



OPEN ACCESS

EDITED BY

Wen Ouyang,
Central South University, China

REVIEWED BY

Sunil Swami,
Blue Health Intelligence (BHI), United States
Vito Domenico Bruno,
IRCCS Galeazzi Sant'Ambrogio—Department
of Minimally Invasive Cardiac Surgery, Italy
Khan Md. Hasib,
Bangladesh University of Business and
Technology, Bangladesh

*CORRESPONDENCE

Tianzuo Li
✉ sjmzltz@163.com
Dongxin Wang
✉ wangdongxin@hotmail.com
Huihui Miao
✉ iverymhh@hotmail.com

RECEIVED 22 October 2023

ACCEPTED 29 December 2023

PUBLISHED 11 January 2024

CITATION

Sheng W, Tang X, Hu X, Liu P, Liu L,
Miao H, Wang D and Li T (2024) Random
forest algorithm for predicting postoperative
delirium in older patients.
Front. Neurol. 14:1325941.
doi: 10.3389/fneur.2023.1325941

COPYRIGHT

© 2024 Sheng, Tang, Hu, Liu, Liu, Miao, Wang
and Li. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Random forest algorithm for predicting postoperative delirium in older patients

Weixuan Sheng¹, Xianshi Tang², Xiaoyun Hu¹, Pengfei Liu¹,
Lei Liu³, Huihui Miao^{1*}, Dongxin Wang^{4*} and Tianzuo Li^{1*}

¹Department of Anesthesiology, Beijing Shijitan Hospital, Capital Medical University, Beijing, China, ²Key Laboratory of National Health Commission on Parasitic Disease Control and Prevention, Key Laboratory of Jiangsu Province on Parasite and Vector Control Technology, Jiangsu Institute of Parasitic Diseases, Wuxi, China, ³Department of Science and Technology, Beijing Shijitan Hospital, Capital Medical University, Beijing, China, ⁴Department of Anesthesiology, Peking University First Hospital, Beijing, China

Objective: In this study, we were aimed to identify important variables via machine learning algorithms and predict postoperative delirium (POD) occurrence in older patients.

Methods: This study was to make the secondary analysis of data from a randomized controlled trial. The Boruta function was used to screen relevant basic characteristic variables. Four models including Logistic Regression (LR), K-Nearest Neighbor (KNN), the Classification and Regression Tree (CART), and Random Forest (RF) were established from the data set using repeated cross validation, hyper-parameter optimization, and Smote technique (Synthetic minority over-sampling technique, Smote), with the calculation of confusion matrix parameters and the plotting of Receiver operating characteristic curve (ROC), Precision recall curve (PRC), and partial dependence graph for further analysis and evaluation.

Results: The basic characteristic variables resulting from Boruta screening included grouping, preoperative Mini-Mental State Examination (MMSE), CHARLSON score, preoperative HCT, preoperative serum creatinine, intraoperative bleeding volume, intraoperative urine volume, anesthesia duration, operation duration, postoperative morphine dosage, intensive care unit (ICU) duration, tracheal intubation duration, and 7-day postoperative rest and move pain score (median and max; VAS-Rest-M, VAS-Move-M, VAS-Rest-Max, and VAS-Move-Max). And Random Forest (RF) showed the best performance in the testing set among the 4 models with Accuracy: 0.9878; Matthews correlation coefficient (MCC): 0.8763; Area under ROC curve (AUC-ROC): 1.0; Area under the PRC Curve (AUC-PRC): 1.0.

Conclusion: A high-performance algorithm was established and verified in this study demonstrating the degree of POD risk changes in perioperative elderly patients. And the major risk factors for the development of POD were CREA and VAS-Move-Max.

KEYWORDS

postoperative delirium, random forest, confusion matrix, partial dependence graph, older patient

1 Background

Postoperative delirium (POD) refers to delirium that occurs after surgery and is defined as an acute mental disorder characterized by disturbances of consciousness, attention, and cognition. The reported incidence ranges from 5 to 52% in elderly patients (1). The mechanism of delirium is not totally clear, but is generally believed to be a result of co-action of predisposing factors and external stress. However, its highly preventable nature determines that early and effective intervention would reduce its occurrence and related treatment costs which are estimated to exceed 100 billion US dollars (2) annually across the globe. Studies have shown that early preventive measures decreased the odds of POD by about 30% in high risk patients (3, 4). Therefore, it is crucial to identify and control the relevant factors contributing to the POD development.

As a major branch of artificial intelligence (AI), machine learning has the advantages in establishing models with more stable and accurate prediction, and is therefore increasing used for such purposes as clinical prediction. The use of artificial intelligence to solve clinical problems and the construction of a precision medical research model based on complex data acquisition and integration utilization will drive innovation and development in clinical medicine. In this study, based on the reviewed database of elderly patients, we used machine learning algorithms to screen risk factors and to predict the risk of POD, in order to assist clinicians in developing personalized management plans for patients in a timely manner.

2 Materials and methods

2.1 Study design and subjects

The data in this article came from delirium in Older Patients after Combined Epidural-General Anesthesia or General Anesthesia for Major Surgery: a Randomized Trial (5). This was a secondary analysis of database from a previous trial. The trial protocol was approved by the Institutional Review Committee of Peking University (Approval No. 00001052-11048) and the ethics committees of five participating centers, and was registered with the Chinese Clinical Trial Registry (www.chictr.org.cn; identifier: ChiCTR-TRC-09000543) and ClinicalTrials.gov (identifier: NCT01661907). The original trial was conducted in five tertiary hospitals in Beijing, China. All participants provided written informed consent.

During the original trial, we enrolled patients aged 60–90 years who underwent elective non-cardiac thoracic or abdominal surgery for at least 2 h and required patient-controlled analgesia after surgery. We excluded those who had severe nervous system disease, acute myocardial infarction or stroke, severe cardiac insufficiency, severe

hepatic insufficiency or renal failure within 3 months, or any contraindication to epidural anesthesia.

2.2 Anesthesia and perioperative care

No premedication was given. Patients were randomized to receive either general anesthesia or combined epidural-general anesthesia in the original trial. For those assigned to general anesthesia alone, patient-controlled intravenous analgesia was provided after surgery. For those assigned to combined anesthesia, epidural block was performed during surgery, followed by patient-controlled epidural analgesia after surgery. Other perioperative care including adverse events were managed per routine.

