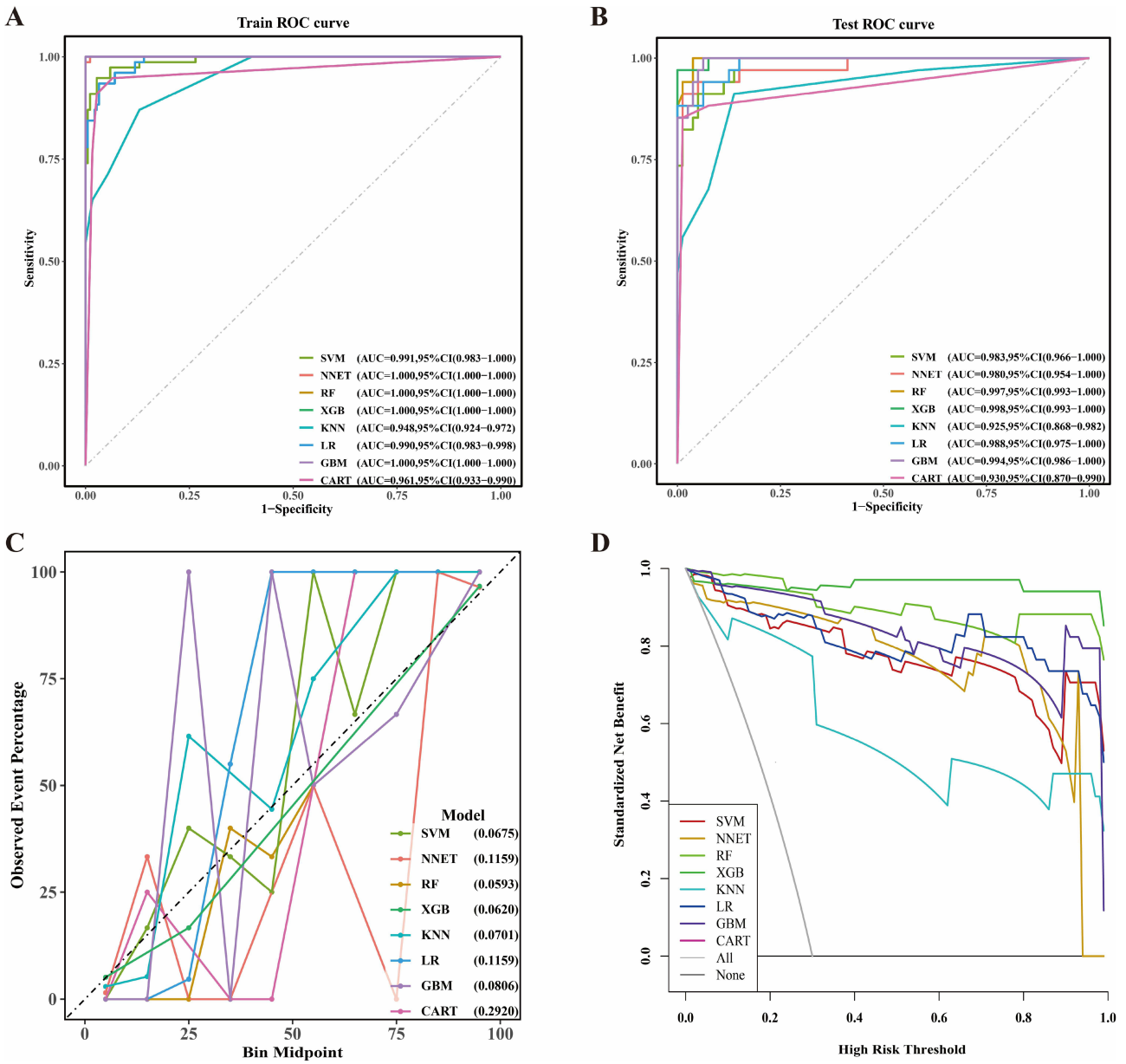


**Figure 3** Evaluating the diagnostic performance of different models. (**A** and **B**) ROC curves of the training and test sets for evaluating the ability of the eight machine learning models to discriminate between PJI and aseptic failures. The AUC values and their 95% confidence intervals are displayed in the figure. (**C**) Calibration curves showing how well the predicted probabilities of the eight models match the actual observed events. The Brier scores obtained by each model are shown on the right side of the figure. XGB obtained better Brier scores 0.062. (**D**) Decision Curve Analysis evaluates the clinical net benefit of each model at different risk thresholds.
**Abbreviations**: AUC, area under curve; ROC; receiver operating characteristic; SVM, Support Vector Machine; NNET, Neural Network; RF, Random Forest; XGB, eXtreme Gradient Boosting; KNN, K-Nearest Neighbors; LR, Logistic Regression; GBM; Gradient Boosting Machine; CART, Classification and Regression Trees.

## SHAP Interpretation of the Model

To improve model interpretability and avoid the black-box phenomenon in machine learning, we applied SHAP to explain XGB's decision-making process. SHAP allows for quantifying the contribution and importance of each feature in the model's predictions and reveals how each feature influences predictions at the individual sample level. Figure 4A shows the feature importance of the seven variables based on SHAP values and their impact on the prediction results. High values of SE-IL6 and DD were significantly positively correlated with PJI, while high values of ALB were negatively correlated with predicting PJI. SE-IL6, SE-CRP, SF-IL6, DD, and ALB had a greater impact on the model's prediction results, while ESR and PMN% had relatively smaller effects (Figure 4A and B). Additionally, we randomly

**Table 3** Performance Metrics of Eight Models in the Test Cohort

| Model | AUC (95% CI) | Sensitivity | Specificity | Accuracy | PPV | NPV | F1 Score |
|---|---|---|---|---|---|---|---|
