Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine learning-based identification of perioperative associated factors and early prediction of postoperative delirium in older adults with renal calculi

Xiao-ying Ma
The First Affiliated Hospital of Guangxi Medical University
Wen-xing Xu
Liuzhou People's Hospital
Rui-jia Sun
The First Affiliated Hospital of Guangxi Medical University
Lian Zeng
gmuahzl@163.com

The First Affiliated Hospital of Guangxi Medical University

Research Article

Keywords: Postoperative delirium, machine learning, older adults, renal calculi

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11163782/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Identifying key perioperative associated factors, incorporating both baseline and early postoperative changes, is crucial for early risk warning and targeted care.

Methods

We performed a retrospective study on older adults (aged ≥ 65 years) who underwent surgery for renal calculi between January 2023 and August 2024. Conventional logistic regression and ten machine-learning algorithms were employed to construct models integrating preoperative, intraoperative, and early postoperative variables (up to POD1) to identify core factors associated with POD. Feature importance was interpreted using Shapley Additive Explanations (SHAP).

Results

A total of 321 patients were enrolled, with the incidence of POD being 14.95% (48/321). Preoperative sleep disturbance and lower baseline preoperative C-reactive protein (CRP) were identified as core predictive factors via SHAP analysis. Among the constructed models, Random Forest (AUC = 0.883), Extreme Random Trees (AUC = 0.879), and Gaussian Naive Bayes (AUC = 0.869) were the preferred models for POD risk prediction, with Random Forest demonstrating the highest discriminative ability.

Conclusion

Eleven prediction models (one logistic regression and ten machine-learning models) were evaluated for early postoperative risk assessment of POD in older adults with renal calculi. By integrating dynamic changes on the first postoperative day, these models offer an exploratory tool for evaluating personalized POD probability and identifying high-risk patients experiencing abnormal acute physiological responses to surgery.

Introduction

Renal calculi are a common condition among older adults. With the extension of life expectancy, an increasing number of older patients undergo surgical treatments, such as lithotripsy and stone extraction, to pursue better prognostic outcomes[(Gauhar et al., 2024; He et al., 2024)]. Urological surgeries for renal calculi possess unique pathophysiological characteristics, including subclinical visceral stress caused by frequent urinary tract instrumentation, hemodynamic fluctuations due to intraoperative irrigation fluid absorption, and distinct postoperative visceral stretch pain. These specific factors interact with the fragile neurological compensatory mechanisms of older patients, constituting a unique POD

vulnerability profile for this demographic. However, current prediction models are predominantly designed for general or cardiac surgeries, lacking personalized risk assessment tools tailored for patients undergoing minimally invasive renal calculi surgery. This study aims to bridge this gap by identifying the core associated factors specific to this vulnerable population. Postoperative delirium (POD) is one of the common and severe complications after surgery in patients with renal calculi, typically occurring 2–7 days after surgery (Alhammadi et al., 2024; Deschamps et al., 2024; Nag et al., 2016). The incidence and adverse outcomes of POD in older urological patients are significant. Postoperative delirium is associated with a decline in patients' independent living ability and an increased risk of long-term functional disability (Emmady et al., 2025; Schweickert et al., 2009). Early identification is essential, as the Society of Critical Care Medicine (SCCM) clinical practice guidelines emphasize the importance of managing modifiable associated factors such as sleep, delirium, and pain (Rieck et al., 2020; Schofield–Robinson et al., 2018; Vardy & Thompson, 2020).

In multiple clinical fields, machine learning methods have effectively facilitated the prediction of target events. Previous studies using traditional logistic regression to predict POD in urological patients have achieved an area under the receiver operating characteristic curve (AUC) in the range of 0.779 to 0.790 (Lee et al., 2023; Shi et al., 2025; Song et al., 2024; Wu et al., 2022). Therefore, there is a clear clinical need to utilize machine learning, combined with early dynamic postoperative physiological changes, to improve the accuracy of POD identification. This approach not only provides robust clinical decision support for older adults with renal calculi but also addresses the limitations of traditional preoperative models by capturing the acute response to surgical trauma (Song et al., 2024; Xie et al., 2022; Zhang et al., 2023).

Materials and Methods

Study design and patients

Our study was retrospective. We retrospectively analyzed a cohort of older adult patients who underwent surgery for renal calculi at a tertiary hospital in Guangxi, China, from January 2023 to August 2024. The inclusion criteria were: (1) age ≥ 65 years; (2) undergoing surgery for renal calculi under general anesthesia. Exclusion criteria: (1) Patients with incomplete electronic medical records or missing key baseline data; (2) Patients with confirmed preoperative delirium or a history of severe neurological diseases (e.g., stroke, dementia, Parkinson’s disease); (3) Patients unable to cooperate with the completion of study-related assessments or data collection due to severe cognitive impairment, communication barriers, or other reasons.

Ethics statements

According to the Declaration of Helsinki, the study was approved by the Ethics Committee Board of a grade-A tertiary hospital located in Guangxi Zhuang Autonomous Region, China. (Number: 2023-K324-01). All data were anonymized before analysis, and patient consent was waived due to the retrospective

study design. This retrospective study was registered on the Chinese Clinical Research Network (https://www.chictr.org.cn/), with the registration number: ChiCTR2500101636.

Data Collection

The dataset was established from the medical record system. We collected patient identification information, preoperative parameters, intraoperative data, and postoperative assessment indicators as follows:

For the basic characteristics of patients, the collected information included ID, age, body mass index (BMI), sex, smoking status, alcohol consumption status, and medical history of hypertension, diabetes mellitus, and coronary heart disease. In addition, we also recorded the patients' preoperative sleep disturbance status (assessed using the Athens Insomnia Scale (AIS), with a score of ≥ 6 explicitly defined as the presence of sleep disturbance), a tool designed to quantify sleep quality and identify sleep disorders by evaluating indicators such as sleep duration, sleep continuity, and nighttime awakening frequency), educational level, and American Society of Anesthesiologists (ASA) physical status classification.