2.3 Data collection and outcome assessment

Baseline data included demographic characteristics, preoperative comorbidity, surgical diagnosis, and main laboratory test results. General health status was evaluated with the Charlson Comorbidity Index and ASA physical status classification. Cognitive function was evaluated with the Mini-Mental State Examination (MMSE). Anxiety and depression were evaluated with the Hospital Anxiety and Depression Scale (HAD). Intraoperative data included type and duration of anesthesia, type and dose of medications, circulation parameters, and type and duration of surgery.

After surgery, patients were followed up twice daily during the first 7 days, and then weekly until 30 days. Pain severity was assessed with the Numeric Rating Scale during the first 7 postoperative days. Delirium assessment: delirium for patients in ICU was assessed by the Confusion Assessment Method for the ICU (CAM-ICU) (6), which has been validated in Chinese patients in the ICU setting (7) and the feasibility of which had been established in our prior studies (8, 9). For patients did not admitted to ICU, delirium was assessed by CAM.

Fifty-eight potentially useful characteristics considered in this study included the followings: basic personal information including age and gender; preoperative comorbidities; Charlson Comorbidities Index scores; MMSE and Hospital Anxiety and Depression Scale (HAD) score; preoperative laboratory examination results; intraoperative anesthesia medication, circulation parameters and grouping (simple general anesthesia group and general anesthesia combined with epidural anesthesia group); postoperative NRS, worst APACHE II score, etc.

2.4 Statistical analysis and sample size

R (version 4.2.2) and RStudio (version 2023.06.0+421) were used for data statistical analysis. The normal distribution of numeric variables was tested by the Shapiro–Wilk test. Continuous variables with a normal distribution were presented as mean \pm standard deviation (SD) and compared using the independent-sample *t*-test. Continuous variables with non-normal distribution were presented as median (IQR) and compared using Mann–Whitney U test. Categorical data were expressed as number (%) and analyzed using the chi-square

Abbreviations: MMSE, Mini-mental state examination; Charlson, Charlson comorbidity index; HCT, Hematocrit; ALB, Albumin; Glu, Blood glucose; CREA, Serum creatinine; BUN, Blood urea nitrogen; MAP, Mean arterial pressure; MHR, Mean heart rate; APACHE II, Acute Physiology and Chronic Health Evaluation II; VAS, Visual analogous scale; TIA, Transient ischemic attack; COPD, Chronic obstructive pulmonary disease; CHD, Coronary heart disease; HT, Hypertension; NYHA, New York Heart Association; DM, Diabetes mellitus; HLP, Hyperlipidemia; NSAIDs, Non-steroid anti-inflammatory drugs.

test or Fisher's exact probability test. Parameters with missing data of more than 20% were excluded from the final dataset. Parameters with missing data of less than 20% were interpolated using the missForest package. MissForest package is a non-parametric method that utilizes random forests to impute missing values, suitable for both continuous and categorical variables. Its core algorithm is to use known variables as independent variables and variables containing missing values as dependent variables to establish a random forest to predict missing values. It yields an out-of-bag (OOB) imputation error estimate.

Independent influencing factors were derived from POD related important characteristic variables used for Boruta screening. Creates possibly balanced samples by Smote technique (the synthetic minority over-sampling technique, Smote) (10–12). The basic idea of Smote technique is to analyze minority samples and manually synthesize new samples based on minority samples to add to the dataset.

The selection of model hyperparameter optimization used repeated k -fold cross validation (folds = 10, repeats = 10) on data set. Repetitive k -fold cross validation is an extension of k -fold cross validation. In this article, it divides the dataset into 10 mutually exclusive subsets of the same size. Each subset is used as a validation dataset to validate the model, while the other nine subsets are used as training datasets to train the model. The appeal process is repeated 10 times. Meanwhile, the performance of the machine model is directly related to hyperparameters. The better the hyperparameter tuning, the better the resulting model.

The mlr3verse package was used to successively build models from the data set via repeated k -fold cross-validation, hyperparameter optimization, and Smote technique 4 machine learning prediction models (LR, KNN, CART, and RF) were included and analyzed, with RF proven better than the other 3 in terms of misclassification rate. RF as the optimal model is further analyzed and evaluated by calculating the parameters of confusion matrix and drawing ROC and PRC. The iml and DALEX package was used to draw importance ranking for important characteristic variables, partial dependence graph, and break down profile to interpret the optimal model.

Basic Principles of Random Forest:

$$Y = H(x) = \arg \max_y \sum_{k=1}^n I(h_k(x) = y)$$

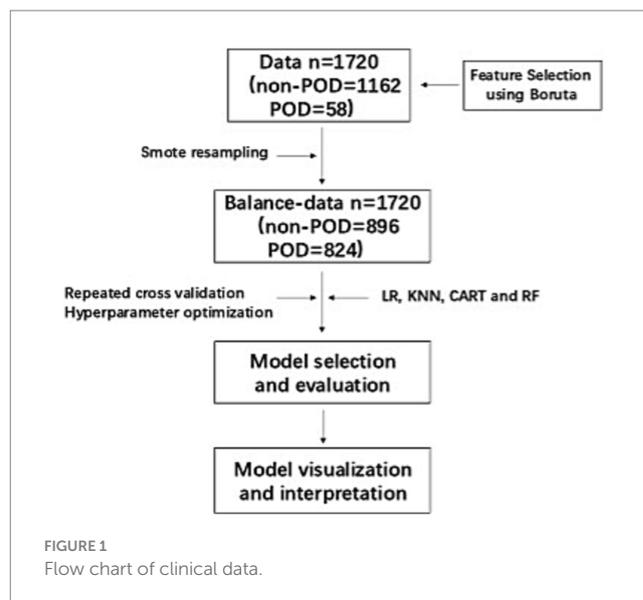
$H(x)$ is a combination classification model; Y is the final classification result; $h_k(x)$ is a single decision tree classifier; y is the classification result of a single decision tree classifier; $I(\cdot)$ is an indicative function.

For the binary classification prediction model, the calculated final sample size obtained through the pmsampsize function of RStudio was 1,459, less than the sample size 1,720 included in this study, with function parameters set as follows: the adjusted maximum R2: 0.327; The number of independent variable parameters to be included: 30; Incidence of postoperative delirium: 0.05 (13–15).