| NNET | 0.980 (0.954–1.000) | 0.912 | 0.988 | 0.965 | 0.969 | 0.963 | 0.939 |
| RF | 0.997 (0.993–1.000) | 1.000 | 0.963 | 0.974 | 0.919 | 1.000 | 0.958 |
| SVM | 0.983 (0.966–1.000) | 0.912 | 0.950 | 0.939 | 0.886 | 0.962 | 0.899 |
| XGB | 0.998 (0.993–1.000) | 0.971 | 1.000 | 0.991 | 1.000 | 0.988 | 0.985 |
| KNN | 0.925 (0.868–0.982) | 0.912 | 0.863 | 0.877 | 0.738 | 0.958 | 0.816 |
| LR | 0.988 (0.975–1.000) | 0.882 | 1.000 | 0.965 | 1.000 | 0.952 | 0.938 |
| GBM | 0.994 (0.986–1.000) | 1.000 | 0.938 | 0.956 | 0.872 | 1.000 | 0.932 |
| CART | 0.930 (0.870–0.990) | 0.853 | 0.988 | 0.947 | 0.967 | 0.940 | 0.906 |

**Abbreviations**: CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; SVM, Support Vector Machine; NNET, Neural Network; RF, Random Forest; XGB, eXtreme Gradient Boosting; KNN, K-Nearest Neighbors; LR, Logistic Regression; GBM; Gradient Boosting Machine; CART, Classification and Regression Trees.

selected two typical examples to demonstrate the interpretability of the model: one PJI patient with a high SHAP prediction score (0.73) (Figure 4C), and another aseptic failure patient with a low SHAP score (−0.335) (Figure 4D).

## Prospective Validation

To ensure the reliability of the model in future clinical applications, we prospectively collected 81 patients as an external validation cohort, of which 25 were diagnosed with PJI (Figure 1). The variable characteristics of the patients are shown in Supplementary Table 2. Our proposed machine learning model achieved an accuracy of 95.06%, with sensitivity and specificity of 92.00% and 96.43%, respectively, in the prospective dataset. In comparison, the MSIS 2013 criteria had an accuracy, sensitivity, and specificity of 83.13%, 70.37%, and 89.29%, respectively, while the ICM 2018 criteria had an accuracy, sensitivity, and specificity of 92.59%, 84.00%, and 96.43%, respectively. In summary, the machine learning model demonstrated higher accuracy and sensitivity in preoperative diagnosis compared to the MSIS 2013 and ICM 2018 criteria, with similar specificity (Table 4).

