



OPEN ACCESS

EDITED BY

Alessandro Gialluisi,
LUM University "Giuseppe Degennaro", Italy

REVIEWED BY

Vahid Rashedi,
University of Social Welfare and Rehabilitation
Sciences, Iran
Antonietta Pepe,
LUM University "Giuseppe Degennaro", Italy

*CORRESPONDENCE

Tianyao Zhang
✉ tianyaozhang123@163.com

†These authors have contributed equally to
this work

RECEIVED 01 April 2025

ACCEPTED 29 July 2025

PUBLISHED 18 August 2025

CITATION

Qian Z, Wu X, He K, Lin K, Luo X and Zhang T
(2025) Application of machine learning in
predicting perioperative neurocognitive
disorders in elderly patients: the impact of
sarcopenia-related features.
Front. Med. 12:1604333.
doi: 10.3389/fmed.2025.1604333

COPYRIGHT

© 2025 Qian, Wu, He, Lin, Luo and Zhang.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Application of machine learning in predicting perioperative neurocognitive disorders in elderly patients: the impact of sarcopenia-related features

Zhengyu Qian^{1†}, Xiaochu Wu^{1,2,3†}, Kunyang He¹, Kaijie Lin¹,
Xiaobei Luo⁴ and Tianyao Zhang^{1,2*}

¹School of Clinical Medicine, Chengdu Medical College, Chengdu, China, ²The First Affiliated Hospital of Chengdu Medical College, Chengdu, China, ³National Clinical Research Center for Geriatrics, West China Hospital, Sichuan University, Chengdu, Sichuan, China, ⁴Suzhou Medical College, Soochow University, Suzhou, Jiangsu, China

Background: Older surgical patients present with diverse clinical profiles, yet research indicates a significant correlation between sarcopenia-related features and the incidence of perioperative neurocognitive disorder (PND). The integration of machine learning techniques offers a promising avenue for identifying older surgical patients at elevated risk of PND, particularly those exhibiting sarcopenia-associated characteristics. This approach enhances preoperative risk stratification and patient selection, thereby improving the precision of clinical management and treatment decisions.

Methods: Data were collected from patients undergoing non-cardiac surgery at the First Affiliated Hospital of Chengdu Medical College to develop and validate a predictive model. Five machine learning models—Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Gradient Boosting Machine (GBM), Adaptive Boosting (AdaBoost), and Random Forest—were constructed to evaluate the risk of PND in older surgical patients. Sarcopenia-related features were incorporated as key variables in these models. The SHapley Additive exPlanations (SHAP) method was subsequently utilized to interpret the most effective model.

Results: A total of 443 patients were included in the study. Among the five models, AdaBoost performed best, achieving an AUC of 0.95. The six most important features identified by SHAP were 6-meter walking speed, preoperative MMSE score, maximum grip strength, appendicular skeletal muscle mass, and sarcopenia assessment age. These results demonstrate AdaBoost's excellent predictive performance, with high interpretability and reliability.

Conclusion: Machine learning models, particularly AdaBoost integrated with SHAP, show significant potential in predicting PND in older surgical patients. The model's ability to clarify the impact of sarcopenia-related features enhances its clinical utility in preoperative risk assessment.

KEYWORDS

machine learning, sarcopenia, postoperative cognitive dysfunction, SHapley Additive exPlanations, perioperative neurocognitive disorders

1 Introduction

Perioperative Neurocognitive Disorders (PND) encompass postoperative cognitive impairments such as memory, attention, and language deficits, predominantly observed in older surgical patients (1, 2). These disorders often manifest initially as postoperative delirium (POD) or emergence agitation, evolving into more chronic forms such as postoperative cognitive dysfunction (POCD) (3). PND is associated with prolonged hospitalization, increased postoperative complications, and significantly diminishes the quality of life, imposing substantial burdens on families and societal resources (4).

A global increase in the elderly population (5) corresponds with heightened surgical demand (6), notably among those over 60, with dementia symptoms and PND rates reaching up to 65% (7). Research indicates that 41.4% of patients older than 60 suffer from PND post-surgery, with a notable percentage not fully recovering within 3 months (8), underscoring the necessity for preemptive risk management in anesthesia.