3 Results

3.1 Flow chart and baseline of clinical data

A total of 1720 patients were included in the original data. The process of data inclusion, model establishment, and evaluation were



presented in Figure 1. First, Boruta was applied to filter characteristic variables in the original data. Then, Smote technology was used to establish a balanced-data ($n = 1720$, non-POD = 896, POD = 824) on top of the original data. Finally, repeated k -fold cross-validation and hyperparameter optimization were used to obtain the optimal model in balance-data, and to perform validation and interpretation.

Among the included patients, 58 (3.372%) developed delirium within the first 7 postoperative days. Comparison of preoperative, intraoperative, and postoperative variables between POD and non-POD patients were listed in Supplementary Table 1. The CREA was significantly lower in POD group than non-POD group [73.50 (61.00–93.00) vs. 86.00 (76.00–99.00), $p < 0.001$]. The depression score was significantly higher in patients with POD. The patients with POD had lower nitrosoxide (%), postoperative morphine (mg), GEA-PCEA (%) and higher seveflurane (%), midazolam (mg), urine (mL), bleeding (mL), MAP (mmHg), MHR (times/min), perioperative morphine (mg), ICU duration (min), intubation duration (min), APACHE-II, VAS-Rest-M, VAS-Move-M, VAS-Rest-Max, VAS-Move-Max, VAS-Rest-Min, VAS-Move-Min, GA-PCIA (%), and intraoperative hypotension (%).

3.2 Screening of characteristic variables using Boruta

Boruta analysis showed that grouping, preoperative MMSE, CHARLSON, preoperative HCT, preoperative serum creatinine, intraoperative bleeding volume, intraoperative urine volume, anesthesia duration, operation duration, postoperative morphine dosage, ICU admission, VAS-Rest-M, VAS-Move-M, VAS-Rest-Max, and VAS-Move-Max were 16 characteristic variables included in the model in Figure 2.

3.3 Model establishment, selection, and evaluation

After identifying these 16 variables, machine learning models were used to predict POD. CE, AU-ROC, and AUC-PRC were

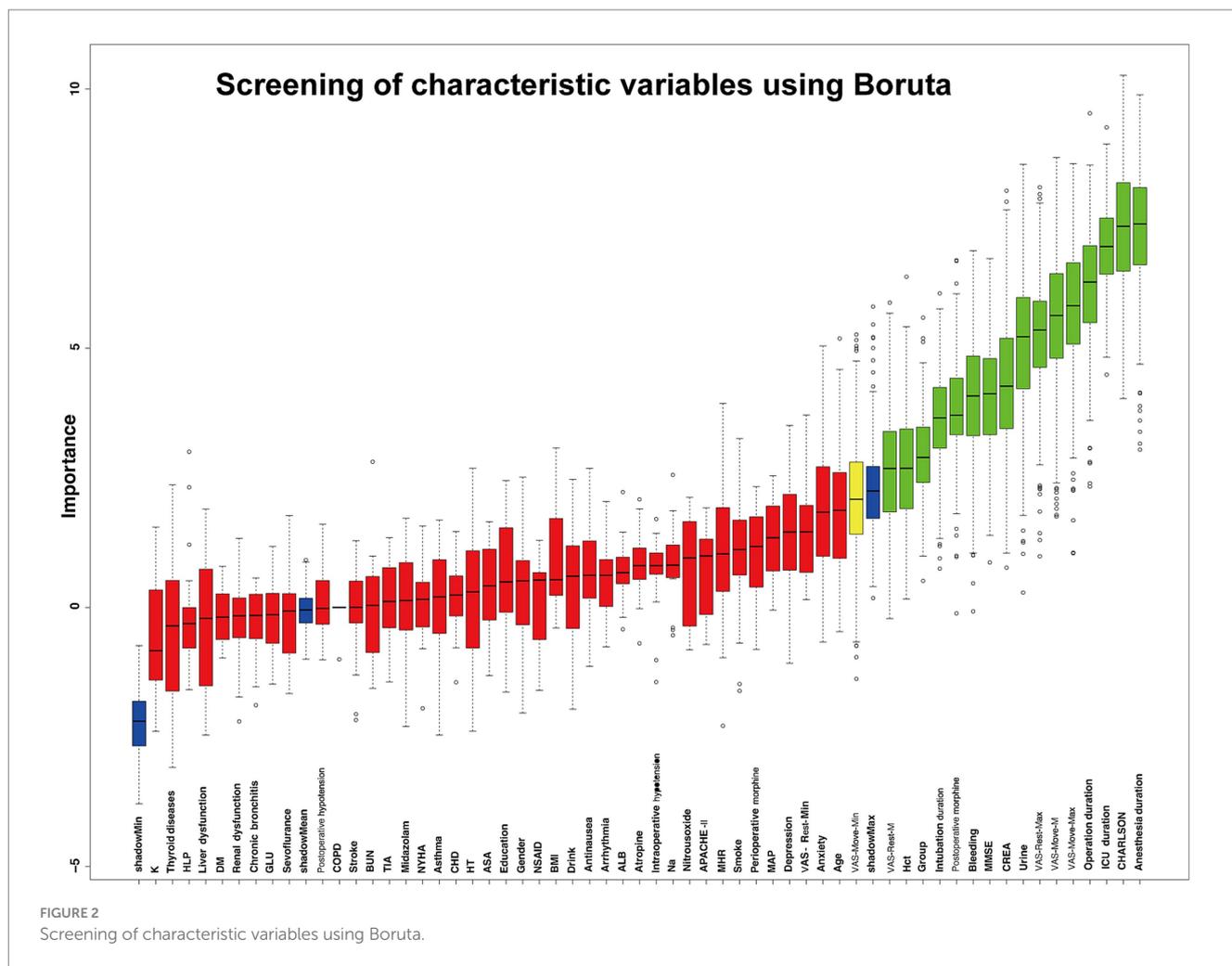


FIGURE 2 Screening of characteristic variables using Boruta.

important indicators used to evaluate prediction models. Among the four models established, RF showed the best performance in error rate (Ce), ROC, and PRC from Figure 3. From Figure 3A, it can be determined that Ce of RF is the lowest among the four models. From Figures 3B,C, it can be determined that the AUC-ROC and AUC-PRC of RF are the highest among the four models. Finally, we calculated the parameters of confusion matrix in RF: Accuracy 0.998, MCC 0.997, AUC-ROC 1.0, and AUC-PRC 1.0. The above data showed that the random forest model had excellent performance in accuracy, overall performance, overall discrimination, and positive result discrimination.