# Discussion

The diagnosis of PJI has always been one of the main challenges in clinical practice, and with societal development and population growth, the incidence of PJI has been rising annually.[27] Currently, two-stage revision surgery is still one of the primary approaches to treating PJI, but infection must be fully controlled preoperatively.[17] Therefore, whether PJI patients can be diagnosed early and rapidly determines the timing for clinicians to initiate prompt anti-infection treatment and proceed with two-stage revision surgery, thereby improving the success and cure rates of the surgery.[28] This study's ML-based preoperative PJI diagnostic model, built using patient baseline data, preoperative serum, and synovial fluid markers, achieved high sensitivity (92.00%) and specificity (96.43%), outperforming the traditional MSIS 2013 criteria (sensitivity of 70.37% and specificity of 89.29%) and the ICM 2018 criteria (sensitivity of 92.59%, specificity of 96.43%). This provides a new reference for the preoperative diagnosis of PJI, helping doctors formulate more accurate treatment strategies.

Currently, the PJI diagnostic criteria established by associations such as MSIS, ICM, and EBJIS are widely used, but the combination of indicators in these standards is primarily based on expert consensus and literature reviews, lacking systematic evaluation of various diagnostic combinations.[20] Diagnostic indicators may carry different weights in different cases, and existing diagnostic thresholds can sometimes lead to ambiguous results in borderline patients. In this study, we used ML to identify relationships among various indicators and their importance in diagnosing PJI, ultimately selecting seven key variables and constructing the model based on their weights. Additionally, unlike traditional diagnostic criteria based on set thresholds, our model dynamically calculates the contribution of each indicator to PJI diagnosis through functional relationships, enhancing its ability to diagnose complex cases. Studies have shown that the application of ML models has demonstrated accurate PJI diagnostic capabilities and promising clinical potential. Both the Yuh-Jyh Hu team and the Puyi Sheng team have used machine learning to develop PJI diagnostic tools that are comparable to existing

