BMC Musculoskeletal Disorders

RESEARCH

Open Access

Predicting periprosthetic joint infection in primary total knee arthroplasty: a machine learning model integrating preoperative and perioperative risk factors

Yuk Yee Chong[1], Chun Man Lawrence Lau[1*], Tianshu Jiang[2], Chunyi Wen[2], Jiang Zhang[3], Amy Cheung[4], Michelle Hilda Luk[4], Ka Chun Thomas Leung[4], Man Hong Cheung[1], Henry Fu[1], Kwong Yuen Chiu[1] and Ping Keung Chan[1*]

Abstract

Background Periprosthetic joint infection leads to significant morbidity and mortality after total knee arthroplasty. Preoperative and perioperative risk prediction and assessment tools are lacking in Asia. This study developed the first machine learning model for individualized prediction of periprosthetic joint infection following primary total knee arthroplasty in this demographic.

Methods A retrospective analysis was conducted on 3,483 primary total knee arthroplasty (81 with periprosthetic joint infection) from 1998 to 2021 in a Chinese tertiary and quaternary referral academic center. We gathered 60 features, encompassing patient demographics, operation-related variables, laboratory findings, and comorbidities. Six of them were selected after univariate and multivariate analysis. Five machine learning models were trained with stratified 10-fold cross-validation and assessed by discrimination and calibration analysis to determine the optimal predictive model.

Results The balanced random forest model demonstrated the best predictive capability with average metrics of 0.963 for the area under the receiver operating characteristic curve, 0.920 for balanced accuracy, 0.938 for sensitivity, and 0.902 for specificity. The significant risk factors identified were long operative time (OR, 9.07; $p=0.018$), male gender (OR, 3.11; $p<0.001$), ASA > 2 (OR, 1.68; $p=0.028$), history of anemia (OR, 2.17; $p=0.023$), and history of septic arthritis (OR, 4.35; $p=0.030$). Spinal anesthesia emerged as a protective factor (OR, 0.55; $p=0.022$).

Conclusion Our study presented the first machine learning model in Asia to predict periprosthetic joint infection following primary total knee arthroplasty. We enhanced the model's usability by providing global and local interpretations. This tool provides preoperative and perioperative risk assessment for periprosthetic joint infection and opens the potential for better individualized optimization before total knee arthroplasty.

*Correspondence:
Chun Man Lawrence Lau
laucml@hku.hk
Ping Keung Chan
cpk464@yahoo.com.hk

Full list of author information is available at the end of the article

Keywords Periprosthetic joint infection, Primary total knee arthroplasty, Joint replacement, Machine learning, Artificial intelligence, Prediction model, Preoperative factor, Perioperative factor

Background

An escalating number of total knee arthroplasty (TKA), the definitive treatment for end-stage knee osteoarthritis, is being performed due to the aging population [1]. Periprosthetic joint infection (PJI), the most common complication leading to TKA revision, results in significant morbidity and mortality, with one-year and five-year mortality rates reaching up to 10.6% and 25.9% respectively [2–4]. Other than inferior clinical outcomes like pain and poor function, PJI also increases hospital resource utilization due to the complexities of revision surgery, extended antibiotic treatment, and prolonged hospital stays [4, 5]. The number of PJI is forecasted to increase due to the rise in the number of TKA performed. By 2030, the anticipated total annual hospital cost in the United States for PJI following TKA is estimated to reach $1.1 billion, posing a substantial economic burden [6]. Strategies to reduce and optimize the management of PJI in TKA would thus provide significant patient- and society-level benefits.

In order to minimize the chance of having PJI and optimize TKA care, it is crucial to assess preoperative and perioperative risk factors. The evaluation facilitates the development of targeted strategies to avoid PJI and enables risk prediction on an individual case basis. While preoperative assessment and perioperative management are performed to screen for and optimize risk factors in clinical practice, the interaction and relative contribution of these risk factors to PJI are not well described. An integrated model that can combinedly assess patients' risk factors and make relevant risk predictions would hence be useful in counseling patients and supporting clinical decision-making, especially in assessing an individual's subsequent risk of PJI.

Machine learning (ML) is a subset of artificial intelligence that learns from historical data to identify patterns and make logical decisions with improvement along the learning process [7]. It is rapidly gaining prominence in the orthopedics field due to its promising potential in discerning underlying patterns within data in complex and non-linear tasks [8]. It has demonstrated its utility in the analysis of PJI [9], for example, the prediction of the failure rate of debridement, antibiotics, and implant retention for PJI and the prediction of recurrent PJI following revision TKA [10, 11]. These prediction models, however, did not encounter significant data imbalance, which occurs in the scenario of PJI prediction among the primary TKA population as the number of PJI (minority class) exists in small numbers among the whole primary TKA population. Such data imbalance complicates pattern detection for many standard ML classifiers, which often assume balanced training data and struggle with the inadequate number of positive samples in this context [12]. In this study, we explored the use of advanced ML algorithms, such as the balanced random forest algorithm, to address the challenges posed by the imbalanced data set. In each iteration of this classifier, a bootstrap sample was drawn from the minority class by random sampling with replacement. Then, the majority class was downsampled to form a balanced bootstrap sample for inducing a classification tree. The predictions of all trees were aggregated to reach the final prediction [13]. The imbalance problem could possibly be solved by using balanced subsamples (Fig. 1).

Recent attempts to use ML or statistical methods to stratify the risk of PJI following primary TKA are constrained in that they are developed based on Caucasian and Western world data [14–19]. However, these models may not be relevant to Asians due to geographic or ethnic differences in the occurrence of the disease. For example, the reported prevalence of rheumatoid arthritis in China is lower than that in the United States and Europe [20]. This divergence in rheumatoid arthritis prevalence would possibly influence its significance as a PJI risk factor and its relative contribution to the overall number of PJIs in Asians. Similarly, many other geographic or ethnic differences may also contribute to the inadaptability of the previous ML models in Asia, for example, the difference in the definition of obesity with a lower body mass index (BMI) cut-off, and older age at which TKA is performed, to name but a few [21, 22]. Due to this shift in demographics, the limited factors employed by previous models may not fit well in the Asian population.

Considering these difficulties and limitations, this study aimed to develop a comprehensive ML model that provides an individualized and systematic prediction of PJI risk following primary TKA in Asia, identifying predictors while navigating through the imbalanced classification task of PJI with improved ML algorithms. A broad range of potential risk factors were examined to identify the most relevant ones and combined into one unifying model.

Methods

Patient cohort

This study was approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (IRB/ REC reference number: UW23-328) and was conducted in accordance with the Declaration of Helsinki. A retrospective analysis of