Preoperative laboratory test results were collected, covering hemoglobin (Hb), hematocrit (HCT), serum sodium (Na+), serum potassium (K+), serum calcium (Ca2+), blood glucose (Glu), albumin (ALB), C-reactive protein (CRP), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), and complete blood cell count-related indicators including neutrophil count (N), lymphocyte count (L), monocyte count (M), and platelet count (PLT). Intraoperative data were documented in detail, including fentanyl consumption, total fluid infusion volume, intraoperative blood loss, duration of anesthesia, and the use of anesthetics (propofol, ciprofol, remimazolam) and adjuvant drugs (dexmedetomidine). It should be noted that only the binary usage status (yes/no) of these anesthetics was recorded, while specific administration dosages were not included in the current database. Additionally, the status of transfusion during surgery was recorded.

Postoperative assessment indicators included three types: one was the visual analog scale (VAS) scores; the second was the early postoperative laboratory test results and scoring indicators, specifically CRP, Hb, HCT, N, L, M, PLT and AIS scores. To strictly prevent target leakage, all "POD1" physiological parameters and scale assessments were routinely recorded early in the morning (at 07:00–08:00 AM) on the first postoperative day in the electronic medical records. The third was the assessment of postoperative sleep quality, with the Athens Insomnia Scale (AIS) used to monitor changes in sleep status after surgery. Moreover, the occurrence of postoperative sleep disturbance (PSD) was also recorded.

A total of 400 eligible cases were initially identified. After initial screening, 35 cases were excluded due to incomplete preoperative baseline data. The remaining participants completed preoperative data collection and entered a 3-day postoperative follow-up. During the postoperative observation window, 44

cases were excluded due to incomplete postoperative records or early discharge prior to postoperative day 3, resulting in 321 cases finally included for analysis (Fig. 1).

To preserve the original data distribution and minimize imputation bias, random missing values in the dataset (accounting for less than 5% of the total data) were handled using the Multiple Imputation by Chained Equations (MICE) approach prior to model construction.

Definitions of POD

Postoperative delirium (POD) was identified via computer-assisted screening based on the Confusion Assessment Method (CAM) scale [(Hornor et al., 2020; Zhang et al., 2020)] and confirmed by neurologists using the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) criteria [(Kuhn et al., 2014)]. The formal observation and diagnostic window for POD was strictly defined from the evening of postoperative day 1 (after 20:00 PM) to postoperative day 3. Importantly, no patients in our cohort developed clinically detectable delirium prior to the morning collection of the POD1 baseline variables. This rigorous temporal sequencing guarantees that the early postoperative dynamic features (e.g., AIS_POD1, CRP_POD1) acted purely as predictive predictors rather than concurrent symptoms of an ongoing delirium episode, thereby avoiding target leakage.

Model building strategy

We developed 11 models, including Logistic Regression and 10 machine learning algorithms: Random Forest (RF), XGBoost, LightGBM, Linear Discriminant Analysis (LDA), Neural Network (MLP), CatBoost, Extra Trees, AdaBoost, K-Nearest Neighbors (KNN), and Gaussian Naive Bayes (GNB). To comprehensively understand the factors associated with POD and establish a dynamic risk assessment tool, we included variables across the entire perioperative continuum. Candidate predictors encompassed baseline characteristics, intraoperative parameters, and early postoperative dynamic indicators. The inclusion of early postoperative markers allows the model to act as a robust tool for identifying ongoing pathophysiological shifts strongly associated with POD onset. To address the high dimensionality relative to the number of positive events, we performed Least Absolute Shrinkage and Selection Operator (LASSO) regression with 10-fold cross-validation on the training set. Only features with non-zero coefficients were retained for the subsequent training of the 11 machine learning models. This step significantly reduced the risk of overfitting.

Analyses were conducted using the Python programming language (version 3.9) utilizing machine learning libraries including scikit-learn, xgboost, and catboost. The dataset was split into training and validation sets at a 3:1 ratio.

To address the imbalance in positive cases without inducing data leakage, the Borderline SMOTE algorithm was applied strictly to the training dataset after the train-validation split. Potential residual confounders were adjusted for via stepwise variable selection in logistic regression and comprehensive

feature inclusion in machine learning models. We employed 10-fold cross-validation during training. Model interpretability and specific variable contributions were quantified using SHapley Additive exPlanations (SHAP).

Statistical analysis

In this study, statistical analyses were performed using appropriate methods based on the distribution characteristics of variables: Student's t-tests were used to compare normally distributed continuous variables, which were expressed as mean (standard deviation); Mann–Whitney U tests were applied to compare non-normally distributed continuous variables, which were presented as median and interquartile range; and the χ^2 test or Fisher's exact test was used to compare categorical variables, which were expressed as frequency or percentage. The significance level was set at 0.05, and all statistical tests were two-tailed.

For model development and statistical computing: the logistic regression model was constructed using SPSS 26.0 (IBM Corp., Armonk, NY, USA), a widely used statistical software for data analysis and model building in clinical research; all machine learning models were developed utilizing the aforementioned open-source Python libraries.

Results

Study Population and Baseline Characteristics

A total of 321 elderly patients with renal calculi were enrolled in this study. The overall incidence of postoperative delirium was 14.95%. Baseline characteristics of the study population were summarized in Table 1.