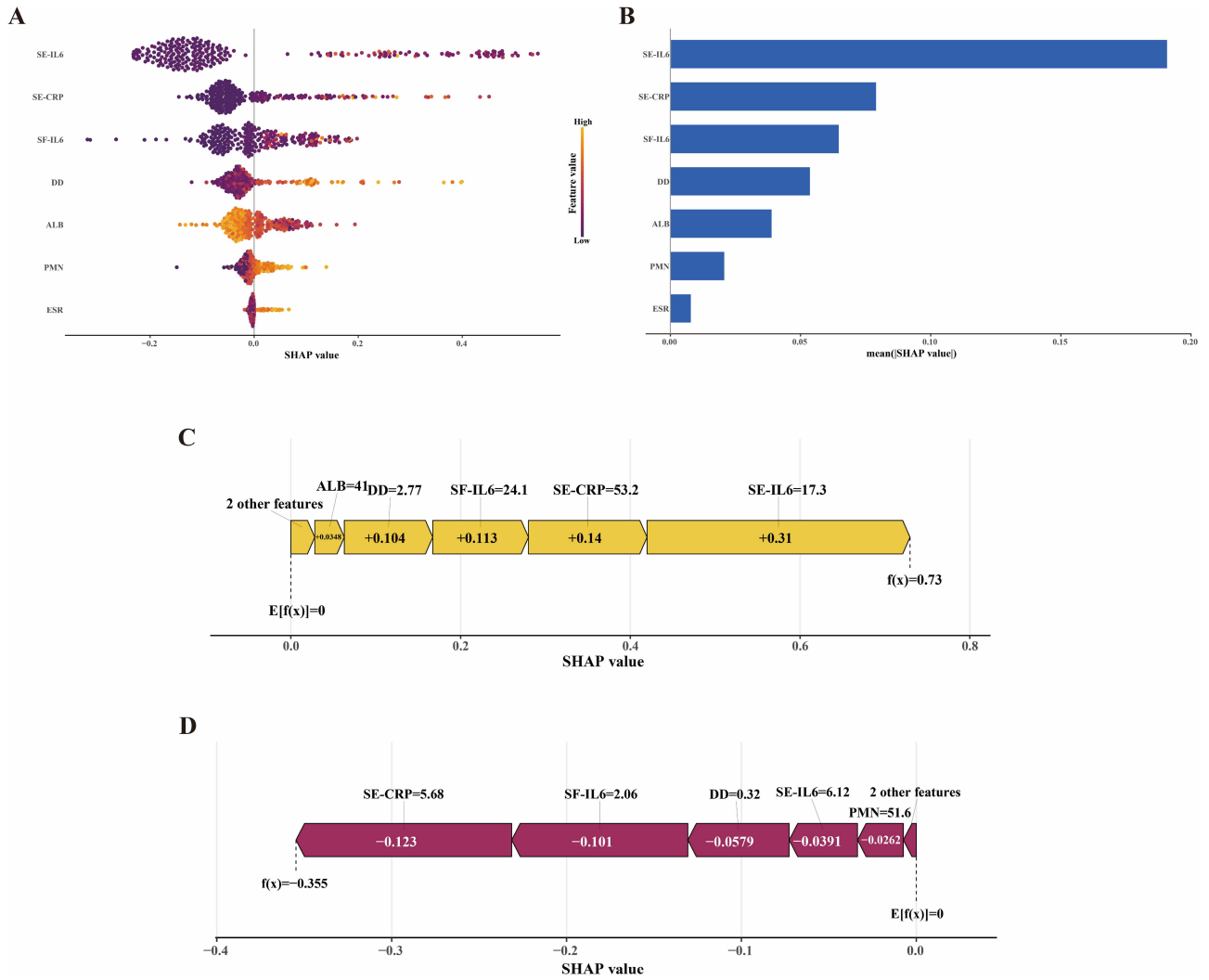


**Figure 4** SHAP interpretation of XGB machine learning model. (**A**) A scatter plot based on SHAP values shows the contribution of each variable to PJI diagnosis. The X-axis represents SHAP values, the Y-axis represents variable names, and the color indicates the contribution range of the variables (from low to high). (**B**) The average SHAP values of each variable reflect their relative importance in the PJI diagnosis model. (**C** and **D**) Individual SHAP force plots for a PJI patient (**C**) and an aseptic failure patient (**D**). The SHAP values represent the diagnostic features of individual patients and the contribution of each variable to diagnosing PJI. f(x) represents the predicted probability of being diagnosed with PJI, with values closer to 1 indicating a higher likelihood of PJI diagnosis; a negative f(x) indicates that the patient's features are pushing the model's diagnosis toward aseptic failure. E[f(x)] is the baseline value of the model when no sample input is provided. The length of the arrows indicates the degree to which the diagnosis is influenced by each variable—the longer the arrow, the greater the impact.

**Abbreviations**: XGB, eXtreme Gradient Boosting; SHAP, Shapley Additive Explanations; PJI, periprosthetic joint infections; ESR, erythrocyte sedimentation rate; SE-CRP, Serum C-reactive protein; DD, D-dimer; SE-IL6, serum interleukin 6; ALB, albumin; PMN%, polymorphonuclear cell percentage; SF-IL6, synovial fluid interleukin 6.

diagnostic standards.[20,22] However, since a large number of indicators and clinical data were initially included in the model development, their diagnostic models often involve more than a dozen indicators, which requires comprehensive clinical data and increases the financial burden on patients. Additionally, some diagnostic models include indicators that

**Table 4** Diagnostic Performance of the ML Model Versus Traditional Criteria for PJI (External Validation)

| Criteria | PJI(N=25) | | Aseptic(N=56) | | Accuracy | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|---|---|---|---|---|
| | TRUE | FALSE | TRUE | FALSE | | | | | |
| MSIS 2013 | 19 | 8 | 50 | 6 | 83.13% | 70.37% | 89.29% | 76.00% | 86.21% |
| ICM 2018 | 21 | 4 | 54 | 2 | 92.59% | 84.00% | 96.43% | 91.30% | 93.10% |
| ML model | 23 | 2 | 54 | 2 | 95.06% | 92.00% | 96.43% | 92.00% | 96.43% |

**Abbreviations**: PJI, periprosthetic joint infections; PPV, positive predictive value; NPV, negative predictive value; MSIS, Musculo-Skeletal Infection Society; ICM, International Consensus Meetings; ML, machine learning.