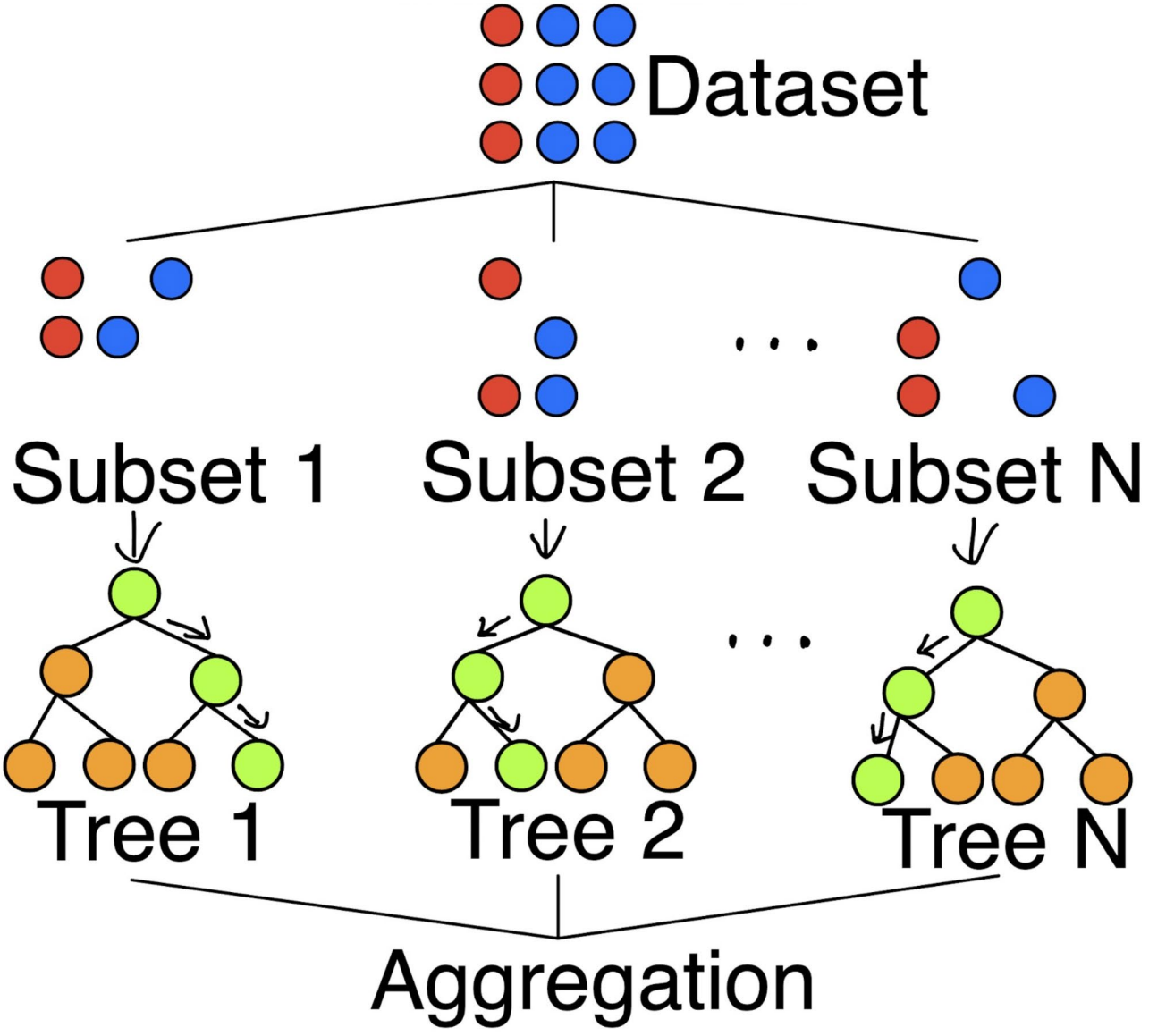

Fig. 1 A diagram illustrating the balanced random forest algorithm's concept. Red dots represent the minority class, whereas blue dots represent the majority class

patients presented to Queen Mary Hospital, a tertiary and quaternary referral academic center in Hong Kong, between 1998 and 2021 was conducted. In this cohort study, patients receiving primary TKA were studied with a mean follow-up time of 8.6 ± 3.9 years, examining the association between the exposure to different risk factors or protective factors and the risk of developing PJI after TKA. Data retrieval was performed using the electronic patient record (ePR) system of Hong Kong public hospitals and its accompanying retrieval system, the Clinical Data Analysis and Reporting System (CDARS, Hospital Authority, Hong Kong). TKA was identified with *the* ICD-9-CM procedure codes 81.41 and 81.54 (total knee replacement), and further supplemented by detailed patient record analysis. The inclusion criteria were ipsilateral, staged, or simultaneous bilateral TKAs, with complete operations documentation and at least one year of clinical follow-up, resulting in 3710 cases of TKA. Non-primary (e.g. revision) and partial (e.g. unicompartmental, bicompartmental) knee arthroplasty procedures were excluded. The final cohort of 3,483 cases of TKA remained for the model development. The cohort selection process is outlined in Fig. 2.

Among the final cohort, cases were divided into the infected class (PJI developed following TKA) and non-infected class (no evidence of PJI following TKA). Possible PJI cases were identified via ICD-9-CM and ICD-10-CM diagnosis codes 996.66 and T84.5 (infection

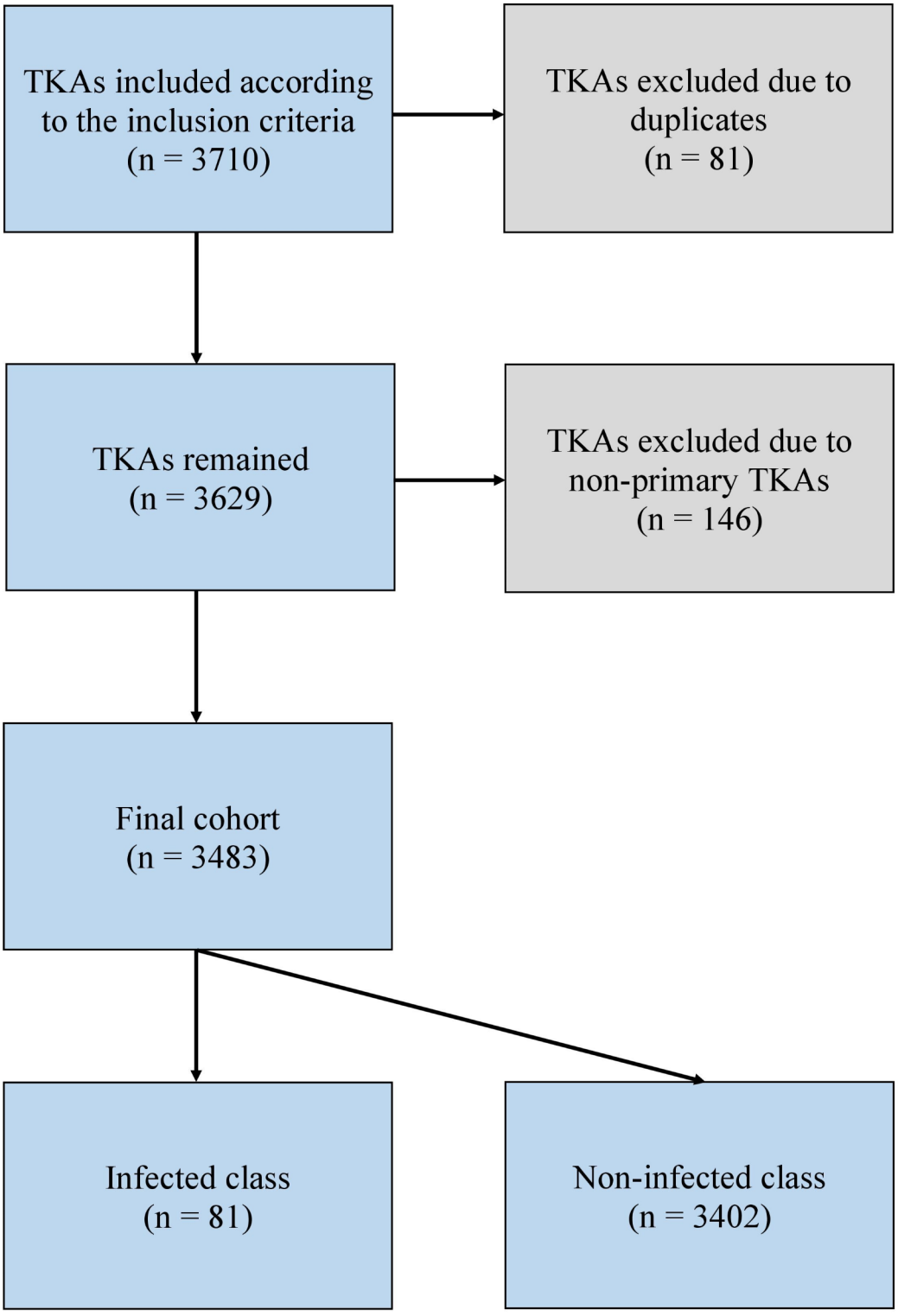

Fig. 2 A flowchart illustrating the cohort selection process

and inflammatory reaction due to internal joint prosthesis), as well as V54 and Z47 (other orthopedic aftercare/follow-up care). These cases were retrieved by manual chart review and assessed according to the 2013 International Consensus Meeting criteria, which is a modified definition from the Musculoskeletal Infection Society criteria of PJI, to confirm the PJI status [23]. Cases with confirmed PJI status were classified as the infected class, while cases without confirmed PJI status were classified as the non-infected class.

Feature collection and preprocessing

60 features, including patient demographics (e.g. age, gender, ethnicity, and American Society of Anesthesiology (ASA) score), operation-related variables (e.g. operative time, anesthesia type, and indication for operation), laboratory findings (e.g. preoperative albumin, hemoglobin, and the international normalized ratio), and comorbidities (e.g. diabetes mellitus (DM), anemia, and septic arthritis) were retrieved from the aforementioned hospital systems (ePR and CDARS). The characteristics of the cohort were summarized in Table 1.

We used Python (version 3.11.0; Python Software Foundation, Wilmington, DE, USA) along with the libraries Scikit-learn (version 1.2.2), NumPy (version 1.21.5), and Pandas (version 1.4.4) for data preprocessing, statistical analysis, model development, and evaluation. Missing data was imputed with the mean for continuous variables and mode for discrete variables (Supplementary Table 1). The effect of outliers was reduced by Winsorization. MaxAbsScaler was applied to rescale continuous variables by their maximum absolute value.

For feature selection, we conducted a univariate analysis using logistic regression with an alpha level of 0.05, selecting features with p-values < 0.05. We then performed an iterative multivariate analysis using logistic regression, where in each iteration, the feature with the maximum p-value was removed if it exceeded 0.1. The final set of predictors obtained from the iterative multivariate analysis had p-values < 0.05, as summarized in Table 1.