Table 1
Patient Demographics and Perioperative Clinical Characteristics

Variables	Total (n = 321)	Non-POD (n = 273)	POD (n = 48)	Statistic	*P*
Age, Mean ± SD	71.05 ± 4.85	70.81 ± 4.76	72.44 ± 5.19	t=-2.15	**0.032**
Gender, n(%)				χ^2=0.44	0.509
Female	221 (68.85)	186 (68.13)	35 (72.92)		
Male	100 (31.15)	87 (31.87)	13 (27.08)		
BMI, Mean ± SD	24.96 ± 3.63	24.97 ± 3.56	24.88 ± 4.04	t = 0.15	0.877
Education, n(%)				χ^2=2.90	0.234
<6 years	145 (45.17)	118 (43.22)	27 (56.25)		
6–12 years	155 (48.29)	137 (50.18)	18 (37.50)		
>12 years	21 (6.54)	18 (6.59)	3 (6.25)		
Hypertension, n(%)				χ^2=0.25	0.620
No	170 (52.96)	143 (52.38)	27 (56.25)		
Yes	151 (47.04)	130 (47.62)	21 (43.75)		
Diabetes, n(%)				χ^2=0.56	0.453
No	284 (88.47)	240 (87.91)	44 (91.67)		
Yes	37 (11.53)	33 (12.09)	4 (8.33)		
Coronary heart disease, n(%)				χ^2=0.00	0.958
No	297 (92.52)	252 (92.31)	45 (93.75)		
Yes	24 (7.48)	21 (7.69)	3 (6.25)		
Smoking, n(%)				χ^2=0.00	1.000
No	289 (90.03)	246 (90.11)	43 (89.58)		
Yes	32 (9.97)	27 (9.89)	5 (10.42)		
Drinking, n(%)				χ^2=1.84	0.175
No	282 (87.85)	237 (86.81)	45 (93.75)		
Yes	39 (12.15)	36 (13.19)	3 (6.25)		

Note: t: t-test, χ^2: Chi-square test, SD: standard deviation.

Variables	Total (n = 321)	Non-POD (n = 273)	POD (n = 48)	Statistic	*P*
Preoperative sleep disturbance, n(%)				χ^2=0.58	0.448
No	170 (52.96)	147 (53.85)	23 (47.92)		
Yes	151 (47.04)	126 (46.15)	25 (52.08)		
ASA, n(%)				χ^2=0.00	0.972
2	188 (58.57)	160 (58.61)	28 (58.33)		
3	133 (41.43)	113 (41.39)	20 (41.67)		
Total fluid infusion volume, Mean ± SD	1368.88 ± 618.10	1372.67 ± 632.21	1347.33 ± 536.23	t = 0.26	0.794
Intraoperative blood loss, Mean ± SD	152.18 ± 221.83	152.71 ± 232.32	149.17 ± 150.59	t = 0.10	0.919
Transfusion, n(%)				χ^2=0.52	0.470
No	199 (61.99)	167 (61.17)	32 (66.67)		
Yes	122 (38.01)	106 (38.83)	16 (33.33)		
Postoperative sleep disturbance, n(%)				χ^2=35.53	**< .001**
No	124 (38.63)	124 (45.42)	0 (0.00)		
Yes	197 (61.37)	149 (54.58)	48 (100.00)		
Note: t: t-test, χ^2: Chi-square test, SD: standard deviation.					

Preoperative laboratory indicators, including hemoglobin, hematocrit, albumin, and high-density lipoprotein cholesterol levels, did not show statistically significant differences between the POD and non-POD groups (all P > 0.05, Supplement Table 1), indicating comparable baseline laboratory profiles. Similarly, preoperative C-reactive protein (CRP) and triglycerides did not differ significantly in univariate analysis. Demographic data including age, body mass index, and sex distribution were detailed, alongside the prevalence of hypertension, diabetes mellitus, coronary heart disease, and preoperative sleep disturbance.

Educational level and self-report physical status classification distributions were also recorded.

Intraoperative parameters such as fentanyl consumption, fluid infusion volume, blood loss, and anesthesia duration were compared between groups. Usage rates of anesthetics including propofol, ciprofol, remimazolam, It is important to note that propofol and ciprofol were used as mutually exclusive alternatives for anesthesia induction and maintenance in our clinical practice. To prevent multicollinearity, their usage was consolidated into a single categorical variable (propofol/ciprofol use)

during the machine learning model construction, and adjuvant drug dexmedetomidine differed, with variations in blood transfusion status noted. The main drug usage during the operation is presented in Table 2. VAS scores and AIS scores were presented, along with postoperative sleep disturbance incidence. Notably, the incidence of postoperative sleep disturbance demonstrated the most striking clinical association, reaching 100% (48/48) in the POD group compared to 54.58% (149/273) in the non-POD group ($P < 0.001$) (Table 1). Detailed preoperative and postoperative laboratory indicators were listed in Supplement Table 1. The detailed results of VAS and AIS scores are presented in Supplement Table 2.

Table 2
The use of drugs in the surgery

Variables	Total (n = 321)	Non-POD (n = 273)	POD (n = 48)	Statistic	P
Propofol, n(%)				χ^2=0.14	0.703
No	108 (33.64)	93 (34.07)	15 (31.25)		
Yes	213 (66.36)	180 (65.93)	33 (68.75)		
Ciprofol, n(%)				χ^2=0.14	0.703
No	213 (66.36)	180 (65.93)	33 (68.75)		
Yes	108 (33.64)	93 (34.07)	15 (31.25)		
Fentanyl consumption, Mean ± SD	0.15 ± 0.06	0.15 ± 0.07	0.13 ± 0.06	t = 1.37	0.172
Remimazolam, n(%)				χ^2=0.03	0.859
No	244 (76.01)	208 (76.19)	36 (75.00)		
Yes	77 (23.99)	65 (23.81)	12 (25.00)		
Dexmedetomidine, n(%)				χ^2=2.62	0.105
No	114 (35.51)	92 (33.70)	22 (45.83)		
Yes	207 (64.49)	181 (66.30)	26 (54.17)		

Note: t: t-test, χ^2: Chi-square test, SD: standard deviation. Propofol and Ciprofol were administered in a mutually exclusive manner; no patient received both.