Sarcopenia, defined by progressive loss of skeletal muscle mass and strength, occurs in up to 33% of individuals over 65 and significantly correlates with PND (9). This condition contributes to cognitive decline through mechanisms such as reduced neurogenesis and lower levels of critical neurotrophic factors, including brain-derived neurotrophic factor (BDNF) (10). Furthermore, sarcopenia is associated with prolonged postoperative recovery, heightened susceptibility to complications, including increased infection risk and extended hospitalization periods. Notably, one study documented a 15% increase in PND incidence among patients exhibiting sarcopenia prior to surgery (11). An animal study also underscored reduced skeletal muscle content as a significant preoperative risk factor for PND (10). Therefore, factors affecting muscle mass, such as inflammation and insulin resistance, are closely associated with PND, suggesting that sarcopenia-related features could be critical risk factors for PND in the older adult (12, 13).

Machine Learning (ML) advancements have significantly transformed medical diagnostics and treatment planning. ML algorithms surpass traditional logistic regression in detecting intricate non-linear relationships and managing complex datasets, thereby enhancing prediction accuracy for clinical outcomes. Current research predominantly targets older surgical patients undergoing cardiac and non-cardiac surgeries, with ML models evaluating factors such as type of surgery, preoperative cognitive status, and anesthesia methods to provide a reliable level of predictive accuracy (14, 15). Despite progress, research tailored to high-risk elderly populations with specific risk profiles, such as those with sarcopenia, remains limited.

This study develops an ML-based predictive model targeting sarcopenia-related risk factors, aiming to optimize perioperative management and enhance postoperative outcomes for older surgical patients. By conducting detailed preoperative assessments, the proposed model intends to predict PND risk accurately, thus refining management strategies and enhancing the postoperative quality of life for these high-risk groups.

2 Materials and methods

2.1 Ethical approval

This study adhered to the principles of the Declaration of Helsinki and was approved by the Medical Ethics Committee of the First Affiliated Hospital of Chengdu Medical College (Approval No: 2021CYFYIRB-BA-67-01) on November 28, 2021. All participants provided written informed consent prior to their inclusion in the study.

2.2 Study design and population

In this prospective cohort study, clinical electronic medical records and relevant data were systematically collected from patients undergoing surgical procedures at the First Affiliated Hospital of Chengdu Medical College via the iMedical system between March 1, 2022, and February 28, 2023. A total of 480 patients were initially enrolled, of whom 443 satisfied the inclusion criteria for model development and internal validation. Preoperative demographic and clinical characteristics were documented at the time of patient admission, while perioperative treatment data and postoperative clinical outcomes were meticulously recorded thereafter. All predictive variables were gathered prior to the onset of perioperative neurocognitive disorder (PND) to facilitate the adjustment of perioperative management strategies by anesthesiologists, based on real-time risk evaluations.

Inclusion criteria included patients aged ≥ 60 years undergoing elective non-cardiac surgery, with specific sarcopenia-related indicators (e.g., appendicular skeletal muscle mass) collected. Exclusion criteria comprised inability to complete gait speed and grip strength assessments; presence of severe systemic diseases affecting major organs (heart, lungs, liver, kidneys); visual, hearing, or communication impairments hindering questionnaire completion or examination cooperation; history of significant preoperative mental illness or comprehensive neurological disorders, such as cognitive impairment; and cases with severe data loss in the study.

2.3 Diagnostic criteria for PND and sarcopenia

Given the absence of a universal diagnostic standard for PND, this study defined PND as a decrease of ≥ 2 points in the Mini-Mental State Examination (MMSE) score 3 days postoperatively compared to the preoperative score (16). MMSE scores were measured 1 day before surgery and on the third postoperative day for all cases included in this study. The diagnostic criteria for sarcopenia were based on the 2019 standards established by the Asian Working Group for Sarcopenia (AWGS) (17): 1. Muscle strength: grip strength (men < 28 kg, women < 18 kg); 2. Physical function: 6-meter walking speed < 1.0 m/s; 3. Appendicular skeletal muscle mass: BIA (men < 7.0 kg/m², women < 5.7 kg/m²).