Model development

In this research, we employed five supervised ML algorithms: (1) balanced random forest, (2) gradient boosting machine, (3) logistic regression, (4) Gaussian Naïve Bayes, and (5) support vector machine to predict the risk of PJI following primary TKA. These five candidate algorithms were inspired by prior studies that demonstrated strong discriminative capability in prediction tasks. A stratified 10-fold cross-validation method was used to train all five models on the same dataset with hyperparameters tuned and assess their performances.

Performance evaluation

All models' performances were evaluated by several metrics in each iteration: the area under the receiver operating characteristic curve (AUC), balanced accuracy, sensitivity, specificity, F1 score, Brier score, calibration slope, and calibration intercept. For discrimination, AUC was mainly used to evaluate the model. It is the area under a probability curve plotted with true positive rates against false positive rates, representing the degree of predictive capability. The AUC ranges from 0 to 1. A model with an AUC higher than 0.7 is considered to have good performance with clinical significance, while an AUC of 1 is considered perfect [24]. For calibration, a calibration plot with calibration slope and intercept was used for evaluation. The ideal model has a calibration slope of 1 and a calibration intercept of 0 [25]. For the rest of the metrics, all metrics range from 0 to 1 with a higher score indicating better performance, except for the Brier score, which reflects better performance when approaching 0. Among the metrics, we put more emphasis on the AUC and balanced accuracy during performance assessment due to the imbalanced nature of the classification task. An average of the scores in each iteration resulted in the final performance of the model. The model with the best performance among the five trained models was selected as our final model.

Model interpretability

In attempting to increase the transparency and interpretability of the final prediction model, both global and local explanations were provided. A global explanation was delivered through SHapley Additive exPlanations (SHAP) summary plots, and local explanations for individual TKA cases were delivered through SHAP waterfall plots.

Results

Cohort characteristics

A total of 3,483 cases of TKA were studied. In the cohort, 81 (2.3%) TKAs developed PJI. Most of the patients were Chinese (98.2%) and female (73%). The mean age was 70.4 ± 10.0 years, with a mean BMI of 27.8 ± 4.6 kg/m^2. A significant proportion of the cohort was ASA class 2 (57.5%) and ASA class 3 (30.9%). The mean duration of TKA operations was 116.6 ± 75.8 min, with spinal anesthesia being the most employed anesthesia type (48.9%), followed by general anesthesia (35.4%) as the second most utilized anesthesia option. Primary osteoarthritis was the major indication for operation (92.6%). The mean preoperative hemoglobin level was 12.8 ± 1.4 g/dL. Certain comorbidities showed a notably higher prevalence among our cohort: hypertension (71.3%), DM (41.3%), and hyperlipidemia (11.3%). The other comorbidities had substantially lower prevalence in the study group (Table 1).

Table 1 Characteristics of the cohort

Characteristics	Primary total knee arthroplasty cases (*N* = 3483)	*P*-values	
		Univariate analysis	Multivariate analysis (Final iteration)
Patient demographics			
Age (years)	70.4 ± 10.0	**< 0.001**	
Male gender	939 (27.0%)	**< 0.001**	**< 0.001**
Ethnicity	Chinese 3421 (98.2%); Other Asians 33 (0.9%); Non-Asians 29 (0.8%)	0.436	
American Society of Anesthesiology score	ASA 1, 146 (4.2%); ASA 2, 2002 (57.5%); ASA 3, 1077 (30.9%); ASA 4, 3 (0.1%); Missing, 255 (7.3%)	ASA > 2: **0.016**	ASA > 2: **0.028**
Body mass index (kg/m^2)	27.8 ± 4.6	0.148	
Operation-related variables			
Laterality	Left 1458 (41.9%); Right 1457 (41.8%); Bilateral 568 (16.3%)	0.229	
Operative time (minutes)	116.6 ± 75.8	**< 0.001**	**0.018**
Preoperative length of stay (days)	1.2 ± 1.8	0.081	
Anesthesia type			
General anesthesia	1233 (35.4%)	**0.002**	
Spinal anesthesia	1703 (48.9%)	**0.003**	**0.022**
Combined-spinal epidural anesthesia	428 (12.3%)	0.987	
Epidural anesthesia	22 (0.6%)	0.497	
Other	97 (2.8%)	0.861	
Indication for operation			
Primary osteoarthritis	3226 (92.6%)	**< 0.001**	
Rheumatoid arthritis	114 (3.3%)	0.826	
Neoplasm	56 (1.6%)	**< 0.001**	
Secondary osteoarthritis	89 (2.6%)	**0.008**	
Other	32 (0.9%)	0.157	
Laboratory findings			
Preoperative albumin (g/L)	42.3 ± 3.3	**0.014**	
Preoperative hemoglobin (g/dL)	12.8 ± 1.4	0.058	
Preoperative international normalized ratio	1.0 ± 0.1	**0.028**	
Preoperative absolute lymphocyte count (×10^9/L)	1.8 ± 0.6	0.102	
Comorbidities			
Diabetes mellitus	1438 (41.3%)	0.742	
Uncomplicated	1339 (38.4%)	0.469	
Complicated	99 (2.8%)	0.258	
Hypertension	2482 (71.3%)	0.157	
Hyperlipidemia	392 (11.3%)	0.692	
Anemia	215 (6.2%)	**0.002**	**0.023**
Thrombocytopenia	16 (0.5%)	**0.018**	
Agranulocytosis	12 (0.3%)	**0.006**	
Neoplasm	339 (9.7%)	0.056	
Non-malignant	206 (5.9%)	0.565	
Malignant	133 (3.8%)	**0.027**	
Renal disease	109 (3.1%)	0.995	
Renal failure	60 (1.7%)	0.999	
Renal impairment	49 (1.4%)	1.000	
Liver disease	90 (2.6%)	0.523	
Heart failure	107 (3.1%)	0.751	
Chronic ischemic heart disease	251 (7.2%)	0.716	
Atrial fibrillation	137 (3.9%)	0.914	
Dysrhythmia	53 (1.5%)	0.831	

Table 1 (continued)

Characteristics	Primary total knee arthroplasty cases (*N* = 3483)	*P*-values	
		Univariate analysis	Multivariate analysis (Final iteration)
Myocardial infarction	17 (0.5%)	1.000	
Angina pectoris	36 (1.0%)	0.857	
Atherosclerosis	38 (1.1%)	0.241	
Peripheral vascular disease	6 (0.2%)	0.052	
Asthma	90 (2.6%)	0.450	
Chronic obstructive pulmonary disease	38 (1.1%)	0.900	
Cerebrovascular disease	85 (2.4%)	0.460	
Organic brain syndrome	14 (0.4%)	0.258	
Epilepsy	13 (0.4%)	0.976	
Hemiplegia	10 (0.3%)	0.996	
Alzheimer's disease	12 (0.3%)	0.999	
Dementia	7 (0.2%)	0.999	
Depression	110 (3.2%)	0.359	
Psychosis	79 (2.3%)	0.385	
Anxiety	54 (1.6%)	0.999	
Tuberculosis	156 (4.5%)	0.384	
Rheumatoid disease	181 (5.2%)	0.689	
Septic arthritis	21 (0.6%)	**0.002**	**0.030**
Systemic lupus erythematosus	11 (0.3%)	0.995	

*Bolded *p*-values are $p < 0.05$

Table 2 The average discrimination and calibration performance of the five models across all ten folds of cross-validation

Metrics	Balanced Random Forest	Gradient Boosting Machine	Logistic Regression	Gaussian Naïve Bayes	Support Vector Machine
AUC	0.963	0.931	0.728	0.719	0.701
Balanced Accuracy	0.920	0.844	0.654	0.655	0.644
Sensitivity	0.938	0.889	0.744	0.708	0.572
Specificity	0.902	0.800	0.564	0.602	0.717
F1 score	0.314	0.174	0.074	0.076	0.084
Brier score	0.097	0.198	0.432	0.395	0.287
Calibration slope	1.362	2.431	1.198	-0.854	1.264
Calibration intercept	-0.009	-0.034	-0.005	0.043	-0.007

AUC = Area under the receiver operating characteristic curve

Significant predictors

By univariate analysis, 16 features ($p < 0.05$) were selected for an iterative multivariate analysis. Six significant features ($p < 0.05$) were selected as our final set of predictors: operative time, male gender, ASA > 2, spinal anesthesia, history of anemia, and history of septic arthritis (Table 1).

Model performance

The AUC of the models ranged from 0.963 to 0.701. The balanced random forest model achieved the highest AUC, showing excellent discriminative capability among the five models. The balanced accuracy of the models ranged from 0.920 for the balanced random forest model to 0.644 for the support vector machine model. The sensitivity ranged from 0.938 to 0.572, with the highest sensitivity achieved by the balanced random forest model. The specificity ranged from 0.902 for the balanced random forest model to 0.564 for the logistic regression model (Table 2). Overall, the balanced random forest model demonstrated the best discrimination performance among all five models, showing stable and robust performances across all ten folds of cross-validation (Fig. 3), hence being selected as our final model. In terms of calibration, the balanced random forest model had a calibration slope of 1.362 and a calibration intercept of -0.009, suggesting a slightly moderate risk estimation (Fig. 4; Table 2).