Prediction Model Development and Performance

Model Performance in Training and Validation

11 machine learning models were developed, forming a comprehensive model pool. The specific LASSO regression feature screening process, detailed hyperparameter settings for each model, training set performance metrics, and model calibration curves are comprehensively provided in the Supplementary

Tables 3 and 4. The training and validation datasets were randomly split at a 3:1 ratio, with 10-fold cross-validation applied for model training. Borderline SMOTE was used to address dataset imbalance. While individual algorithms such as K-Nearest Neighbors exhibited severe overfitting on the training set with an AUC of 1.00 (Supplement Table 5), tree-based ensemble models maintained robust generalization capabilities in the independent validation set. The area under the receiver operating characteristic curve values of the top-performing models exceeded 0.80, with Random Forest and XGBoost showing superior discriminative ability. Logistic regression achieved an AUC of 0.738. Detailed performance metrics including sensitivity, precision, and AUC for all models were summarized in Table 3.

Table 3
Parameters of each model

Model	AUC	Accuracy	Sensitivity	Specificity	Precision	F1 Score
Logistic Regression	0.738	0.702	0.685	0.72	0.698	0.691
Random Forest	0.883	0.835	0.82	0.85	0.842	0.831
XGBoost	0.847	0.801	0.793	0.81	0.805	0.799
LightGBM	0.776	0.738	0.725	0.751	0.74	0.732
Linear Discriminant Analysis (LDA)	0.794	0.755	0.742	0.768	0.758	0.75
Neural Network (MLP)	0.673	0.64	0.628	0.652	0.635	0.631
CatBoost	0.812	0.773	0.76	0.786	0.775	0.767
Extra Trees	0.879	0.828	0.815	0.841	0.833	0.824
AdaBoost	0.810	0.77	0.756	0.784	0.772	0.764
K-Nearest Neighbors (KNN)	0.634	0.605	0.59	0.62	0.6	0.595
Gaussian Naive Bayes (GNB)	0.869	0.819	0.806	0.832	0.825	0.815

Note: All performance metrics (AUC, Accuracy, Sensitivity, etc.) reported in this table are derived exclusively from the independent validation set (25% of the total cohort) to ensure an unbiased evaluation of model generalization.

Hosmer–Lemeshow goodness-of-fit testing confirmed adequate calibration of the logistic regression model. Receiver operating characteristic (ROC) curve analysis demonstrated the predictive performance across different threshold probabilities for all models, with results visualized in Fig. 2A.

Performance Comparison of Machine Learning Models

Among the machine learning models, Random Forest, XGBoost, and CatBoost showed comparable high performance. Extreme Random Trees and Gaussian Naive Bayes also presented favorable AUC values above 0.85. Neural Network and K-Nearest Neighbors exhibited relatively lower discriminative ability with AUC below 0.70.

Model Interpretability

Before utilizing SHAP for detailed interpretation, the relative feature importance was evaluated across the ensemble models. As shown in Fig. 2B, Preoperative Sleep Disturbance and Preoperative CRP were identified as the most critical clinical variables for predicting POD, followed by Drinking, Hb_POD1, and Dexmedetomidine. Subsequently, the SHAP analysis further revealed that alongside preoperative sleep disturbance and preoperative CRP, early postoperative physiological responses, specifically the Athens Insomnia Scale score on POD1 (AIS_POD1) and parameters (Hb_POD1, CRP_POD1, VAS_POD1), played significant roles in the models. This highlights that POD is not solely predetermined by baseline status but is deeply influenced by the acute postoperative inflammatory and physiological trajectory.

We selected the model with the best AUC performance and conducted SHapley Additive exPlanations (SHAP) explanations for the random forest model. SHAP was used to assess feature importance and model interpretability. Individual feature importance scores and their impact on model outputs were visualized in Fig. 3, showing consistent contribution patterns across different models. Among these, preoperative sleep disturbance and preoperative CRP stand out as the most prominent factors: Fig. 3 collectively highlights that preoperative sleep disturbance has the largest mean SHAP value (Fig. 3B) and a notable impact on model outputs (Fig. 3A), while a lower baseline Preoperative CRP also exhibits a high degree of importance, underscoring their key roles in the dynamic risk assessment.

Figure 3C presents a SHAP force plot. For the analysis, a single sample was randomly selected, and the SHAP values of this patient were first standardized before being applied to dynamically evaluate the patient's probability of experiencing POD during the 3-day postoperative observation window. As visualized in the plot, the base value for the model's output stood at -3.157. Through the combined "push" or "pull" effects of SHAP values from various influencing factors, such as the risk-increasing effects of certain variables counterbalanced by the risk-reducing effects of others, the final model output value was adjusted to -5.05. A more negative output value relative to the base value indicates a significantly lower probability of developing POD for this specific patient. This clearly illustrates the clinical utility of SHAP in personalized dynamic risk assessment.

Discussion

The primary objective of this study was to identify key perioperative factors independently associated with POD in older adults with renal calculi and to develop a comprehensive risk stratification model (Hua et al., 2022). The exploratory value of this study lies in the construction of an internal risk assessment framework that integrates preoperative baseline data with early postoperative dynamic indicators. Unlike traditional prediction models that rely solely on static preoperative data, we purposefully incorporated

physiological parameters from the first postoperative day (e.g., Hb_POD1, CRP_POD1) into our algorithms. This 'dynamic early-warning' strategy better reflects the pathophysiological evolution during the perioperative period. Although this design shortens the early-warning time window, it effectively captures the acute inflammatory fluctuations and metabolic disturbances triggered by surgical trauma, which have been proven to be key proximal triggers for delirium (Kang et al., 2023; Li et al., 2025; Stollings et al., 2021). Future studies will attempt to establish stratified early-warning models combining preoperative screening with postoperative monitoring to achieve precise, full-process management.