2.4 Feature selection and model development

In the dataset used in this study, only the variable “age” had missing values, affecting two patients (0.45% of the total sample); all other variables were complete. Outliers in continuous variables were identified using the interquartile range (IQR) method and removed to reduce skewness. Continuous variables were normalized using min-max scaling for comparability across models. Missing values were imputed using multivariate imputation by chained equations (MICE, 4 iterations) (38).

Initially, we performed univariate regression analysis on each variable. For categorical variables, differences between the two groups were compared using Fisher’s exact test or chi-square test, depending on sample size and expected frequency. For continuous variables following a normal distribution, an independent samples *t*-test was used for inter-group comparisons; for continuous variables not following a normal distribution, the Mann-Whitney *U*-test (rank-sum test) was employed. The Shapiro-Wilk test was used to assess the normality of the continuous variables to determine the appropriate statistical methods. A *P*-value <0.05 was considered statistically significant.

We conducted logistic regression analysis on important variables from the univariate analysis to identify significant features for subsequent evaluation and construction of ML predictive models. The original dataset was randomly divided into a test set (30%) and a training set (70%). The training set was used to train the ML models, while the test set was used to adjust model parameters and evaluate model performance. All significant features identified by logistic regression were incorporated into five ML models to predict PND. These models included Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Gradient Boosting Machine (GBM), Adaptive Boosting (AdaBoost), and Random Forest. To avoid overfitting, we used a grid search algorithm with 10-fold cross-validation to optimize the training set and find the best hyperparameters for training our ML models, ensuring good generalization ability. Stratified sampling was used during cross-validation to maintain the class distribution between PND and non-PND cases, thereby mitigating imbalance without the need for synthetic oversampling.

For the AdaBoost model, we performed hyperparameter tuning using grid search with 10-fold cross-validation. The tuning grid included 27 combinations of the number of iterations (50, 100, 150), maximum tree depth (1, 2, 3), and learning rate ($\nu = 0.1, 0.5, 1$).

While AUC was the primary metric for model evaluation, F1 score was used as the guiding metric during hyperparameter tuning, as it provides a more balanced assessment under class imbalance. The best-tuned model achieved an AUC of 0.952, substantially outperforming the default AdaBoost configuration (AUC = 0.910) under identical validation conditions.

2.5 Model evaluation and interpretation

To evaluate the robustness of the model with respect to missing data imputation, we conducted a sensitivity analysis comparing

the full dataset ($n = 443$) and the complete-case dataset ($n = 441$). Excluding the two patients with missing age values led to a decrease in AUC from 0.952 to 0.892. Both excluded cases were PND-positive and exhibited atypical profiles, including high MMSE scores and good physical performance, suggesting that their inclusion enhanced model discrimination. These findings support the stability of the AdaBoost model and justify the use of the imputed dataset for subsequent model training and interpretation. Details are provided in [Supplementary Table S1](#) and [Supplementary Figure S1](#).

Our evaluation of model performance focused on discrimination and calibration. We used the area under the receiver operating characteristic curve (AUROC) to reflect discriminative ability and described the statistical differences in AUROC using the Delong test (18). Additionally, when evaluating the performance of machine learning models, we calculated accuracy, precision, recall, and F1 score, and referenced the Youden index to optimize and select the best cutoff value to balance sensitivity and specificity. These metrics comprehensively measure model performance from different perspectives, ensuring that we can identify the strengths and weaknesses of the models in various application scenarios and make specific improvements to optimize overall performance. The interpretability of machine learning models has always been challenging. To further explain the impact and contribution of each feature variable to the final model, we used the SHAP method to interpret the best-performing black-box model (19). This ranking process is based on the mean absolute SHAP values of all individuals. These interpretability techniques were implemented in R using the “iml” package version 0.11.3.

2.6 Language editing assistance

During the preparation of this manuscript, we used ChatGPT to assist with language editing and to improve the readability of the text. All scientific content and data analysis were conducted independently by the authors.

3 Results

3.1 Patient characteristics

A total of 443 older surgical patients were included in this study, of whom 121 developed PND, resulting in an incidence rate of 27.3%. The Activities of Daily Living (ADL) scale is commonly used to assess an individual’s ability to perform basic and instrumental activities, with a score of 100 indicating good self-care ability. The Instrumental Activities of Daily Living (IADL) scale assesses the ability to perform more complex daily activities. The total IADL score ranges from 0 to 8, with a score of 8 indicating that the individual can independently complete all activities, demonstrating a high level of self-care ability. A frailty score >0 indicates some degree of frailty or a pre-frailty state. Detailed clinical characteristics are presented in [Table 1](#).