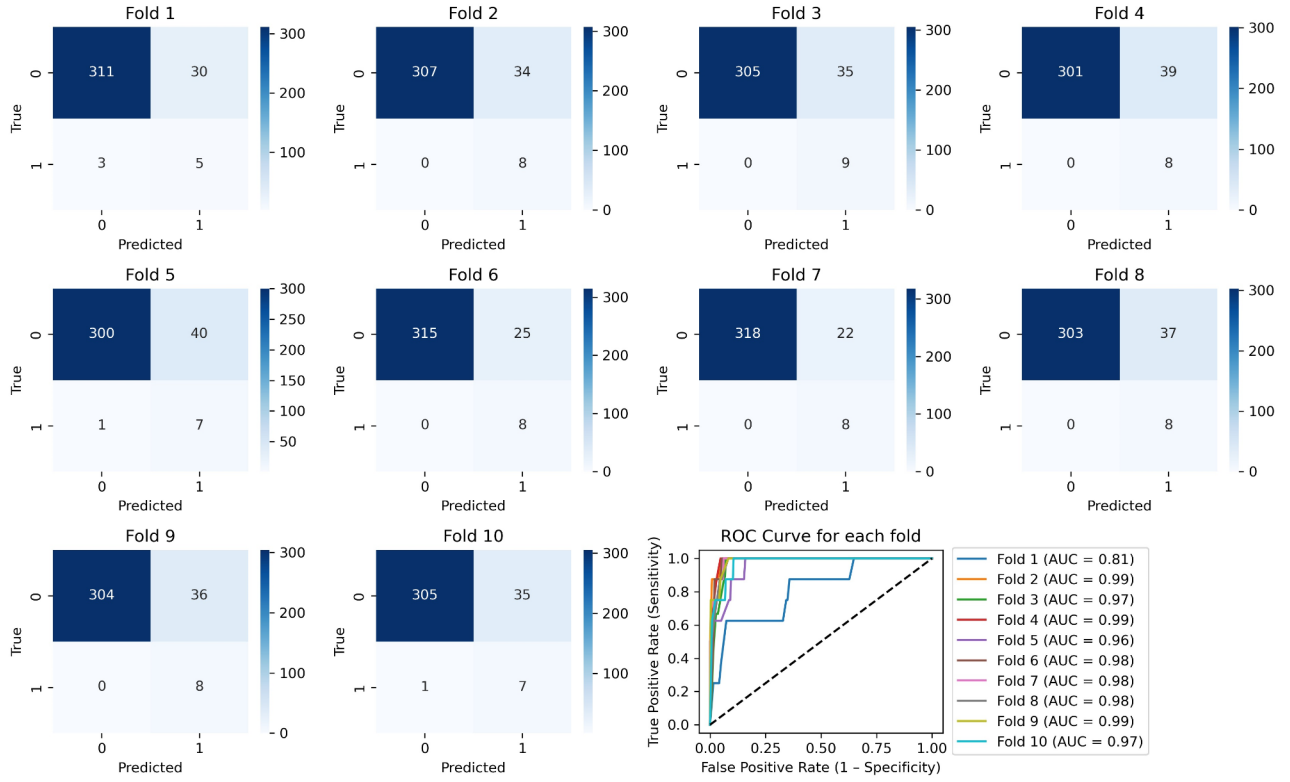

Fig. 3 Performance of the balanced random forest model, showing the confusion matrices and receiver operating characteristic curves for all ten folds of cross-validation. In the confusion matrices, TKA with PJI was classified as Class 1, and TKA without PJI was classified as Class 0

Model interpretability

The SHAP summary plots provided a global explanation of the model decision, visualizing the feature importance of each feature (Figs. 5 and 6). In descending levels of significance, the model used operative time (OR, 9.07; 95% CI, 1.47–56.14; $p = 0.018$), male gender (OR, 3.11; 95% CI, 1.97–4.90; $p < 0.001$), ASA > 2 (OR, 1.68; 95% CI, 1.06–2.67; $p = 0.028$), spinal anesthesia (OR, 0.55; 95% CI, 0.33–0.92; $p = 0.022$), history of anemia (OR, 2.17; 95% CI, 1.11–4.24; $p = 0.023$), and history of septic arthritis (OR, 4.35; 95% CI, 1.15–16.41; $p = 0.030$) as predictors. In general, all predictors were risk factors for PJI except spinal anesthesia, which was a protective factor.

We presented two local SHAP waterfall plots for a high-risk and a low-risk patient respectively, illustrating the individual patient-level explanations for the final model predictions (Figs. 7 and 8). The 8th TKA patient in our cohort was a low-risk patient with a predicted PJI risk of 33.3% (Fig. 7). In this individual case, being a male was the most significant risk factor, whereas spinal anesthesia administration and TKA lasting for 70 min (0.08 after rescaling) were the two most important protective factors. In another example, the 32nd TKA patient in our cohort was a high-risk patient with a predicted PJI risk of 93.9% (Fig. 8). In this individual case, TKA lasting for 117 min (0.134 after rescaling) and non-spinal anesthesia administration were the two most important risk factors, whereas being a female was the most significant protective factor.

Discussion

Model performance

This study developed the first PJI prediction model based on the Chinese population, marking the first such model in Asia. Recent studies have attempted to stratify PJI risk following primary TKA by statistical or ML-based methods with the use of Western world data [15, 16, 19, 26] (Table 3). In particular, our ultimate ML model employing the balanced random forest algorithm exhibited superior performance compared to prior models, demonstrating its exceptional discriminative capability and potential for identifying patients with an elevated PJI risk. After extensive investigation of potential predictors to ensure the comprehensiveness of our study, our model concentrated on six selected predictors that are most relevant to the Chinese population. This design allows for easy adaptation by other hospitals, as it necessitates minimal data variable collection.

Among the five candidate algorithms, our study demonstrated that the balanced random forest algorithm stands out as an optimal solution for addressing the imbalanced classification problem in predicting PJI

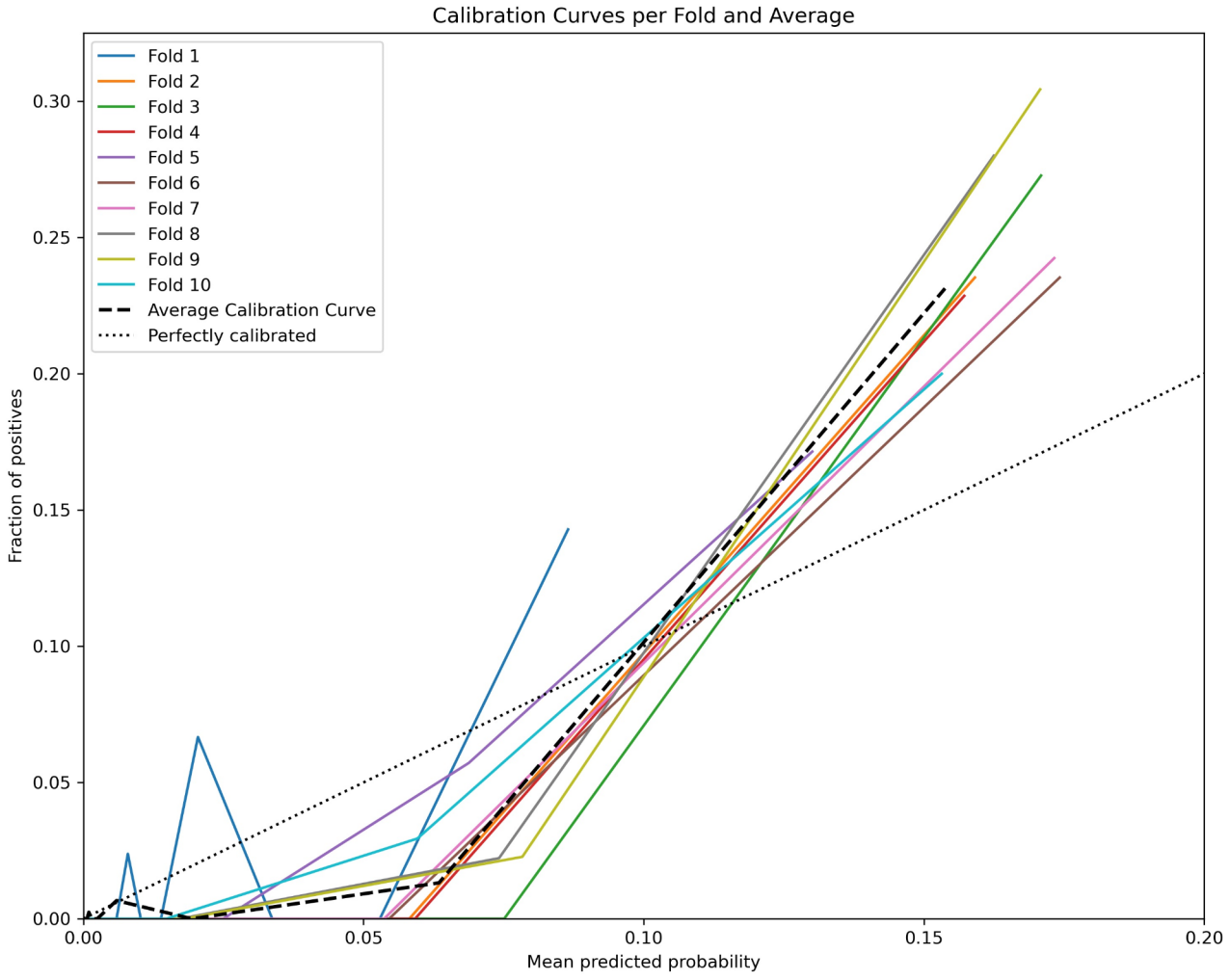

Fig. 4 A calibration plot of the balanced random forest model

following primary TKA. This algorithm was specifically designed to address the imbalance issue by utilizing balanced subsamples.