Our SHAP analysis emphasized preoperative sleep disturbance and lower baseline CRP as core associated variables. Clinically, a relatively lower baseline CRP prior to surgical trauma may reflect a state of immunosenescence or blunted physiological reserve in older frail patients [24–27]. We observed a trend of a higher acute postoperative inflammatory surge on the first day after surgery in the POD group. This dynamic low baseline to high reactivity inflammatory pattern may mediate neurocognitive dysfunction. Given the retrospective design, this finding should be interpreted cautiously and requires future validation with cytokine profiling (Brummel et al., 2024; Dodsworth et al., 2023; Dogan et al., 2024; Huang et al., 2022). Furthermore, we observed a 100% incidence of postoperative sleep disturbance in the POD group. This perfect co-occurrence suggests a high degree of symptom overlap between severe postoperative sleep deprivation and early delirium, highlighting the critical value of perioperative sleep optimization.

While our univariate analysis identified a 100% incidence of postoperative sleep disturbance in the POD group (the strongest clinical association), the SHAP analysis assigned a higher predictive weight to preoperative sleep disturbance. In the clinical pathophysiological sequence, preoperative sleep disturbance acts as the distal vulnerability, leading to acute postoperative sleep disruption (the proximal trigger), which ultimately manifests as POD. The machine learning model prioritizes the distal baseline vulnerability as the core predictor, making the model more suited for early warning rather than merely diagnosing an already occurring acute symptom.

By comparing 11 prediction models, our findings explicitly demonstrate the superiority of tree-based machine learning approaches (AUC > 0.80) over conventional logistic regression (AUC 0.738).

Several limitations must be acknowledged. First, the limited number of positive events introduces a risk of overfitting when training complex algorithms. Although we applied LASSO regression and Borderline SMOTE, the findings remain exploratory. Future studies require larger multicenter cohorts to validate the true performance of the Random Forest model. Second, the database only recorded the binary administration status of anesthetics without specific infusion dosages or durations. Because drugs like dexmedetomidine exhibit highly dose-dependent effects on neurocognitive function, this binary categorization may obscure their true clinical impact and introduce selection bias. Third, regarding clinical utility, a high AUC does not perfectly equate to clinical efficacy. The models require further evaluation of positive and negative predictive values, and the current complexity of SHAP visualizations may hinder direct deployment in grassroots hospitals without dedicated software interfaces. Lastly, by incorporating variables from the first postoperative morning, the model functions strictly as an early

postoperative dynamic assessment tool rather than a preoperative screening tool. It is designed to assist with early ward rounds by capturing acute postoperative physiological shifts.

Conclusion

This study identified preoperative sleep disturbance and preoperative CRP as core predictive factors for POD. The developed dynamic machine learning models effectively utilize routinely collected clinical and early postoperative variables to provide personalized early warning of POD risk. Pending future prospective external validation, this exploratory approach may assist clinicians in identifying vulnerable patients to optimize early postoperative care strategies.

Declarations

CRediT authorship contribution statement: Xiao-ying Ma: Conceptualization, Methodology, Writing - Original Draft. Wen-xing Xu: Data curation, Formal analysis. Rui-jia Sun: Visualization, Validation. Lian Zeng: Supervision, Writing - Review & Editing.

Author Details

Xiao-ying Ma[1] (1227012141@qq.com)

Wen-xing Xu[2] (xuwenxingde@163.com)

Rui-jia Sun[1] (gxmurj1@163.com)

Lian Zeng[1] (gmuahzl@163.com)

[1] Department of Anesthesiology, The First Affiliated Hospital of Guangxi Medical University, Nanning, GuangXi, China

[2] Department of Anesthesiology, Liuzhou People's Hospital, Liuzhou, GuangXi, China

Conflict of interest statement

All authors declare no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

Funding

This study was supported by 2023 Guangxi Traditional Chinese Medicine Suitable Technology Development and Promotion Project-GZSY23-56. The funder had no role in the study design, data collection, analysis, interpretation, manuscript preparation, or the decision to submit for publication.

Ethical Statement

All procedures involving human participants in this study were conducted in compliance with the ethical standards of the institutional and/or national research committee, as well as the 1964 Declaration of Helsinki and its subsequent amendments or equivalent ethical guidelines.

The study was approved by the Ethics Committee Board of the First Affiliated Hospital of Guangxi Medical University (Approval No.: 2023-K324-01). Due to its retrospective study design, the requirement for patient informed consent was waived. All data were anonymized prior to analysis to ensure participant privacy. The protocol for this retrospective study has been registered on the Chinese Clinical Research Network at https://www.chictr.org.cn/, with the registration number: ChiCTR2500101636.

Data Availability Statement

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request, subject to institutional and ethical restrictions.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used no generative AI or AI-assisted technologies in the writing process.

References

1. Alhammadi E, Kuhlmann JM, Rana M, Frohnhofen H, Moellmann HL. Postoperative delirium in oral and maxillofacial surgery: a scoping review. Head Face Med. 2024;20(1):39. https://doi.org/10.1186/s13005-024-00439-9.
2. Brummel NE, Hughes CG, McNeil JB, Pandharipande PP, Thompson JL, Orun OM, Raman R, Ware LB, Bernard GR, Harrison FE, Ely EW, Girard TD. Systemic inflammation and delirium during critical illness. Intensive Care Med. 2024;50(5):687–96. https://doi.org/10.1007/s00134-024-07388-6.
3. Deschamps A, Ben Abdallah A, Jacobsohn E, Saha T, Djaiani G, El-Gabalawy R, Overbeek C, Palermo J, Courbe A, Cloutier I, Tanzola R, Kronzer A, Fritz BA, Schmitt EM, Inouye SK, Avidan MS. Electroencephalography-Guided Anesthesia and Delirium in Older Adults After Cardiac Surgery: The ENGAGES-Canada Randomized Clinical Trial. JAMA. 2024;332(2):112–23. https://doi.org/10.1001/jama.2024.8144. Canadian Perioperative Anesthesia Clinical Trials.
4. Dodsworth BT, Reeve K, Falco L, Hueting T, Sadeghirad B, Mbuagbaw L, Goettel N, Gelsomino S, N. Development and validation of an international preoperative risk assessment model for postoperative delirium. Age Ageing. 2023;52(6). https://doi.org/10.1093/ageing/afad086.
5. Dogan RB, Yavuz BG, Satilmis D, Afacan MA, Colak S. CRP/albumin, Glasgow prognostic score, and prognostic nutritional index as a predictor of mortality among delirium patients. Ir J Med Sci. 2024;193(1):469–76. https://doi.org/10.1007/s11845-023-03443-1.
6. Emmady PD, Schoo C, Tadi P. (2025). Major Neurocognitive Disorder (Dementia). In *StatPearls*. https://www.ncbi.nlm.nih.gov/pubmed/32491376