3.4 Importance ranking and partial dependency graph of characteristic variables

Ranking and partial dependency graphs of 16 characteristic variables were established through RF model in Figures 4, 5. In importance ranking, it could be intuitively seen how much each characteristic variable contributes to the predicted variable. In our study, the level of CREA and VAS-Move-Max ranked first and second in importance. Partial dependency graph was used to analyze RF model, showing the reflect the influence of each feature

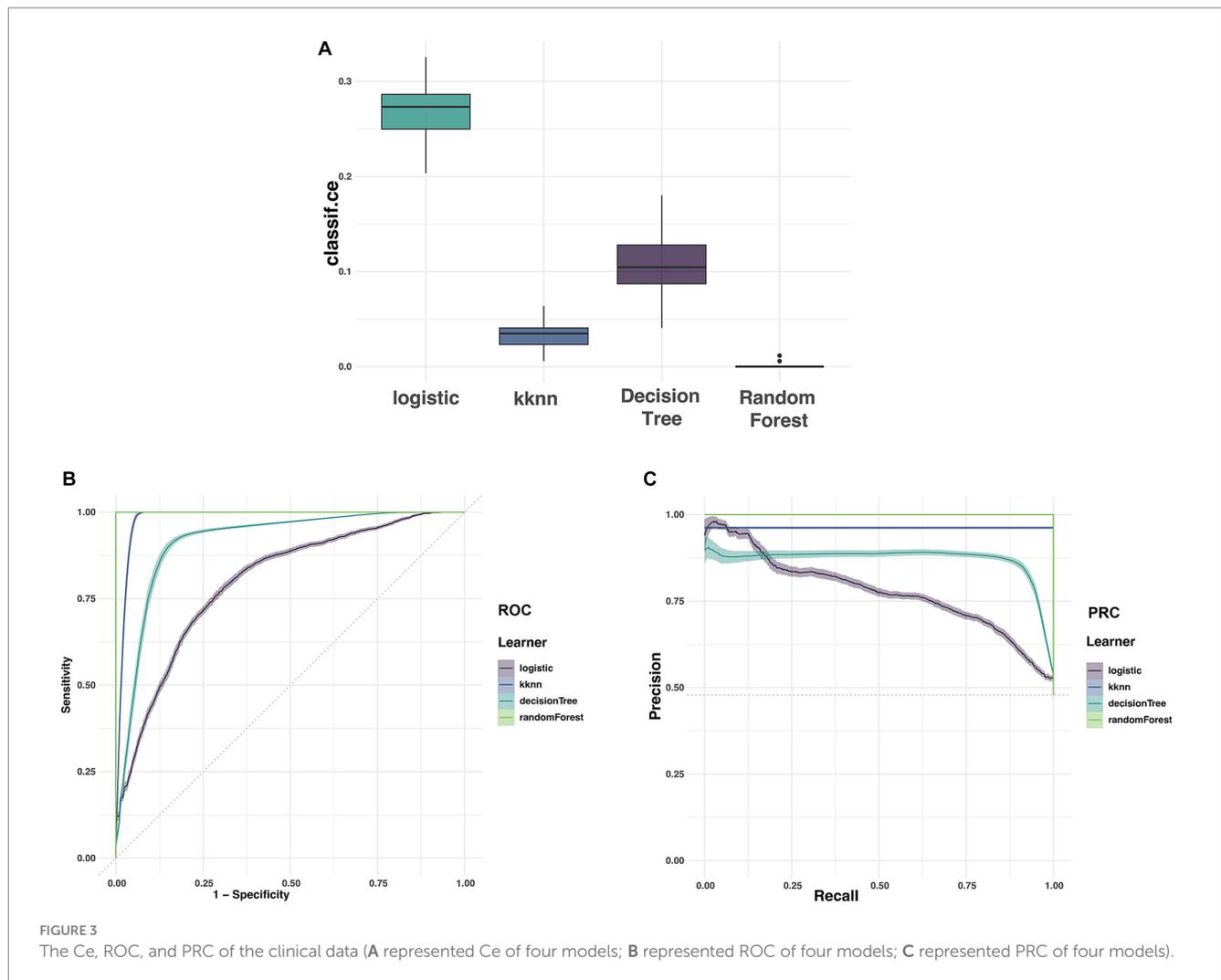
in the sample and also showing the positive and negative influences. At the same time, when the characteristic variable was above or below the cutoff value, the predictive variable would undergo a qualitative transformation.

3.5 Break down profile to explain a single sample in RF

The Breakdown profile visualizes the contribution of each variable to the prediction for a single sample in Figure 6. The model predicts that the value of a sample (delirium as an outcome variable) is 0.915, and the red or blue bars display the impact of each variable on the prediction. The predicted value is equal to the sum of the contributions of each feature.

4 Discussion

In importance ranking, it can be intuitively seen that Creatinine, as a feature variable, has the greatest contribution to the predictive variable. Furthermore, through partial dependency graphs, POD occurrence probability was increased no matter when preoperative creatinine levels were lower or higher than normal, which is consistent



with the published results (16), demonstrating that kidney function has an impact on brain cognitive ability. The possible reason is that metabolic disorders of renal function (abnormal levels of creatinine, etc.) affect the cognitive function.

As the influencing factors before surgery, the partial dependence graph of preoperative MMSE used for cognitive function assessment showed that patients with MMSE < 20 and diagnosed with moderate to severe cognitive impairment before surgery would have the significantly increased probability of POD occurrence, confirming studies reporting the correlation between preoperative MMSE and POD (17) and identifying the preexisting cognitive impairment as the important basis of POD. This study showed that CHARLSON score was positively correlated with postoperative cognitive impairment, with the cutoff value of 100 in the partial dependence graph, which is consistent with former literature reports (18) and possibly related to the stress state caused by the existing physical illness of patients.