Significant predictors

The existing literature supports the six selected predictors. In this study, operative time emerged as the strongest predictor with a longer duration found to be associated with higher PJI risk. With a 15-minute increase in operative time, an 18% (95% CI, 11–26) increased PJI risk was found by a retrospective study of 11,840 primary TKAs performed between 2014 and 2017 [27], and a 9% (95% CI, 4–13) increased deep wound infection risk was found by a registry-based study of 56,216 primary TKAs performed between 2001 and 2009 [28]. The underlying mechanism may be multifactorial. As a result of a longer operative time, an open TKA incision wound is more likely to be contaminated by airborne bacteria [29, 30], and tissue desiccation around incisions is more likely to occur [31], which delays wound healing and increases infection risk [32]. A longer tourniquet duration may also prolong wound hypoxia and increase infection risk [33].

In accordance with prior works, we found that the male gender played the second most significant role in predicting PJI risk following primary TKA. Male gender was revealed to be a significant PJI risk factor in an analysis of 64,566 TKAs from the New Zealand Joint Registry performed between 1999 and 2012 (OR, 1.84; 95% CI, 1.24–2.73) [34] and an analysis of 56,216 TKAs from an American registry performed from 2001 to 2009 (HR, 1.89) [35]. The underlying mechanism is controversial. Male gender itself and its hormonal factors might or might not be a direct risk factor for PJI, but it is undoubtedly associated with multiple behavioral factors that remain unrecorded in an electronic health system. For example, males in Asian populations, traditionally assuming the role of primary breadwinners, tend to adopt

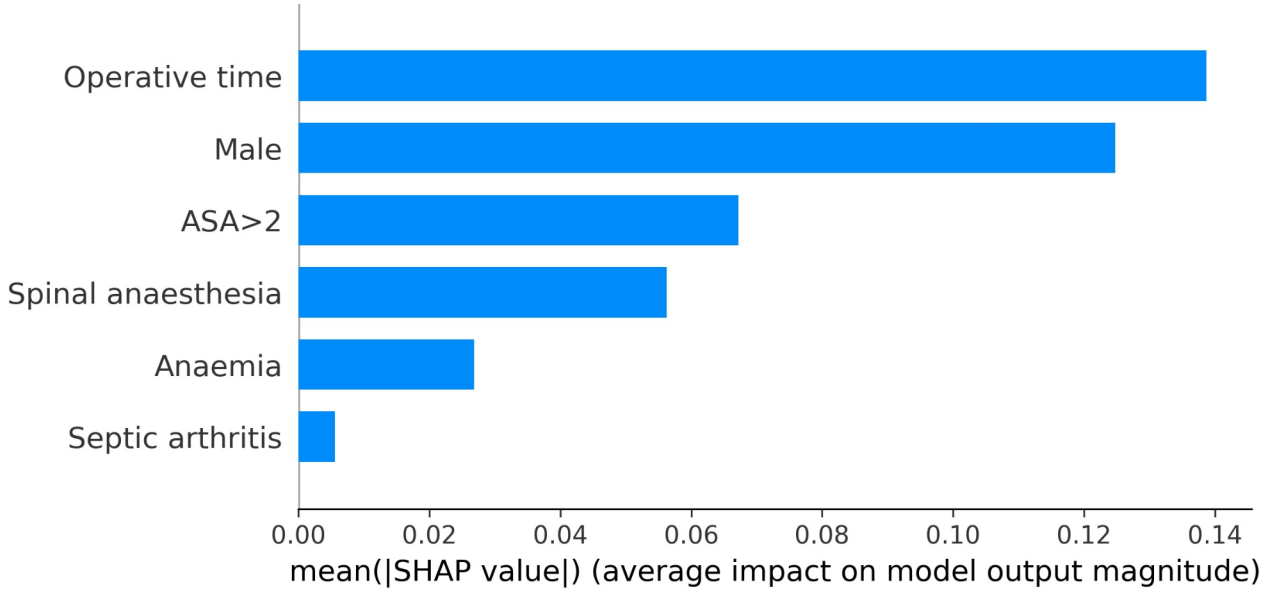

Fig. 5 A SHAP summary bar plot, ranking the average importance level of each predictor on the model prediction in descending order

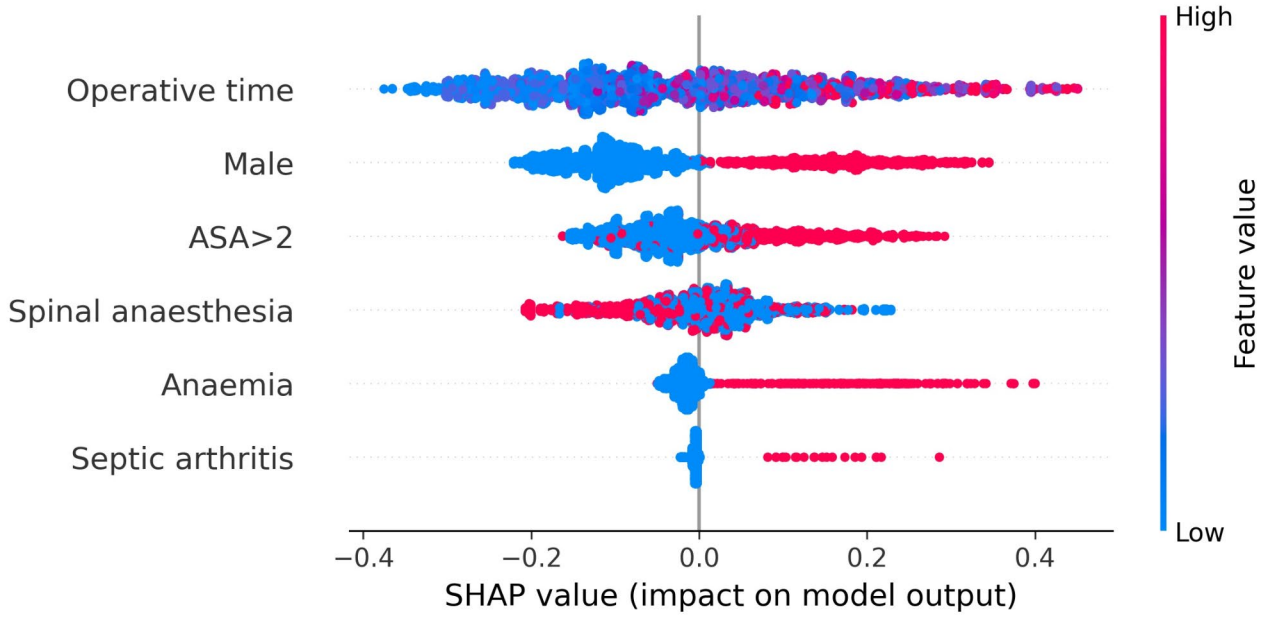

Fig. 6 A SHAP summary dot plot, showing the impact of each predictor on the model output. A positive SHAP value represents an increase in PJI risk, whereas a negative SHAP value represents a decrease in PJI risk. For continuous feature (operative time), the feature value ranges from red to purple to blue, with red dots indicating a higher value of the feature, and blue dots indicating a lower value. For discrete features (the rest of the predictors), the feature value is represented by red and blue dots, with red dots indicating the presence of the feature, and blue dots indicating the absence of the feature

a more sedentary lifestyle, encounter greater exposure to occupational environments with suboptimal hygiene like construction sites, and expose more to alcohol and smoking, etc. These factors tend to be more prevalent in males than females and were not accounted for in this study [36, 37].