7. Gauhar V, Traxer O, Castellani D, Fong KY, Hamri B, Gokce S, Gadzhiev MI, Corrales N, Malkhasyan M, Ragoori V, Soebhali D, Tan B, Chai K, Tursunkulov CA, Tanidir AN, Persaud Y, Elshazly S, Kamal M, Tefik W, Somani T, B. K. Operative outcomes 24 hours after retrograde intrarenal surgery for solitary renal calculi using a flexible and navigable suction ureteral access sheath. A prospective global multicenter study by the European Association of Urology Section on Urolithiasis. Minerva Urol Nephrol. 2024;76(5):625–34. https://doi.org/10.23736/S2724-6051.24.05961-5.
8. He M, Dong Y, Cai W, Cai J, Xie Y, Yu M, Li C, Wen L. Recent advances in the treatment of renal stones using flexible ureteroscopys. Int J Surg. 2024;110(7):4320–8. https://doi.org/10.1097/JS9.0000000000001345.
9. Hornor MA, Ma M, Zhou L, Cohen ME, Rosenthal RA, Russell MM, Ko CY. Enhancing the American College of Surgeons NSQIP Surgical Risk Calculator to Predict Geriatric Outcomes. J Am Coll Surg. 2020;230(1):88–e100101. https://doi.org/10.1016/j.jamcollsurg.2019.09.017.
10. Hua Y, Chen S, Xiong X, Lin C, Li D, Tu P. Risk factors for postoperative delirium in elderly urological patients: A meta-analysis. Med (Baltim). 2022;101(38):e30696. https://doi.org/10.1097/MD.0000000000030696.
11. Huang X, Li L, Feng Q. (2022). Correlation Analysis of Inflammatory Markers CRP and IL-6 and Postoperative Delirium (POD) in Elderly Patients: A Meta-Analysis of Observational Studies. *J Environ Public Health*, *2022*, 1136386. https://doi.org/10.1155/2022/1136386
12. Kang J, Cho YS, Lee M, Yun S, Jeong YJ, Won YH, Hong J, Kim S. Effects of nonpharmacological interventions on sleep improvement and delirium prevention in critically ill patients: A systematic review and meta-analysis. Aust Crit Care. 2023;36(4):640–9. https://doi.org/10.1016/j.aucc.2022.04.006.
13. Kuhn E, Du X, McGrath K, Coveney S, O'Regan N, Richardson S, Teodorczuk A, Allan L, Wilson D, Inouye SK, MacLullich AM, Meagher D, Brayne C, Timmons S, Davis D. Validation of a consensus method for identifying delirium from hospital records. PLoS ONE. 2014;9(11):e111823. https://doi.org/10.1371/journal.pone.0111823.
14. Lee DY, Oh AR, Park J, Lee SH, Choi B, Yang K, Kim HY, Park RW. Machine learning-based prediction model for postoperative delirium in non-cardiac surgery. BMC Psychiatry. 2023;23(1):317. https://doi.org/10.1186/s12888-023-04768-y.
15. Li J, Fan Y, Luo R, Yin N, Wang Y, Jing J, Zhang J. The Impact of Non-Pharmacological Sleep Interventions on Delirium Prevention and Sleep Improvement in Postoperative ICU Patients: A Systematic Review and Network Meta-Analysis. Intensive Crit Care Nurs. 2025;87:103925. https://doi.org/10.1016/j.iccn.2024.103925.
16. Nag DS, Chatterjee A, Samaddar DP, Singh H. Sepsis associated delirium mimicking postoperative delirium as the initial presenting symptom of urosepsis in a patient who underwent nephrolithotomy. World J Clin Cases. 2016;4(5):130–4. https://doi.org/10.12998/wjcc.v4.i5.130.
17. Rieck KM, Pagali S, Miller DM. (2020). Delirium in hospitalized older adults. *Hosp Pract (1995)*, *48*(sup1), 3–16. https://doi.org/10.1080/21548331.2019.1709359