TABLE 1 *P*-values were obtained by comparing the PND and non-PND groups using the Mann-Whitney *U*-test a; Chi-square test or Fisher's exact test b; independent samples *t*-test c; Variables marked with * were included in the logistic regression model (*P* < 0.05).

Variables	PND (<i>N</i> = 121, 27.3%)	Non-PND (<i>N</i> = 322, 72.6%)	<i>P</i> -values
Demographics			
Age, year, median (IQR)	73 (69–73)	71 (68–75)	0.020 ^{a*}
Gender, <i>n</i> (%)			0.383 ^{b*}
Male	67 (52.9)	187 (41.9)	
Female	54 (47.1)	135 (58.1)	
Activities of daily living, <i>n</i> (%)			
ADL score (<100)	14 (11.6)	30 (9.3)	<0.001 ^{b*}
IADL score (<8)	52 (43.0)	142 (44.1)	<0.001 ^{b*}
Frailty level (>0)	60 (49.6)	126 (39.1)	0.139 ^b
Frailty score (>0)	72 (59.5)	165 (51.2)	0.002 ^{a*}
Preoperative cognitive function variables, median (IQR)			
Preoperative orientation	9 (7.0–10)	8 (7.0–10.0)	0.032 ^{b*}
Preoperative immediate memory	3 (3.0–3.0)	3 (3.0–3.0)	<0.001 ^{b*}
Preoperative attention and calculation	5 (3.0–5.0)	5 (4.0–5.0)	<0.001 ^{b*}
Preoperative delayed memory	3 (3.0–3.0)	3 (3.0–3.0)	0.482 ^{b*}
Preoperative language ability	7 (5.0–8.0)	7 (5.0–8.0)	0.605 ^{b*}
Preoperative visuospatial ability	1 (0.0–1.0)	1 (0.0–1.0)	<0.001 ^{b*}
Preoperative cognitive impairment present	0 (0.0–0.0)	0 (0.0–0.0)	1 ^{b*}
Preoperative MMSE score	26 (24.0–28.0)	26 (24.0–27.0)	0.15 ^{b*}
Physical function variables			
Mini Nutritional Assessment, median (IQR)	13 (11.0–14.0)	13 (11.0–14.0)	0.436 ^a
6-meter walking speed, median (IQR)	0.81 (0.62–0.94)	0.95 (0.79–1.10)	<0.001 ^{c*}
Maximum grip strength, median (IQR)	23.5 (17.2–26.9)	24.9 (20.33–31.48)	<0.001 ^{c*}
Appendicular skeletal muscle mass, median (IQR)	6.5 (5.80–7.31)	7.0 (6.23–7.77)	<0.001 ^{c*}
Sarcopenia assessment, <i>n</i> (%)	58 (47.9)	65 (20.2)	<0.001 ^{b*}
Anesthesia and surgery-related variables			
ASA classification, <i>n</i> (%)			0.367 ^b
I	69 (57.0)	183 (56.8)	
II	51 (42.1)	139 (43.2)	
III	1 (0.09)	0 (0.0)	

(Continued)

TABLE 1 (Continued)

Variables	PND (<i>N</i> = 121, 27.3%)	Non-PND (<i>N</i> = 322, 72.6%)	<i>P</i> -values
Use of dexmedetomidine, <i>n</i> (%)	37 (30.6)	103 (32.0)	0.865 ^{b*}
Combined nerve block, <i>n</i> (%)	52 (43.0)	144 (44.7)	0.824 ^b
Type of anesthesia, <i>n</i> (%)			<0.001 ^{b*}
Local anesthesia	3 (2.5)	1 (0.04)	
Subarachnoid anesthesia	32 (26.5)	71 (22.0)	
General anesthesia	86 (71.0)	250 (77.6)	
Use of analgesic pump, <i>n</i> (%)	63 (52.1)	182 (56.5)	<0.001 ^{b*}
Duration of surgery, min, median (IQR)	106 (64–185)	110 (63–210)	0.564 ^{a*}
Duration of anesthesia, min, median (IQR)	150 (101–237)	150 (100–264)	0.455 ^{a*}
Blood loss, ml, median (IQR)	20 (10–50)	20 (10–50)	0.999
Fluid infusion, ml, median (IQR)	1100 (1000–2050)	1200 (1000–2100)	0.447
Medical history, <i>n</i> (%)			
Diabetes	37 (30.5)	85 (26.4)	0.448 ^{b*}
Hypertension	58 (47.9)	127 (39.4)	<0.001 ^{b*}