ASA score was the third significant predictor of PJI risk following primary TKA. This score, evaluated by an anesthetist preoperatively, serves as an indicator of the patient's physical status and plays a role in estimating comorbidity and preoperative risk. ASA > 2 indicates the presence of significant systemic disease [38]. We found that an ASA score > 2 was associated with an increased risk of PJI following TKA. This observation was consistent with prior studies. In a meta-analysis, ASA score > 2 was concluded to be a high-risk factor for PJI (OR, 2.06;

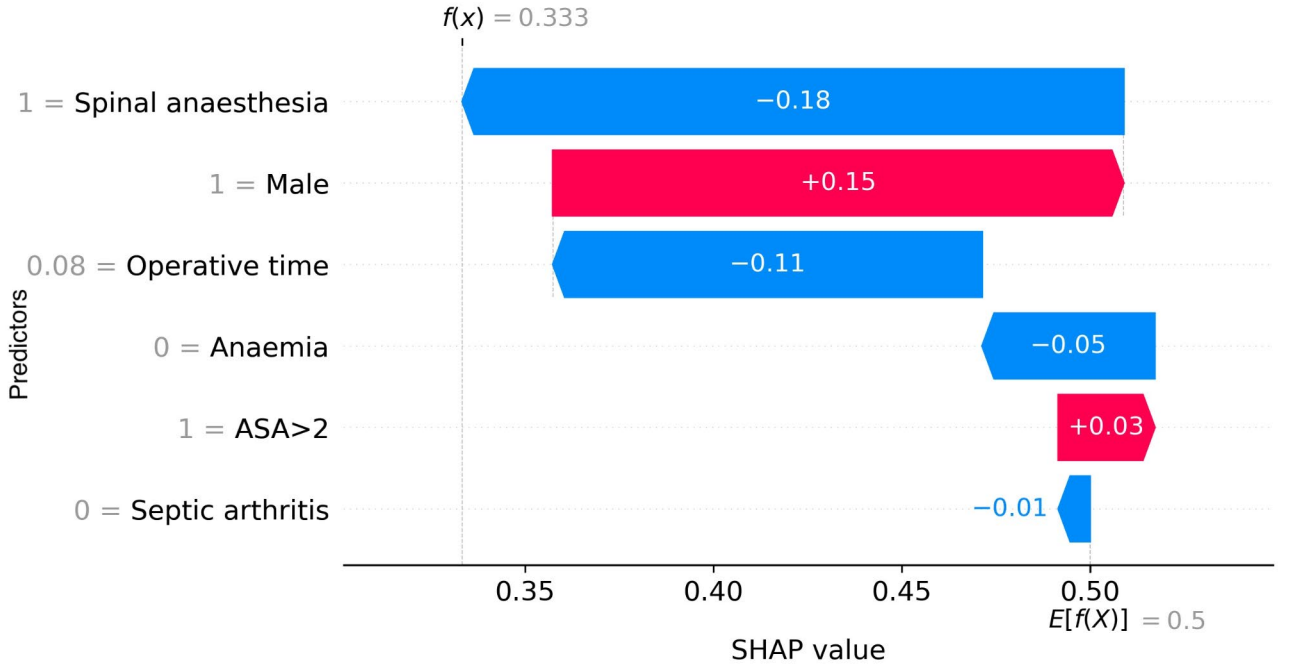

Fig. 7 A local SHAP waterfall plot of the 8th TKA patient. The number on the left of each predictor is the input value, with 0 representing the absence of predictors, and 1 representing the presence of predictors. For continuous feature, the input value of the operative time was rescaled from the original duration. The impact of predictors on the model output was ranked in descending order of significance, with spinal anesthesia being the strongest predictor and septic arthritis being the least important predictor in this TKA case. Red arrows represent an increase in PJI risk, whereas blue arrows represent a decrease in PJI risk. The sum of the base value (E[f(X)] = 0.5) and SHAP values of all predictors is the predicted probability (f(x) = 0.333) which is less than the base value. Hence, he was classified as class 0 with a low PJI risk of 33.3%

95% CI, 1.77–2.39) [38]. In another multivariable analysis, ASA class 2 was found to have a 1.2 lower hazard ratio than ASA > 2 for PJI following total hip arthroplasty [39]. ASA score may also serve as a comprehensive indicator encompassing the influence of age-related comorbidities on PJI risk. The inclusion of ASA as a risk factor addresses the exclusion of age as a direct risk factor in our model, as multiple studies have indicated that age alone does not exhibit significant predictive value for PJI risk after adjusting for confounding variables [40–42]. These findings highlight the importance of considering broader health status factors, which ASA score reflects, when assessing the risk of PJI following TKA.

Apart from risk factors, spinal anesthesia was the only protective factor identified among all predictors. Scholten et al. [43] reported a lower association between spinal anesthesia and early PJI following TKA in comparison to general anesthesia, which has an odds ratio of 2.0 (95% CI, 1.0–3.7) relative to spinal anesthesia. Similar results were obtained from several studies [44–46]. Although the underlying mechanism is not fully understood, the association between spinal anesthesia and less blood loss, fewer blood transfusions required, and less incidence of hyperglycemia are some possible reasons for its protective nature against PJI because these factors suppress immunity [43, 47, 48]. Besides, due to the limited operative time allowed by spinal anesthesia, it may also imply that the TKA operation went smoothly without unexpected complications. Therefore, it was also in echo with our above finding that longer operative time was a risk factor for PJI.

While the identified predictors are familiar to orthopedic surgeons, the interplay between these factors and their relative contribution to PJI remains understudied. Our model represents a novel approach by integrating these significant parameters into a unified risk prediction tool. By accounting for the complex interactions among predictors, our model offers rapid and personalized predictions tailored to individual patients' cases.

Among the selected predictors, three of them (operative time, spinal anesthesia, and history of anemia) were potentially modifiable. By applying our model to real clinical settings, we could identify high-risk patients and their modifiable risk factors preoperatively. For example, in a male patient with an ASA score above 2, surgeons should pay attention to the possibly heightened risk of PJI. In addition to standard preoperative health assessment, they can further optimize the patient by paying particular attention to the reduction of operative time, choice of anesthesia (spinal anesthesia if possible), and necessity of correcting anemia for a period before TKA. Patients could also make a thorough preoperative

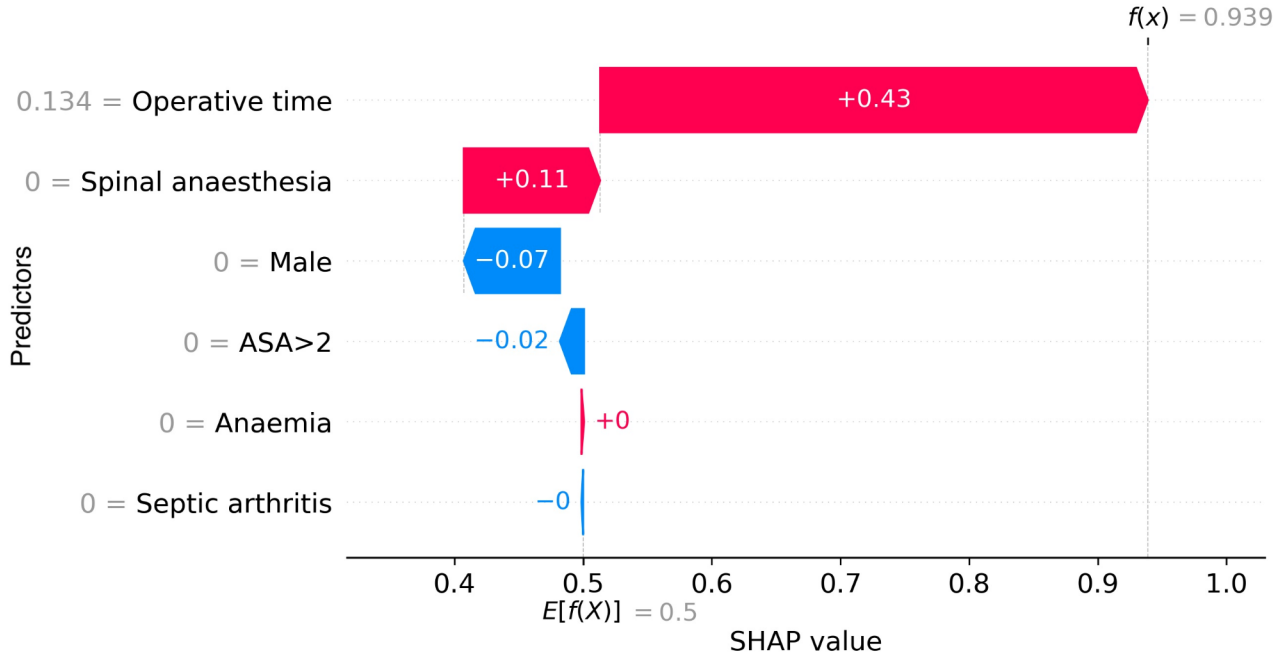

Fig. 8 A local SHAP waterfall plot of the 32nd TKA patient. The number on the left of each predictor is the input value, with 0 representing the absence of predictors, and 1 representing the presence of predictors. For continuous feature, the input value of the operative time was rescaled from the original duration. The impact of predictors on the model output was ranked in descending order of significance, with operative time being the strongest predictor and septic arthritis being the least important predictor in this TKA case. Red arrows represent an increase in PJI risk, whereas blue arrows represent a decrease in PJI risk. The sum of the base value (E[f(X)] = 0.5) and SHAP values of all predictors is the predicted probability (f(x) = 0.939) which is higher than the base value. Hence, she was classified as class 1 with a high PJI risk of 93.9%