https://doi.org/10.2147/JIR.S499903

impose high demands on hospital equipment and skilled personnel, especially in resource-limited primary care settings, making it difficult for these models to be applied in routine clinical practice.[29] Currently, intraoperative and postoperative diagnostic indicators are still a crucial part of existing PJI diagnostic standards. Zhang et al demonstrated that sonication fluid culture from antibiotic-loaded bone cement spacers shows high specificity (94%) in determining infection eradication after two-stage revision surgery; however, its utility in early preoperative assessment of infection remains limited.[30] Similarly, PJI diagnostic models reported in studies also rely on intraoperative and postoperative indicators, which prevents clinicians from accurately diagnosing patients preoperatively, potentially leading to missed optimal intervention windows or even misdiagnosis.[20,22,24–26] In this study, to address the above limitations, we referenced the MSIS, ICM, and EBJIS diagnostic criteria and prioritized selecting preoperative indicators that are easy to use in clinical practice. We developed a preoperative diagnostic model that is easy to detect and implement, offers high diagnostic accuracy, and does not rely on microbiological testing.

After LASSO screening, SE-IL6, SE-CRP, ESR, SF-IL6, PMN%, DD, and ALB were selected for model development. Serum ESR and CRP, as systemic inflammatory markers, are widely included in PJI diagnostic criteria due to their ease of access, low cost, and widespread use globally.[13] In this study, serum ESR and CRP were included in the final diagnostic model, and according to SHAP, serum CRP contributed more to disease diagnosis than ESR. This may be due to CRP's high sensitivity to infection and inflammatory responses as an acute-phase protein, whereas ESR is more easily affected by other factors.[13,31] This is consistent with previous research findings, where ESR's sensitivity and specificity for PJI are lower than those of CRP.[9,32,33] IL6 is produced by immune cells such as monocytes and macrophages and is one of the most important mediators of the acute inflammatory response.[34] Previous studies have shown that IL6 levels have good specificity and sensitivity in diagnosing PJI. However, SE-IL6 can be influenced by other systemic inflammations, while SF-IL6 has greater site-specificity.[10] Our previous research also found that combining SE-IL6 and SF-IL6 for diagnosing PJI demonstrated higher diagnostic performance. This is similar to the findings of this study, where both SE-IL6 and SF-IL6 significantly contributed to PJI diagnosis.[4] Synovial PMN% is listed as one of the diagnostic markers in the ICM 2018 criteria, and it was also included in the final diagnostic model we developed.[25] A meta-analysis has shown that D-dimer exhibits lower sensitivity and specificity compared to fibrinogen, and its overall diagnostic performance is inferior to traditional markers such as CRP and ESR. However, in our model, the combined use of D-dimer with CRP, ESR, and other indicators significantly enhanced the overall diagnostic accuracy for PJI, highlighting its clinical value as a supportive biomarker.[35] Infection, especially bacterial infection, triggers an inflammatory response, which stimulates the activation of the fibrinolytic and coagulation systems, ultimately resulting in elevated DD levels.[36] In the ICM 2018 criteria, DD is considered one of the important diagnostic markers for PJI infection, which is consistent with our study's finding that DD contributes significantly to PJI diagnosis.[25] Additionally, our ML model found that ALB, a biochemical marker for assessing nutritional status, is closely associated with the occurrence of PJI. ALB values were negatively correlated with PJI diagnosis, possibly due to infection leading to a negative nitrogen balance and increased protein consumption in patients. Furthermore, a decrease in ALB often indicates compromised immune function, which may exacerbate the infection.[37] In summary, this study identified the most important diagnostic indicators, combining serum and synovial fluid markers. The model was constructed by assigning different weights to each diagnostic test based on their diagnostic ability and characteristics in PJI diagnosis, ultimately improving the model's diagnostic performance. Furthermore, we visualized the ML model using SHAP, which revealed the contribution of each indicator to PJI diagnosis. This provides a reference for clinical testing and improves the model's acceptance in clinical practice. In external validation, our model's preoperative diagnostic performance surpassed that of MSIS 2013 and ICM 2018, enabling clinicians to make better surgical decisions.