As POD predictors (19, 20) directly related to cerebral hypoxia, when preoperative HCT < 30, the risk of POD increased with the decrease of HCT, and when the intraoperative blood loss was less than 500 mL, the curve of the partial dependence graph rose sharply, and then at a gentle pace. The model also predicted that the increased probability of POD would come with the increase of urine volume, with the volume of 500 mL as the visible cutoff value of the curve

which rose sharply when urine volume was between 500 and 1,000 mL and fell slowly when urine volume was over 1,000 mL. The common cause for urine output increase is excessive perioperative fluid load which leads to complications such as heart failure, pulmonary edema, and postoperative cognitive dysfunction (21) and the wide application of goal-oriented fluid therapy (GDFT) in clinical practice would effectively prevent it from happening. The partial dependence graphs showed that either a prolonged anesthesia duration or surgery duration would result in POD risk increase, with the latter's partial dependence graph curve (the cutoff value at about 180 min) steeper than the former's. The impact caused by anesthesia duration could be attributed to the inhibition of sedation and analgesic drugs on the central nervous system, whereas surgery trauma could increase the release of peripheral and central inflammatory factors and cause neuroinflammation and changes in cognitive function (22).

Among the postoperative influencing factors, the risk of POD would be minimized when the postoperative opioid dosage was less than 50 mg (converted to equivalent morphine dosage), but when the dosage was greater than 120 mg, the probability of POD occurrence increased significantly. In order to reduce the side effects of opioid overdose, general anesthesia combined with epidural or nerve block could be given preference during operation for its obvious advantages in perioperative cognitive improvement and POD prevention

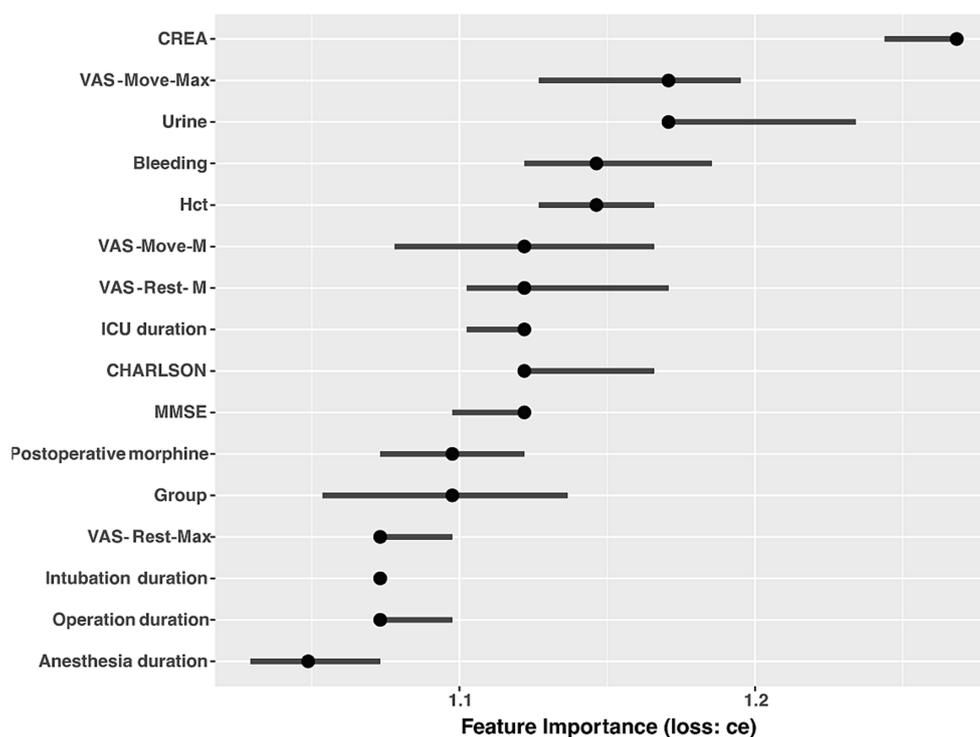
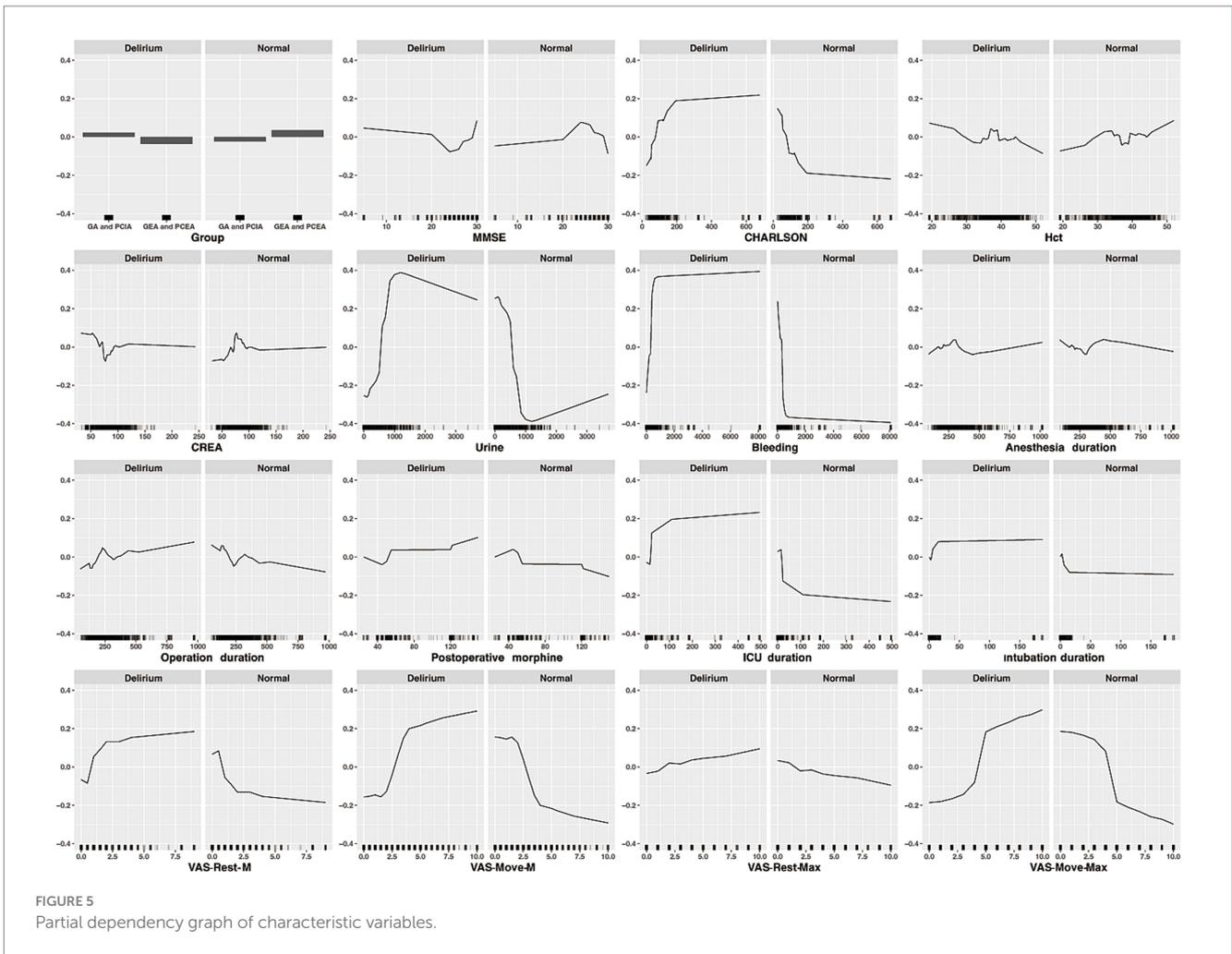