18. Schofield-Robinson OJ, Lewis SR, Smith AF, McPeake J, Alderson P. Follow-up services for improving long-term outcomes in intensive care unit (ICU) survivors. Cochrane Database Syst Rev. 2018;11(11):CD012701. https://doi.org/10.1002/14651858.CD012701.pub2.
19. Schweickert WD, Pohlman MC, Pohlman AS, Nigos C, Pawlik AJ, Esbrook CL, Spears L, Miller M, Franczyk M, Deprizio D, Schmidt GA, Bowman A, Barr R, McCallister KE, Hall JB, Kress JP. Early physical and occupational therapy in mechanically ventilated, critically ill patients: a randomised controlled trial. Lancet. 2009;373(9678):1874–82. https://doi.org/10.1016/S0140-6736(09)60658-9.
20. Shi JY, Gao DP, Chen R, Hu XY, Zhu LY, Zhang Y, Li Q, Mao QH, Ji MH, Fan D, Liu QR. Leveraging data-driven machine learning: From explainable risk prediction to hierarchical clustering-based subtypes of postoperative delirium in a prospective non-cardiac surgery cohort. J Clin Anesth. 2025;107:112006. https://doi.org/10.1016/j.jclinane.2025.112006.
21. Song Y, Zhang D, Wang Q, Liu Y, Chen K, Sun J, Shi L, Li B, Yang X, Mi W, Cao J. Prediction models for postoperative delirium in elderly patients with machine-learning algorithms and SHapley Additive exPlanations. Transl Psychiatry. 2024;14(1):57. https://doi.org/10.1038/s41398-024-02762-w.
22. Stollings JL, Kotfis K, Chanques G, Pun BT, Pandharipande PP, Ely EW. Delirium in critical illness: clinical manifestations, outcomes, and management. Intensive Care Med. 2021;47(10):1089–103. https://doi.org/10.1007/s00134-021-06503-1.
23. Vardy E, Thompson RE. Quality improvement and delirium. Eur Geriatr Med. 2020;11(1):33–43. https://doi.org/10.1007/s41999-019-00268-z.
24. Wu H, Ding T, Yan S, Huang Z, Zhang H. Risk factors for moderate-to-severe postoperative pain after percutaneous nephrolithotomy: a retrospective cohort study. Sci Rep. 2022;12(1):8366. https://doi.org/10.1038/s41598-022-12623-5.
25. Xie Q, Wang X, Pei J, Wu Y, Guo Q, Su Y, Yan H, Nan R, Chen H, Dou X. Machine Learning-Based Prediction Models for Delirium: A Systematic Review and Meta-Analysis. J Am Med Dir Assoc. 2022;23(10):1655–e16681656. https://doi.org/10.1016/j.jamda.2022.06.020.
26. Zhang LM, Hornor MA, Robinson T, Rosenthal RA, Ko CY, Russell MM. Evaluation of Postoperative Functional Health Status Decline Among Older Adults. JAMA Surg. 2020;155(10):950–8. https://doi.org/10.1001/jamasurg.2020.2853.
27. Zhang Y, Hu J, Hua T, Zhang J, Zhang Z, Yang M. Development of a machine learning-based prediction model for sepsis-associated delirium in the intensive care unit. Sci Rep. 2023;13(1):12697. https://doi.org/10.1038/s41598-023-38650-4.

Figures

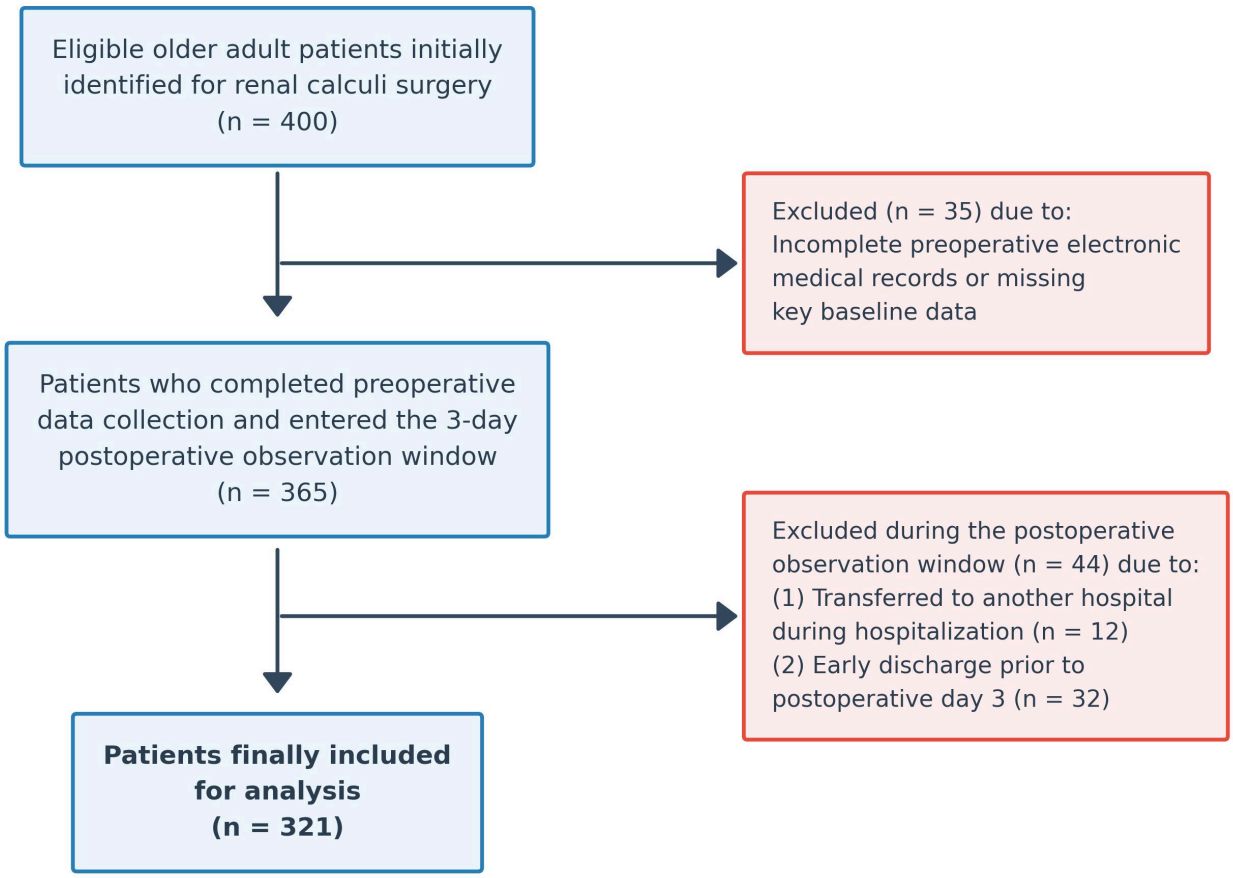


Figure 1

Flowchart of participant recruitment and follow-up

A total of 400 eligible cases were initially identified. After initial screening, 35 cases were excluded due to incomplete preoperative baseline data. The remaining participants completed preoperative data collection and entered a 3-day postoperative follow-up. During the postoperative observation window, 44 cases were excluded due to incomplete postoperative records or early discharge prior to postoperative day 3, resulting in 321 cases finally included for analysis.