Naturally, this study also has some limitations. Due to the widespread acceptance of the 2013 MSIS criteria, we used these standards as the basis for diagnosis. However, there is currently no gold standard for diagnosing PJI, which may lead to missed or incorrect diagnoses, potentially impacting our study results. This is a harsh and challenging situation that all studies evaluating diagnostic tests for PJI infection likely need to address. Furthermore, this is a retrospective study, with patients mainly from Southwest China, and it lacks external validation cohorts from multiple centers. This may introduce regional limitations and potential bias, although the results of this study are promising. Although we have prospectively validated the model in an independent cohort and preliminarily confirmed its satisfactory predictive

performance and stability, large-scale, multicenter prospective studies are still warranted to further evaluate its accuracy and clinical applicability. Lastly, to minimize confounding factors, we excluded patients with underlying diseases that affect inflammatory factor expression. However, this may limit the generalizability of our findings, as real-world practice may involve different conditions.

## Conclusion

Our ML model constructed using SE-IL6, SE-CRP, ESR, SF-IL6, PMN%, DD, and ALB demonstrated good differentiation between PJI and aseptic failure in both internal and external validation. The indicators used in our model are easy to obtain and low-cost, allowing for more accurate preoperative diagnosis. This enables clinicians to make informed decisions before resorting to more expensive and invasive tests, offering high economic value and clinical applicability. The ML model has the potential to become a valuable tool for supporting preoperative PJI diagnosis and surgical decision-making.

## Data Sharing Statement

The data that support the findings of this study are available from the corresponding author, Professor Leilei Qin upon reasonable request.

## Ethical Approval

The ethical approval for this study was obtained from the Institutional Ethics Committee of the First Affiliated Hospital of Chongqing Medical University (Approval No. 20187101 and No. 2021-257), and the study was also registered with the Chinese Clinical Trial Registry (Registration No. ChiCTR1800020440).

## Informed Consent

Informed consent was obtained from all individual participants included in the study.

## Acknowledgments

This study used ChatGPT4.0 to assist the author in improving the language of the first draft after its completion, in order to enhance the fluency and readability of the text. However, the accuracy of the content, viewpoints, and responsibility for the final article are assumed by the author.

## Author Contributions

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

## Funding

This work was supported by Tibet Autonomous Region Natural Science Foundation Medical Aid Group Project (Grant No. XZ2024ZR-ZY102(Z)); National Natural Science Foundation of China (Grant No. 82402836); Chongqing's Special Funding for Postdoctoral Research Projects (Grant No. 2023CQBSHTB3124).

## Disclosure

The authors have not disclosed any competing interests.

## References

1. Wouthuyzen-Bakker M, Sebillotte M, Arvieux C, et al. How to handle concomitant asymptomatic prosthetic joints during an episode of hematogenous periprosthetic joint infection, a multicenter analysis. *Clin Infect Dis*. 2020;73(11):e3820–e3824. doi:10.1093/cid/ciaa1222