FIGURE 4
Importance ranking of characteristic variables.

compared with simple general anesthesia (23). The increased incidence of POD caused by prolonged postoperative ICU treatment and tracheal intubation time could be explained by the ICU environment, stress state during tracheal intubation and the severity of patients' disease *per se* (24). There is a clear correlation between the incidence of postoperative delirium and the degree of postoperative pain. Incomplete postoperative analgesia can enhance the patient's stress response and alter the transmission of neurotransmitters. When postoperative analgesia is insufficient, patients may experience anxiety, irritability, resistance to communication, decreased motor function, slow recovery of gastrointestinal function, and changes in sleep cycle, all of which are factors leading to the occurrence of POD. Studies suggested that postoperative pain management may help reduce the risk of postoperative delirium in the elderly patients (25). In importance ranking, the maximum VAS value of exercise pain within 7 days after surgery, as a characteristic variable, ranks second in contribution to the predictive variable. In addition, this study showed that the median and maximum VAS values of resting pain and exercise pain within 7 days after surgery were the most closely correlated with POD occurrence, and when the pain was controlled within the mild range, the risk of POD was lowered, the risk rising with VAS values. Finally, as reported by Bilotta et al. (26), type of surgery was strong predictor of POD and for some surgical procedures-including orthopedic, abdominal aortic aneurysm, and cardiac thoracic surgery-it links to an increased risk. Compared between cardiac surgery and non-cardiac surgeries, the Odd Ratio of predictors for POD was: 3.5 (1.6–7.4). Therefore, we focus only on the non-cardiac thoracic and abdominal surgeries to reduce the influence of POD incidence by surgery type.

For classification of the imbalanced data in this study caused by the extremely low positive sample number in the data set, the cross validation and Smote technique (Synthetic minority over-sampling technique, Smote) were used to balance the data set and ensure excellent classification results in minority classes during model sampling, via retaining the majority class units and synthesizing new minority class units linearly from those that were set close (27, 28). In RF modeling, the selected ensemble algorithms adopted the data classification strategy of constructing multiple weaker classifiers, combining them into classifiers with strong classifier generalization performance, and forcing the classifiers to focus on minority class samples in the algorithmic level, which is advantageous over the regular approach of establishing a single strong classifier with excellent generalization ability in the training set in terms of unbalanced data modeling (29, 30). Besides, accuracy was not used as the single evaluation indicator in this study, because the overall accuracy of the imbalanced data classification would not accurately reflect the classification situation in minority classes. Instead, confusion matrix parameters (accuracy, AUC-ROC, AUC-PRC, and MMC scores) were adopted to comprehensively evaluate the model (31).

In this study, Boruta was used to screen and include 16 characteristic variables into the prediction model RF where importance ranking and univariate partial dependence graph were made to enhance its intelligibility, visibility, and potential applicability in clinical practice (32). Boruta algorithm generated "shadow attribute" for each variable and calculated the Z-score value for each of them through RF model. When the Z-score value was significantly higher than the highest shadow attribute value, the input variable was viewed and retained as dependent variable related one (33). Boruta



follows all relevant feature selection methods and can capture all features related to the result variable. In contrast, most traditional feature selection algorithms follow a minimum optimization method, relying on a small subset of features and resulting in minimal errors in selecting classification. This method minimizes the error of the model to the greatest extent possible, which will ultimately form a minimum optimal feature subset. This occurs by selecting an overly condensed version of the input dataset, which in turn may result in the loss of some relevant features. On the other hand, Boruta finds all features, regardless of their correlation with the decision variable. This makes it very suitable for application in the field of biomedicine. In this article, POD related risk factors were screened and identified using Boruta, offering guidance for clinicians to take timely intervention measures for high-risk patients and reduce POD occurrence.

The challenges of applying machine learning lie primarily in the lack of interpretability and repeatability of machine learning-generated results, which may limit their application. Interpretable machine learning can effectively open the “black box” of machine learning (32, 34). In this study, the degree of contribution of each feature variable was explained through an importance sorting chart, and the trend of the result variable changing with the feature variable was explained through a univariate partial dependency profile and visualization prediction of random individual samples through a