Table 3 Model performances of prior studies and this study

Author	Year	Algorithm	AUC
Yeo, I. [19]	2023	Artificial neural network	0.840
Tan, T.L. [16]	2018	Logistic regression	0.830
Bilimoria, K.Y. [26]	2013	Logistic regression	0.817
Espindola, R. [15]	2022	Logistic regression	0.730
This study	/	Balanced Random Forest	0.963

treatment decision by weighing TKA risks against benefits through the analysis of their non-modifiable risk factors. For instance, a male patient, with a history of septic arthritis, and a high ASA score, should be informed of the relatively high risk of PJI. If his knee arthritis is relatively tolerable or he is in low demand, then TKA may not be in his best interest. The model could be applied not only in the preoperative period but also postoperatively, where it may help identify high-risk patients who may need enhanced monitoring and screening for potential PJI development, allowing earlier diagnosis and, consequently, earlier treatment. With this model, practice changes are therefore possible to further reduce the risk of PJI, a shared decision can also be reached between surgeons and patients who are now better informed of the potential risk, and high-risk patients could be closely monitored for potential PJI development. Potentially, this could translate into overall better clinical outcomes of TKA and reduce clinical expenditure on managing PJI.

Limitations

Our study has its limitations. Machine learning's 'black box' nature limits transparency in its predictions, which can hinder user understanding [49]. Although we employed SHAP values to enhance interpretability, they only partially clarified the model's decision-making [50]. Future work could incorporate additional methods, alongside SHAP, for a deeper insight into the model. Furthermore, we studied a long period, during which practices may have varied and data documentation could be incomplete. Our study also conducted in a highly specialized academic center that emphasizes preoperative health optimization for TKA. This might introduce confounding temporal factors and biases with some known risk factors, like DM, BMI, etc., being relatively treated and downplayed, hence not reaching statistical significance in our study [51–53]. Future analysis on temporal effects and these risk factors is encouraged. In addition, the highly imbalanced dataset may affect the calibration performance [25], and contribute to a high number of false positives. However, since the primary objective is to flag high-risk patients for preoperative optimization, the model prioritized sensitivity even at the cost of

false positives. Further multi-center studies with a larger cohort could allow external validation and recalibration to enhance model's generalizability and accuracy. Nevertheless, our model provides a solid foundation for further investigation and collaboration among Asian centers. Moreover, to align with existing studies using a one-year follow-up to capture most high-risk period PJIs, we adopted a one-year threshold [10, 15, 54, 55]. While this may limit detection of late-onset PJIs, our extended mean follow-up of 8.6 years mitigates this limitation. Further studies with longer minimum follow-up periods could provide further insights into late-onset PJI risk. Future work can expand on this ML architecture for predicting not only PJI but also other arthroplasty complications, particularly rare complications with imbalanced datasets, such as implant fracture, instability, and bone cement implantation syndrome [56–58], since we demonstrated the feasibility of applying this ML architecture to tackle imbalanced prediction tasks.

Conclusion

We developed the first ML model for individualized prediction of PJI following primary TKA in Asia, identifying operative time and gender as the strongest predictors. This study demonstrated the excellent performance and potential of the balanced random forest algorithm in solving imbalanced classification difficulties. Global and local patient-level interpretations were also provided to increase the transparency and interpretability of the model. Potential areas for further preoperative and perioperative optimization were highlighted by the model.

Abbreviations

ASA	American Society of Anesthesiology
AUC	Area under the receiver operating characteristic curve
BMI	Body mass index
CDARS	Clinical Data Analysis and Reporting System
DM	Diabetes mellitus
ePR	Electronic patient record
ML	Machine learning
PJI	Periprosthetic joint infection
SHAP	SHapley Additive exPlanations
TKA	Total knee arthroplasty

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12891-025-08296-6.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions

YYC: conceptualization, data acquisition, model development, statistical analysis, manuscript drafting. CMLL: manuscript editing, revision comments provision. TSJ: model development, manuscript editing. CYW and JZ: model development. AC, MHL, KCTL, MHC, HF, KYC: revision comments & expert advice provision. PKC: conceptualization, supervision, revision comments provision. All authors read and approved the final manuscript.

Funding

Not applicable.

Data availability

The datasets analyzed during the current study are available from the corresponding authors upon reasonable request.

Declarations

Human ethics and consent to participate

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (IRB/ REC reference number: UW23-328). Informed consent to participate was waived by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster due to the retrospective nature of this study.

Consent for publication

Not applicable

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Orthopedics and Traumatology, The University of Hong Kong, Hong Kong SAR, China
[2]Department of Biomedical Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China
[3]Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong SAR, China
[4]Department of Orthopedics and Traumatology, Queen Mary Hospital, Hong Kong SAR, China

Received: 1 January 2024 / Accepted: 6 January 2025
Published online: 11 March 2025

References

1. Matsuoka H, Nanmo H, Nojiri S, Nagao M, Nishizaki Y. Projected numbers of knee and hip arthroplasties up to the year 2030 in Japan. J Orthop Sci. 2023;28(1):161–6.
2. Bozic KJ, Kurtz SM, Lau E, Ong K, Chiu V, Vail TP, et al. The epidemiology of revision total knee arthroplasty in the United States. Clin Orthop Relat Res. 2010;468(1):45–51.
3. Ahmed SS, Haddad FS. Prosthetic joint infection. Bone Joint Res. 2019;8(11):570–2.
4. Zmistowski B, Karam JA, Durinka JB, Casper DS, Parvizi J. Periprosthetic joint infection increases the risk of one-year mortality. J Bone Joint Surg Am. 2013;95(24):2177–84.
5. Premkumar A, Kolin DA, Farley KX, Wilson JM, McLawhorn AS, Cross MB, et al. Projected Economic Burden of Periprosthetic Joint Infection of the hip and knee in the United States. J Arthroplasty. 2021;36(5):1484–e93.
6. Premkumar A, Kolin DA, Farley KX, Wilson JM, McLawhorn AS, Cross MB, et al. Projected Economic Burden of Periprosthetic Joint Infection of the hip and knee in the United States. J Arthroplasty. 2021;36(5):1484–9. e3.
7. Bini SA, Artificial Intelligence M, Learning. Deep learning, and Cognitive Computing: what do these terms Mean and how will they Impact Health Care? J Arthroplasty. 2018;33(8):2358–61.
8. Lau LCM, Chui ECS, Man GCW, Xin Y, Ho KKW, Mak KKK, et al. A novel image-based machine learning model with superior accuracy and predictability for knee arthroplasty loosening detection and clinical decision making. J Orthop Translat. 2022;36:177–83.
9. Chong YY, Chan PK, Chan VWK, Cheung A, Luk MH, Cheung MH, 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.