2. Senneville E, Dinh A, Ferry T, Beltrand E, Blondiaux N, Robineau O. Tolerance of prolonged oral tedizolid for prosthetic joint infections: results of a multicentre prospective study. *Antibiotics*. 2020;10(1):4. doi:10.3390/antibiotics10010004
3. Almazrou Y, Shibl AM, Alkhlaif R, et al. Epidemiology of invasive pneumococcal disease in Saudi Arabian children younger than 5 years of age. *J Epidemiol Glob Health*. 2016;6(2):95–104. doi:10.1016/j.jegh.2015.08.002
4. Qin L, Li X, Wang J, Gong X, Hu N, Huang W. Improved diagnosis of chronic hip and knee prosthetic joint infection using combined serum and synovial IL-6 tests. *Bone Joint Res*. 2020;9(9):587–592. doi:10.1302/2046-3758.99.BJR-2020-0095.R1
5. Qin L, Yang S, Zhao C, et al. Prospects and challenges for the application of tissue engineering technologies in the treatment of bone infections. *Bone Res*. 2024;12(1):28. doi:10.1038/s41413-024-00332-w
6. Sigmund IK, Puchner SE, Windhager R. Serum inflammatory biomarkers in the diagnosis of periprosthetic joint infections. *Biomedicines*. 2021;9(9):1128. doi:10.3390/biomedicines9091128
7. Zou Y, Yang Y, Yang J, et al. The utility of synovial fluid interleukin-10 in diagnosing chronic periprosthetic joint infection: a prospective cohort study. *Infect Drug Resist*. 2025;18:533–542. doi:10.2147/IDR.S490962
8. Yang Y, Zhou H, Kang R, et al. Quantifying the long and short axes of the external iliac lymph nodes using dual-energy computed tomography: a potential diagnostic approach for periprosthetic joint infection - A prospective study. *Infect Drug Resist*. 2024;17:5605–5617. doi:10.2147/IDR.S497736
9. Qin L, Li F, Gong X, Wang J, Huang W, Hu N. Combined measurement of D-dimer and C-reactive protein levels: highly accurate for diagnosing chronic periprosthetic joint infection. *J Arthroplasty*. 2020;35(1):229–234. doi:10.1016/j.arth.2019.08.012
10. Zhou H, Yang Y, Zhang Y, et al. Current status and perspectives of diagnosis and treatment of periprosthetic joint infection. *Infect Drug Resist*. 2024;17:2417–2429. doi:10.2147/IDR.S457644
11. Cheng Q, Yang Y, Li F, Li X, Qin L, Huang W. Dual-energy computed tomography iodine maps: application in the diagnosis of periprosthetic joint infection in total hip arthroplasty. *J Arthroplasty*. 2025;40(2):499–505. doi:10.1016/j.arth.2024.08.007
12. Chong YY, Chan PK, Chan VWK, et al. Application of machine learning in the prevention of periprosthetic joint infection following total knee arthroplasty: a systematic review. *Arthroplasty*. 2023;5(1):38. doi:10.1186/s42836-023-00195-2
13. Tripathi S, Tarabichi S, Parvizi J, Rajgopal A. Current relevance of biomarkers in diagnosis of periprosthetic joint infection: an update. *Arthroplasty*. 2023;5(1):41. doi:10.1186/s42836-023-00192-5
14. Busch A, Jäger M, Beck S, et al. Metal artefact reduction sequences (Mars) in magnetic resonance imaging (MRI) after total hip arthroplasty (THA): a non-invasive approach for preoperative differentiation between periprosthetic joint infection (PJI) and aseptic complications? *BMC Musculoskelet Disord*. 2022;23(1):620. doi:10.1186/s12891-022-05560-x
15. Tande AJ, Patel R. Prosthetic joint infection. *Clin Microbiol Rev*. 2014;27(2):302–345. doi:10.1128/CMR.00111-13
16. Klemt C, Tirumala V, Smith EJ, Padmanabha A, Kwon Y-M. Development of a preoperative risk calculator for reinfection following revision surgery for periprosthetic joint infection. *J Arthroplasty*. 2021;36(2):693–699. doi:10.1016/j.arth.2020.08.004
17. Yaghi N, Yaghi C, Abifadel M, Boulos C, Feart C. Dietary patterns and risk factors of frailty in Lebanese older adults. *Nutrients*. 2021;13(7):2188. doi:10.3390/nu13072188
18. Ry C, As C, J K-C, Mf C, Jp C. Introduction to machine learning, neural networks, and deep learning. *Trans Vision Sci Technol*. 2020;9(2). doi:10.1167/tvst.9.2.14
19. Guni A, Normahani P, Davies A, Jaffer U. Harnessing machine learning to personalize web-based health care content. *J Med Internet Res*. 2021;23(10):e25497. doi:10.2196/25497