3.2 Model evaluation

Following logistic regression analysis, we incorporated 22 variables into the model, including: age, ADL scale, IADL scale, frailty score, preoperative orientation, preoperative immediate memory, preoperative attention and calculation, preoperative delayed memory, preoperative language ability, preoperative visuospatial ability, preoperative presence of cognitive impairment, preoperative MMSE score, 6-meter walking speed, maximum grip strength, appendicular skeletal muscle mass, sarcopenia assessment, hypertension, anesthesia type, dexmedetomidine usage, analgesia pump usage, surgery duration, and anesthesia duration. Five algorithms were employed: SVM, XGBoost, GBM, AdaBoost, and Random Forest. The results indicated that the AdaBoost model achieved the highest area under the receiver operating characteristic curve (AUC) on the test set (Figure 1). Moreover, the AdaBoost model outperformed the other models in all other performance metrics. The details of the specific model parameters developed using different algorithms are shown in Table 2.

To provide a conventional benchmark, we additionally evaluated a logistic regression model using the same dataset and 10-fold cross-validation strategy. The model yielded an AUC of 0.708 (95% CI: 0.653–0.763), which was substantially lower than the performance of the AdaBoost model (AUC = 0.952, 95% CI: 0.918–0.985). This result highlights the added predictive

value of the machine learning approach over traditional linear modeling techniques.

3.3 Model interpretation

To better understand the relationship between the model and the data, we used the SHAP method to visually interpret the best performing AdaBoost model, illustrating how each variable affects the occurrence of PND. SHAP values were used to explain the 15 assessed risk factors. As shown in Figure 2, the SHAP values on the X-axis represent a unified metric that shows how each feature influences the model’s outcome. For each feature, the impact on the outcome is represented by colored dots, with high feature values indicated by yellow and low feature values by purple. The Y-axis lists the 15 factors that are statistically significant for the PND outcome.

Higher values of features (points closer to yellow) such as maximum grip strength, appendicular skeletal muscle mass, preoperative MMSE score, and 6-meter walking speed were generally associated with negative SHAP values, indicating a protective effect against PND. In contrast, features like older

age, longer surgery and anesthesia duration, hypertension, general anesthesia, and sarcopenia assessment showed positive SHAP contributions, indicating increased risk. Other variables (e.g., IADL, preoperative language ability, immediate memory) exhibited minimal or inconsistent SHAP effects.

To further assess non-linear and interaction effects revealed by the SHAP analysis, we examined the dependence plot for age. The SHAP value for age showed a non-monotonic trend—rising up to ~80 years, then plateauing or slightly declining—possibly reflecting survivor bias or physiological adaptation in the oldest patients.

We also explored the interaction between age and grip strength (Figure 3). Patients with lower grip strength tended to show relatively higher SHAP values for age, suggesting a possible interaction pattern. While this trend was not consistent across all data points, the visual distribution implies that reduced muscle strength may be associated with increased age-related vulnerability to PND. This observation highlights the potential role of sarcopenia not only as an independent risk factor but also as a contributor to age-related susceptibility.

3.4 Comparative AUC performance and DeLong tests

To further quantify performance differences, we compared the AUC values of all models using the DeLong test. As shown in Table 3, AdaBoost significantly outperformed all other models ($P < 0.001$). The logistic regression model yielded an AUC of 0.708 (95% CI: 0.653–0.763), whereas the AUCs for XGBoost, GBM, SVM, and Random Forest ranged from 0.746 to 0.775.