10. Shohat N, Goswami K, Tan TL, Yayac M, Soriano A, Sousa R et al. 2020 Frank Stinchfield Award: identifying who will fail following irrigation and debridement for prosthetic joint infection. Bone Joint J. 2020;102-B(7_Supple_B):11 – 9.
11. Klemt C, Laurencin S, Uzosike AC, Burns JC, Costales TG, Yeo I, et al. Machine learning models accurately predict recurrent infection following revision total knee arthroplasty for periprosthetic joint infection. Knee Surg Sports Traumatol Arthrosc. 2022;30(8):2582–90.
12. Ali A, Shamsuddin SM, Ralescu A. Classification with class imbalance problem. Rev. 2015;7:176–204.
13. Chen C, Breiman L. Using Random Forest to learn Imbalanced Data. Berkeley: University of California; 2004.
14. del Toro MD, Peñas C, Conde-Albarracín A, Palomino J, Brun F, Sánchez S, et al. Development and validation of baseline, perioperative and at-discharge predictive models for postsurgical prosthetic joint infection. Clin Microbiol Infect. 2019;25(2):196–202.
15. Espindola R, Vella V, Benito N, Mur I, Tedeschi S, Rossi N, et al. Preoperative and perioperative risk factors, and risk score development for prosthetic joint infection due to Staphylococcus aureus: a multinational matched case-control study. Clin Microbiol Infect. 2022;28(10):1359–66.
16. Tan TL, Maltenfort MG, Chen AF, Shahi A, Higuera CA, Siqueira M, et al. Development and evaluation of a preoperative risk calculator for Periprosthetic Joint Infection Following Total Joint Arthroplasty. JBJS. 2018;100(9):777–85.
17. Bozic KJ, Lau E, Kurtz S, Ong K, Berry DJ. Patient-related risk factors for postoperative mortality and periprosthetic joint infection in medicare patients undergoing TKA. Clin Orthop Relat Res. 2012;470(1):130–7.
18. Inacio MC, Pratt NL, Roughead EE, Graves SE. Predicting infections after total joint arthroplasty using a prescription based Comorbidity measure. J Arthroplasty. 2015;30(10):1692–8.
19. Yeo I, Klemt C, Robinson MG, Esposito JG, Uzosike AC, Kwon YM. The Use of Artificial neural networks for the prediction of Surgical Site infection following TKA. J Knee Surg. 2023;36(6):637–43.
20. Felson DT. Comparing the prevalence of rheumatic diseases in China with the rest of the world. Arthritis Res Ther. 2008;10(1):106.
21. Cheung A, Chan PK, Fu H, Cheung MH, Chan VWK, Yan CH, et al. Total knee arthroplasty is safe for patients aged >/=80 years in Hong Kong. Hong Kong Med J. 2021;27(5):350–4.
22. Low S, Chin MC, Ma S, Heng D, Deurenberg-Yap M. Rationale for redefining obesity in asians. Ann Acad Med Singap. 2009;38(1):66–9.
23. Parvizi J, Gehrke T. Definition of periprosthetic joint infection. J Arthroplasty. 2014;29(7):1331.
24. Trifonova O, Lokhov P, Archakov AI. Metabolic profiling of human blood. Biomeditsinskaya Khim. 2014;60:281–94.
25. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Bossuyt P, et al. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230.
26. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. J Am Coll Surg. 2013;217(5):833–42.e1-3.
27. Anis HK, Sodhi N, Klika AK, Mont MA, Barsoum WK, Higuera CA, et al. Is operative time a predictor for post-operative infection in primary total knee arthroplasty? J Arthroplasty. 2019;34(7s):S331–6.
28. Namba RS, Inacio MC, Paxton EW. Risk factors associated with deep surgical site infections after primary total knee arthroplasty: an analysis of 56,216 knees. J Bone Joint Surg Am. 2013;95(9):775–82.
29. Whyte W, Hodgson R, Tinkler J. The importance of airborne bacterial contamination of wounds. J Hosp Infect. 1982;3(2):123–35.
30. Lidwell OM, Lowbury EJ, Whyte W, Blowers R, Stanley SJ, Lowe D. Airborne contamination of wounds in joint replacement operations: the relationship to sepsis rates. J Hosp Infect. 1983;4(2):111–31.
31. Haridas M, Malangoni MA. Predictive factors for surgical site infection in general surgery. Surgery. 2008;144(4):496–501. discussion – 3.
32. Scott RD. Surgical pearls in total knee arthroplasty: a lifetime of lessons learned. Am J Orthop (Belle Mead NJ). 2016;45(6):384–8.
33. Clarke MT, Longstaff L, Edwards D, Rushton N. Tourniquet-induced wound hypoxia after total knee replacement. J Bone Joint Surg Br. 2001;83(1):40–4.
34. Tayton ER, Frampton C, Hooper GJ, Young SW. The impact of patient and surgical factors on the rate of infection after primary total knee arthroplasty. Bone Joint J. 2016;98–B(3):334–40.
35. Namba RS, Inacio MCS, Paxton EW. Risk factors Associated with Deep Surgical Site infections after primary total knee arthroplasty: an analysis of 56,216 knees. JBJS. 2013;95(9):775–82.
36. Peersman G, Laskin R, Davis J, Peterson M. Infection in total knee replacement: a retrospective review of 6489 total knee replacements. Clin Orthop Relat Res. 2001(392):15–23.
37. Jämsen E, Huhtala H, Puolakka T, Moilanen T. Risk factors for infection after knee arthroplasty. A register-based analysis of 43,149 cases. J Bone Joint Surg Am. 2009;91(1):38–47.
38. Kong L, Cao J, Zhang Y, Ding W, Shen Y. Risk factors for periprosthetic joint infection following primary total hip or knee arthroplasty: a meta-analysis. Int Wound J. 2017;14(3):529–36.
39. Panula VJ, Alakylä KJ, Venäläinen MS, Haapakoski JJ, Eskelinen AP, Manninen MJ, et al. Risk factors for prosthetic joint infections following total hip arthroplasty based on 33,337 hips in the Finnish Arthroplasty Register from 2014 to 2018. Acta Orthop. 2021;92(6):665–72.
40. Chen J, Cui Y, Li X, Miao X, Wen Z, Xue Y, et al. Risk factors for deep infection after total knee arthroplasty: a meta-analysis. Arch Orthop Trauma Surg. 2013;133(5):675–87.
41. Kunutsor SK, Whitehouse MR, Blom AW, Beswick AD, Team I. Patient-related risk factors for Periprosthetic Joint Infection after total joint arthroplasty: a systematic review and Meta-analysis. PLoS ONE. 2016;11(3):e0150866.
42. Inoue D, Xu C, Yazdi H, Parvizi J. Age alone is not a risk factor for periprosthetic joint infection. J Hosp Infect. 2019;103(1):64–8.
43. Scholten R, Leijtens B, Hannink G, Kamphuis ET, Somford MP, van Susante JLC. General anesthesia might be associated with early periprosthetic joint infection: an observational study of 3,909 arthroplasties. Acta Orthop. 2019;90(6):554–8.
44. Qvistgaard M, Nåtman J, Lovebo J, Almerud-Österberg S, Rolfson O. Risk factors for reoperation due to periprosthetic joint infection after elective total hip arthroplasty: a study of 35,056 patients using linked data of the Swedish hip Arthroplasty Registry (SHAR) and Swedish Perioperative Registry (SPOR). BMC Musculoskelet Disord. 2022;23(1):275.
45. Lenguerrand E, Whitehouse MR, Beswick AD, Kunutsor SK, Foguet P, Porter M, et al. Risk factors associated with revision for prosthetic joint infection following knee replacement: an observational cohort study from England and Wales. Lancet Infect Dis. 2019;19(6):589–600.
46. Rasouli MR, Cavanaugh PK, Restrepo C, Ceylan HH, Maltenfort MG, Viscusi ER, et al. Is neuraxial anesthesia safe in patients undergoing surgery for treatment of periprosthetic joint infection? Clin Orthop Relat Res. 2015;473(4):1472–7.
47. Guay J. The effect of neuraxial blocks on surgical blood loss and blood transfusion requirements: a meta-analysis. J Clin Anesth. 2006;18(2):124–8.
48. Gottschalk A, Rink B, Smektala R, Piontek A, Ellger B, Gottschalk A. Spinal anesthesia protects against perioperative hyperglycemia in patients undergoing hip arthroplasty. J Clin Anesth. 2014;26(6):455–60.
49. Price WN. Big data and black-box medical algorithms. Sci Transl Med. 2018;10(471).
50. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems; Long Beach, California, USA: Curran Associates Inc.; 2017. pp. 4768–77.
51. Chan VWK, Chan PK, Woo YC, Fu H, Cheung A, Cheung MH, et al. Universal haemoglobin A1c screening reveals high prevalence of dysglycaemia in patients undergoing total knee arthroplasty. Hong Kong Med J. 2020;26(4):304–10.
52. Szymski D, Walter N, Alt V, Rupp M. Evaluation of comorbidities as risk factors for fracture-related infection and Periprosthetic Joint Infection in Germany. J Clin Med. 2022;11:17.
53. Shearer J, Agius L, Burke N, Rahardja R, Young SW. BMI is a Better Predictor of Periprosthetic Joint Infection risk than local measures of adipose tissue after TKA. J Arthroplast. 2020;35(6):S313–8.
54. Kuo FC, Hu WH, Hu YJ. Periprosthetic joint infection prediction via machine learning: comprehensible personalized decision support for diagnosis. J Arthroplasty. 2022;37(1):132–41.
55. Parvizi J, Tan TL, Goswami K, Higuera C, Della Valle C, Chen AF, et al. The 2018 definition of Periprosthetic hip and knee infection: an evidence-based and validated Criteria. J Arthroplasty. 2018;33(5):1309–e142.
56. Moldovan F. Bone cement implantation syndrome: a rare disaster following cemented hip Arthroplasties—Clinical considerations supported by Case studies. J Personalized Med. 2023;13(9):1381.
57. Rodriguez-Merchan EC. Instability following total knee arthroplasty. Hss j. 2011;7(3):273–8.

58. Pellegrino A, Coscione A, Santulli A, Pellegrino G, Paracuollo M. Knee periprosthetic fractures in the Elderly: current Concept. Orthop Rev (Pavia). 2022;14(6):38566.

Publisher's note