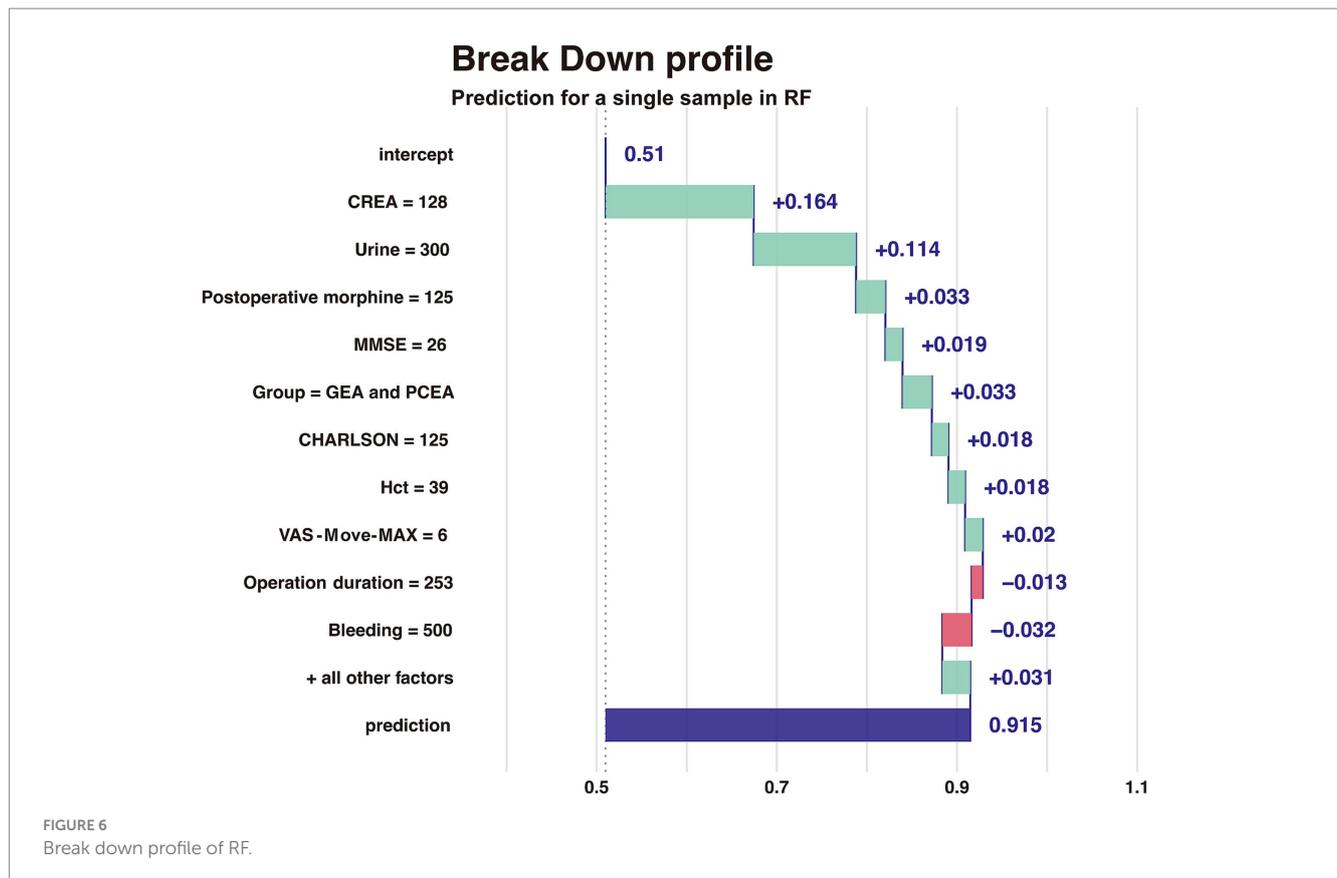
breakdown profile. This solves the problem of lack of interpretability in predictive models.

The following are the weaknesses of the present study that may have affected our results. Firstly, we included multiple risk factors, but did not include laboratory data. Secondly, POD subtypes can be divided into low, high, and mixed types, which we will continue to explore in subsequent studies. Thirdly, this article uses SMOTE technique to process imbalanced datasets, improving model performance while also potentially generating noise. Finally, this model requires an independent dataset to test its extrapolation and generalization capabilities. In the future, we will collect sufficient external validation datasets to further improve this model.

In this study, the major risk factors for the development of postoperative delirium are CREA and VAS-Move-Max. Machine learning algorithm can be established to predict the occurrence of postoperative delirium for older patients who underwent non-cardiac thoracic or abdominal surgery with general anesthesia.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding authors.



Ethics statement

The studies involving humans were approved by The trial protocol was approved by the Institutional Review Committee of Peking University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

WS: Writing – original draft, Formal analysis. XT: Formal analysis, Writing – review & editing. XH: Writing – review & editing, Data curation. PL: Data curation, Writing – review & editing. LL: Writing – review & editing, Formal analysis. HM: Funding acquisition, Writing – original draft. DW: Supervision, Writing – review & editing. TL: Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work is supported by the National Natural Science Foundation of China (82071180 and 82271206) and Natural Science Foundation of Beijing (7212023).

Acknowledgments

We thank DW and his study group for providing the database.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2023.1325941/full#supplementary-material>