20. Chen W, Hu X, Gu C, et al. A machine learning-based model for "in-time" prediction of periprosthetic joint infection. *Digit Health*. 2024;10:20552076241253531. doi:10.1177/20552076241253531
21. Klemt C, Yeo I, Harvey M, et al. The use of artificial intelligence for the prediction of periprosthetic joint infection following aseptic revision total knee arthroplasty. *J Knee Surg*. 2023;37(2):158–166. doi:10.1055/s-0043-1761259
22. Kuo F-C, Hu W-H, Hu Y-J. Periprosthetic joint infection prediction via machine learning: comprehensible personalized decision support for diagnosis. *J Arthroplasty*. 2022;37(1):132–141. doi:10.1016/j.arth.2021.09.005
23. Parvizi J, Zmistowski B, Berbari EF, et al. New definition for periprosthetic joint infection: from the workgroup of the musculoskeletal infection society. *Clin Orthopaedics Related Res*. 2011;469(11):2992. doi:10.1007/s11999-011-2102-9
24. Parvizi J, Gehrke T. Definition of periprosthetic joint infection. *J Arthroplasty*. 2014;29(7):1331. doi:10.1016/j.arth.2014.03.009
25. Parvizi J, Tan TL, Goswami K, et al. The 2018 definition of periprosthetic hip and knee infection: an evidence-based and validated criteria. *J Arthroplasty*. 2018;33(5):1309–1314.e2. doi:10.1016/j.arth.2018.02.078
26. McNally M, Sousa R, Wouthuyzen-Bakker M, et al. The EBJIS definition of periprosthetic joint infection. *Bone Joint J*. 2021;103-B(1):18–25. doi:10.1302/0301-620X.103B1.BJJ-2020-1381.R1
27. Diez-Escudero A, Carlsson E, Andersson B, Järhult JD, Hailer NP. Trabecular titanium for orthopedic applications: balancing antimicrobial with osteoconductive properties by varying silver contents. *ACS Appl Mater Interfaces*. 2022;14(37):41751–41763. doi:10.1021/acsami.2c11139
28. Ascione T, Balato G, Festa E, Pandolfo G, Siciliano R, Pagliano P. Ideal timing of reimplantation in patients with periprosthetic knee infection undergoing 2-stage exchange: a diagnostic scoring system. *J Bone Joint Surg*. 2024;106(11):984–991. doi:10.2106/JBJS.23.00424
29. Yang J, Qin L, Zhao C, et al. Diagnosis of periprosthesis joint infection and selection of replantation timing: a novel nomogram diagnosis model. *Int J Surg*. 2024. doi:10.1097/JS9.0000000000002067
30. Zhang Q, Ding B, Wu J, Dong J, Liu F. Sonication fluid culture of antibiotic-loaded bone cement spacer has high accuracy to confirm eradication of infection before reimplantation of new prostheses. *J Orthop Surg Res*. 2021;16(1):377. doi:10.1186/s13018-021-02520-4
31. Steinvil A, Shapira I, Arbel Y, Justo D, Berliner S, Rogowski O. Determinants of the erythrocyte sedimentation rate in the era of microinflammation: excluding subjects with elevated c-reactive protein levels. *Am J Clin Pathol*. 2008;129(3):486–491. doi:10.1309/U04E2YFJRR6JQQTK
32. Wang H, Qin L, Wang J, Huang W. Synovial fluid IL-1β appears useful for the diagnosis of chronic periprosthetic joint infection. *J Orthop Surg Res*. 2021;16(1):144. doi:10.1186/s13018-021-02296-7
33. Su X, Chen Y, Zhan Q, et al. The ratio of IL-6 to IL-4 in synovial fluid of knee or hip performances a noteworthy diagnostic value in prosthetic joint infection. *J Clin Med*. 2022;11(21):6520. doi:10.3390/jcm11216520
34. Li B, Wang X, Choi IY, et al. miR-146a modulates autoreactive Th17 cell differentiation and regulates organ-specific autoimmunity. *J Clin Invest*. 2017;127(10):3702–3716. doi:10.1172/JCI94012

35. Zhang Q, Dong J, Zhou D, Liu F. Circulating D-dimer versus fibrinogen in the diagnosis of peri-prosthetic joint infection: a meta-analysis. *Surg Infect*. 2021;22(2):200–210. doi:10.1089/sur.2019.298
36. Keshari RS, Silasi R, Lupu C, Taylor FB, Lupu F. In vivo–generated thrombin and plasmin do not activate the complement system in baboons. *Blood*. 2017;130(24):2678–2681. doi:10.1182/blood-2017-06-788216
37. Liang F. Regulatory effect and mechanisms of carbon monoxide-releasing molecule II on hepatic energy metabolism in septic mice. *World J Gastroenterol*. 2014;20(12):3301–3311. doi:10.3748/wjg.v20.i12.3301