4 Discussion

In this retrospective cohort study, we developed a model to predict the risk of PND among older surgical patients, specifically incorporating sarcopenia-related indicators. We used 23 important features to train and validate the model using SVM, XGBoost, GBM, AdaBoost, and Random Forest. Among the different models compared, the AdaBoost model showed the best predictive power, with an AUC of 0.95, precision of 0.89, sensitivity of 0.95.

Data imbalance can significantly affect the accuracy of predictive models. In our study cohort of elderly individuals, the relatively low prevalence of PND (incidence rate of 27.3%) created a notable disparity between normal and affected samples. This class imbalance means that a model could theoretically

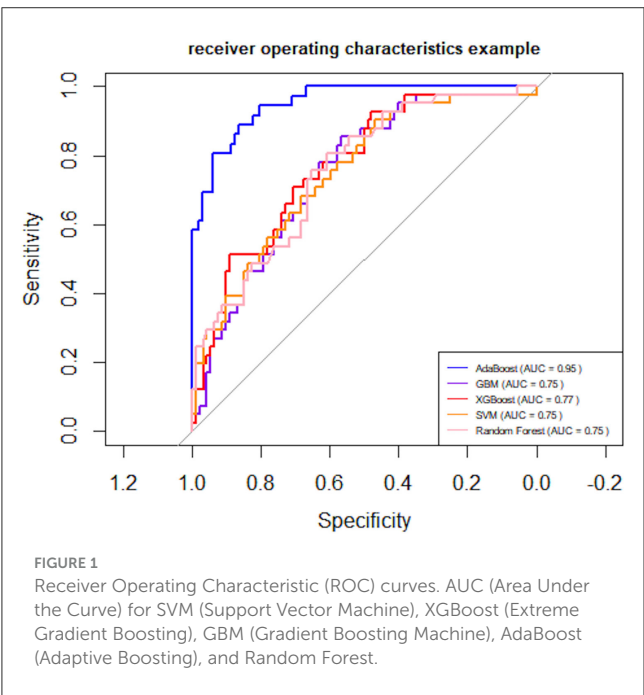
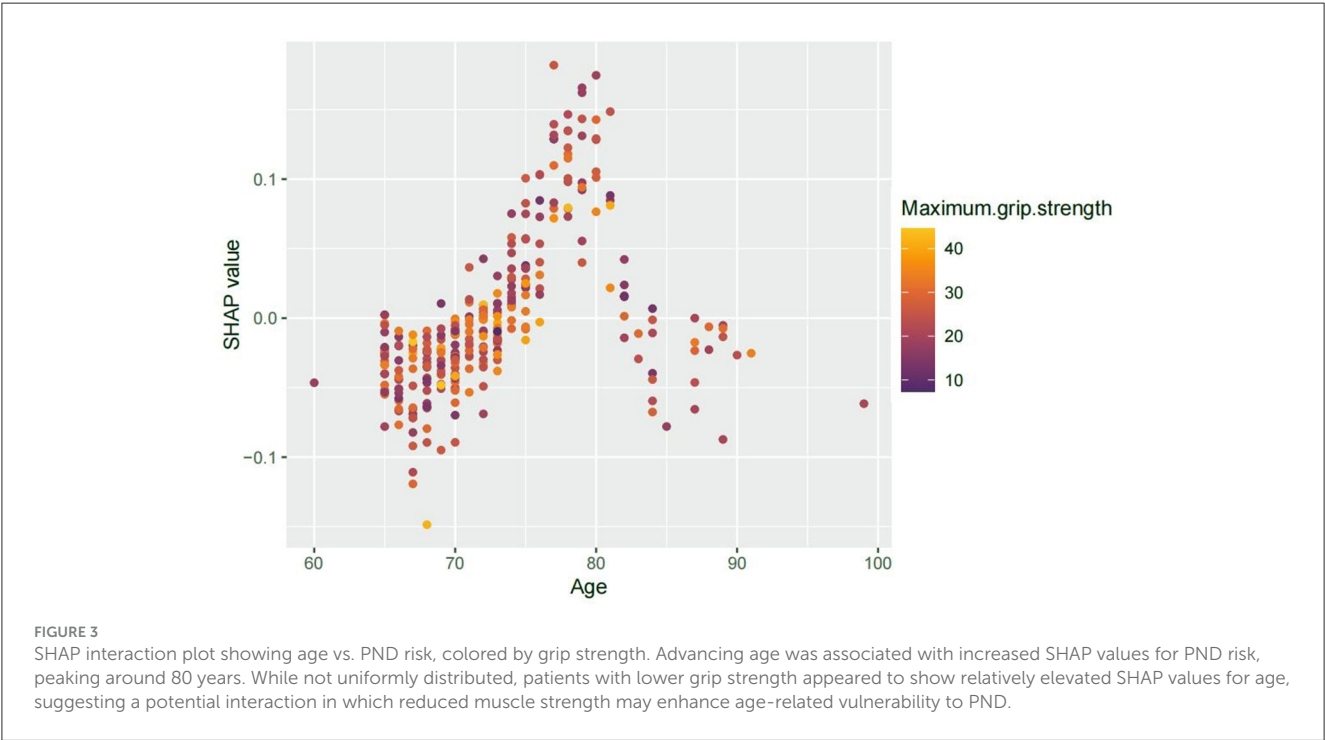