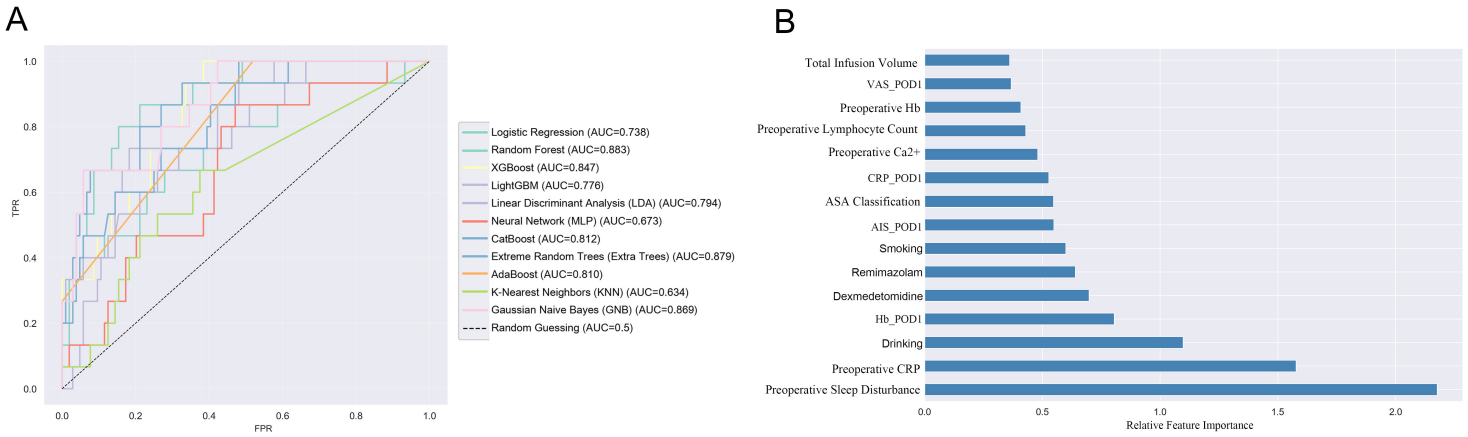


Figure 2

Performance of machine learning models and feature importance analysis

(A) Receiver operating characteristic curves comparing the performance of various machine learning models, with area under the curve values indicated. (B) Bar plot showing the relative feature importance for each variable. The importance score was calculated based on the mean decrease in Gini impurity across all decision trees in the ensemble models, normalized to clearly display the top contributing clinical variables.

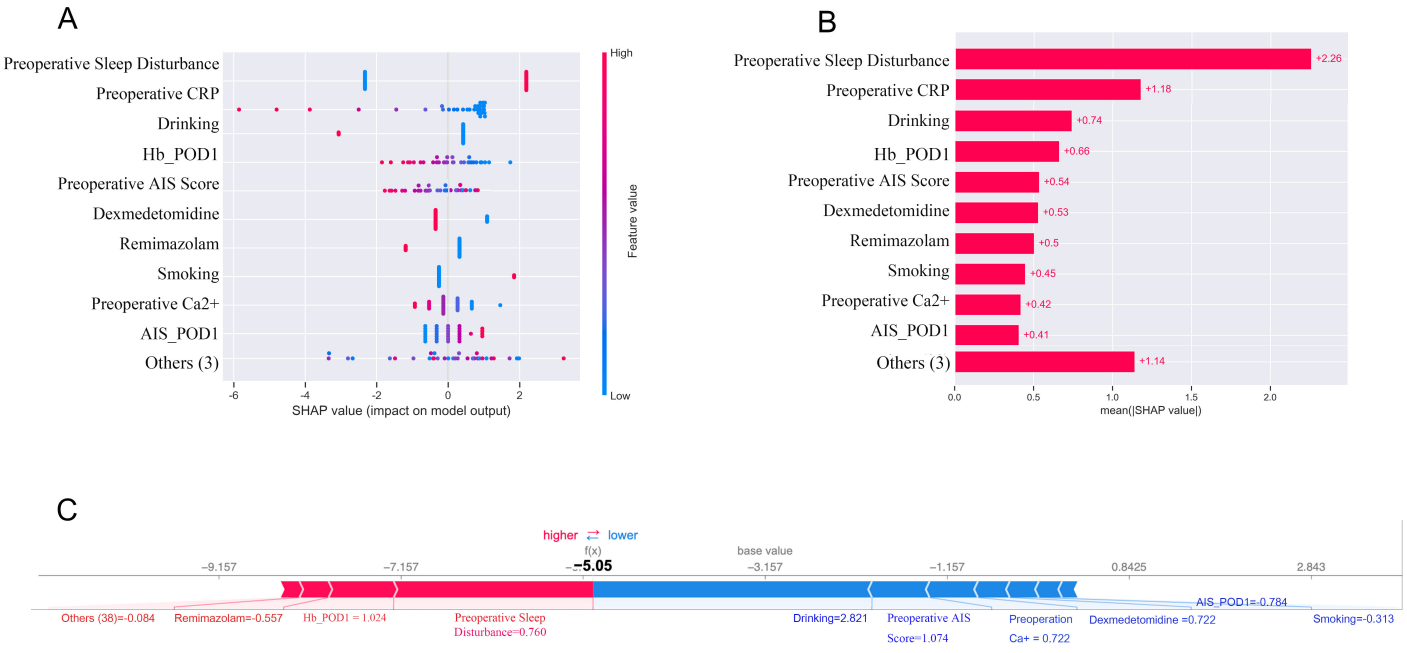


Figure 3

SHAP analysis of feature contributions

(A) SHAP summary plot illustrating the impact of each feature on the model output, where each dot represents a sample and the color indicates feature value (high vs. low). (B) Mean absolute SHAP value bar plot, quantifying the average contribution of each feature to model predictions. (C) SHAP force plot for a single sample, decomposing the model's prediction into the contribution of each feature relative to the base value.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryTable1.docx
- supplementaryTable2.docx
- supplementaryTable3.docx
- supplementaryTable4.docx
- supplementaryTable5.docx