References

- Buchan TA, Sadeghirad B, Schmutz N, Goettel N, Foroutan F, Couban R, et al. Preoperative prognostic factors associated with postoperative delirium in older people undergoing surgery: protocol for a systematic review and individual patient data meta-analysis. *Syst Rev.* (2020) 9:261. doi: 10.1186/s13643-020-01518-z
- Wang YY, Yue JR, Xie DM, Carter P, Li QL, Gartaganis SL, et al. Effect of the tailored, family-involved hospital elder life program on postoperative delirium and function in older adults: a randomized clinical trial. *JAMA Intern Med.* (2020) 180:17–25. doi: 10.1001/jamainternmed.2019.4446
- Salvi F, Young J, Lucarelli M, Aquilano A, Luzi R, Dell'Aquila G, et al. Non-pharmacological approaches in the prevention of delirium. *Eur Geriatr Med.* (2020) 11:71–81. doi: 10.1007/s41999-019-00260-7
- Mart MF, Williams Roberson S, Salas B, Pandharipande PP, Ely EW. Prevention and management of delirium in the intensive care unit. *Semin Respir Crit Care Med.* (2021) 42:112–26. doi: 10.1055/s-0040-1710572
- Li YW, Li HJ, Li HJ, Zhao BJ, Guo XY, Feng Y, et al. Delirium in older patients after combined epidural-general anesthesia or general anesthesia for major surgery: a randomized trial. *Anesthesiology.* (2021) 135:218–32. doi: 10.1097/ALN.0000000000003834
- Ely EW, Inouye SK, Bernard GR, Gordon S, Francis J, May L, et al. Delirium in mechanically ventilated patients: validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU). *JAMA.* (2001) 286:2703–10. doi: 10.1001/jama.286.21.2703
- Wang C, Wu Y, Yue P, Ely EW, Huang J, Yang X, et al. Delirium assessment using confusion assessment method for the intensive care unit in Chinese critically ill patients. *J Crit Care.* (2013) 28:223–9. doi: 10.1016/j.jccr.2012.10.004
- Wang W, Li HL, Wang DX, Zhu X, Li SL, Yao GQ, et al. Haloperidol prophylaxis decreases delirium incidence in elderly patients after noncardiac surgery: a randomized controlled trial*. *Crit Care Med.* (2012) 40:731–9. doi: 10.1097/CCM.0b013e3182376e4f
- Mu DL, Wang DX, Li LH, Shan GJ, Li J, Yu QJ, et al. High serum cortisol level is associated with increased risk of delirium after coronary artery bypass graft surgery: a prospective cohort study. *Crit Care.* (2010) 14:R238. doi: 10.1186/cc9393
- Adnan M, Alarood AAS, Uddin MI, Ur Rehman I. Utilizing grid search cross-validation with adaptive boosting for augmenting performance of machine learning models. *PeerJ Comput Sci.* (2022) 8:e803. doi: 10.7717/peerj-cs.803
- Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* (2002) 16:321–57. doi: 10.1613/jair.953
- Hui H, Wang W, Mao B (2005). "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning" in *Advances in Intelligent Computing, International Conference on Intelligent Computing, ICIC 2005, Hefei, China.* August 23–26, 2005, Proceedings, Part I. 3644: 878–887.
- Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ.* (2020) 368:m441. doi: 10.1136/bmj.m441
- van Smeden M, Moons KG, de Groot JA, Collins GS, Altman DG, Eijkemans MJC, et al. Sample size for binary logistic prediction models: beyond events per variable criteria. *Stat Methods Med Res.* (2019) 28:2455–74. doi: 10.1177/0962280218784726
- Riley RD, Van Calster B, Collins GS. A note on estimating the cox-Snell R2 from a reported C statistic (AUROC) to inform sample size calculations for developing a prediction model with a binary outcome. *Stat Med.* (2021) 40:859–64. doi: 10.1002/sim.8806
- Liu J, Li J, Gao D, Wang J, Liu M, Yu D. High ASA physical status and low serum uric acid to creatinine ratio are independent risk factors for postoperative delirium among older adults undergoing urinary calculi surgery[J]. *Clin Interv Aging.* (2023) 18:81–92. doi: 10.2147/CIA.S395893
- Hu XY, Liu H, Zhao X, Sun X, Zhou J, Gao X, et al. Automated machine learning-based model predicts postoperative delirium using readily extractable perioperative collected electronic data. *CNS Neurosci Ther.* (2022) 28:608–18. doi: 10.1111/cns.13758
- Brown CH, Edwards C, Lin C, Jones EL, Yanek LR, Esmaili M, et al. Spinal anesthesia with targeted sedation based on bispectral index values compared with general anesthesia with masked Bispectral index values to reduce delirium: the SHARP randomized controlled trial. *Anesthesiology.* (2021) 135:992–1003. doi: 10.1097/ALN.0000000000004015
- Kung WM, Yuan SP, Lin MS, Wu CC, Islam MM, Atique S, et al. Anemia and the risk of cognitive impairment: An updated systematic review and meta-analysis. *Brain Sci.* (2021) 11:777. doi: 10.3390/brainsci11060777
- Gao H, Ma HJ, Li YJ, Yin C, Li Z. Prevalence and risk factors of postoperative delirium after spinal surgery: a meta-analysis. *J Orthop Surg Res.* (2020) 15:138. doi: 10.1186/s13018-020-01651-4
- Wang A, Peng W, Zhang L, Huang S. Application of phenylephrine combined with goal-directed fluid therapy in elderly patients undergoing hip arthroplasty: a randomized controlled trial. *Altern Ther Health Med.* (2022) 28:132–8.
- Lin X, Chen Y, Zhang P, Chen G, Zhou Y, Yu X. The potential mechanism of postoperative cognitive dysfunction in older people. *Exp Gerontol.* (2020) 130:110791. doi: 10.1016/j.exger.2019.110791
- Jipa M, Isac S, Klimko A, Simion-Cotorogea M, Martac C, Cobilinschi C, et al. Opioid-sparing analgesia impacts the perioperative anesthetic Management in Major Abdominal Surgery[J]. *Medicina.* (2022) 58:487. doi: 10.3390/medicina58040487
- Khaled M, Sabac D, Marcucci M. Postoperative pain and pain management and neurocognitive outcomes after non-cardiac surgery: a protocol for a series of systematic reviews[J]. *Syst Rev.* (2022) 11:280. doi: 10.1186/s13643-022-02156-3
- Halaszynski TM. Pain management in the elderly and cognitively impaired patient: the role of regional anesthesia and analgesia. *Curr Opin Anaesthesiol.* (2009) 22:594–9. doi: 10.1097/ACO.0b013e32833020dc
- Bilotta F, Lauretta MP, Borozdina A, Mizikov VM, Rosa G. Postoperative delirium: risk factors, diagnosis and perioperative care. *Minerva Anesthesiol.* (2013) 79:1066–76.
- Hasib KM, Towhid NA, Faruk KO, Al Mahmud J, Mridha MF. Strategies for enhancing the performance of news article classification in Bangla: handling imbalance and interpretation. *Eng Appl Artif Intell.* (2023) 125:106688. doi: 10.1016/j.engappai.2023.106688
- Hasib K.M., Iqbal M.S., Shah F.M., Mahmud J.A., Popel M.H., Showrov M.I.H., et al. (2020). A survey of methods for managing the classification and solution of data imbalance problem. arXiv [Preprint]. doi: 10.3844/jcsp.2020.1546.1557
- Wu Z, Lin W, Ji Y. An integrated ensemble learning model for imbalanced fault diagnostics and prognostics. *NJ IEEE Access.* (2018) 6:8394–402. doi: 10.1109/ACCESS.2018.2807121
- Hasib KM, Islam MR, Sakib S, Akbar MA, Razzak I, Alam MS. Depression detection from social networks data based on machine learning and deep learning techniques: An interrogative survey. *IEEE Trans Comput Soc Syst.* (2023) 10:1568–86. doi: 10.1109/TCSS.2023.3263128
- Gao Q, Jin X, Xia E, Wu X, Gu L, Yan H, et al. Identification of orphan genes in unbalanced datasets based on ensemble learning. *Front Genet.* (2020) 11:820. doi: 10.3389/fgene.2020.00820
- Petch J, Di S, Nelson W. Opening the black box: the promise and limitations of explainable machine learning in cardiology. *Can J Cardiol.* (2022) 38:204–13. doi: 10.1016/j.cjca.2021.09.004
- Kursa MB, Rudnicki WR. Feature selection with the Boruta package. *J Stat Softw.* (2010) 36:1–13. doi: 10.18637/jss.v036.i11
- Ayano YM, Schwenker F, Duferra BD, Debeleg TG. Interpretable machine learning techniques in ECG-based heart disease classification: a systematic review. *Diagnostics.* (2022) 13:111. doi: 10.3390/diagnostics13010111