TABLE 2 Performance Metrics of Different Machine Learning Models for PND.

Model	Accuracy (95% CI)	Precision (95% CI)	Recall (95% CI)	F1 Score (95% CI)
AdaBoost	0.88 (0.82-0.93)	0.89 (0.83-0.95)	0.95 (0.91-0.99)	0.92 (0.87-0.97)
XGBoost	0.77 (0.70-0.84)	0.79 (0.72-0.86)	0.90 (0.85-0.95)	0.84 (0.78-0.90)
GBM	0.69 (0.61-0.77)	0.69 (0.62-0.77)	0.99 (0.97-1.01)	0.82 (0.75-0.88)
SVM	0.70 (0.62-0.78)	0.70 (0.62-0.78)	0.99 (0.97-1.01)	0.82 (0.75-0.88)
RF	0.74 (0.66-0.81)	0.75 (0.67-0.82)	0.93 (0.89-0.98)	0.83 (0.77-0.89)



achieve an accuracy of 88% by predicting all outcomes as normal. However, such predictions are misleading, as they fail to adequately identify PND cases, leading to a bias toward predicting normal outcomes and reducing precision. Therefore, using the F1 score (0.95) provides a more balanced evaluation of model performance in cases with significant sample imbalance. In contrast, accuracy

can be misleading in imbalanced datasets, as it may overestimate performance by favoring the majority class. In this study, we used five different machine learning algorithms, achieving AUC values from 0.75 to 0.95. Initially, the inclusion of only statistically significant features ($P < 0.05$) yielded a lower AUC, as these variables alone did not

TABLE 3 AUC values represent model discrimination performance in the test set.

Model	AUC	P-value vs. AdaBoost
AdaBoost	0.95	-
XGBoost	0.77	0.0002
GBM	0.75	2.7e-05
SVM	0.75	4.6e-05
RF	0.75	1.6e-05

P-values were obtained by comparing each model with the AdaBoost model using the DeLong test.

capture the complexity of interactions affecting PND risk. Adding statistically insignificant variables helped the model account for subtle relationships, improving performance. Specifically, adding variables such as dexmedetomidine usage, anesthesia duration, and surgery duration ($P > 0.05$) improved the model's AUC.

To evaluate potential multicollinearity, we computed GVIFs for the retained non-significant variables. All $GVIF^{(1/2Df)}$ values were below 2, except for anesthesia duration (4.30), which was retained due to its clinical relevance. As AdaBoost is a tree-based method inherently robust to multicollinearity, model integrity was not compromised (Supplementary Table S2). Additionally, we conducted a likelihood ratio test comparing nested logistic regression models—one with only significant variables and another including all 22 features. While the full model showed reduced residual deviance, the difference was not statistically significant ($P = 0.136$). Nonetheless, these variables may enhance non-linear models like AdaBoost by capturing complex interactions and improving predictive performance.

Research shows dexmedetomidine reduces perioperative neurocognitive disorders by inhibiting inflammation and enhancing neuroprotection in older surgical patients and animal models (20). Dexmedetomidine, being negatively correlated with PND, helps to better differentiate between patient outcomes, thus improving the model's accuracy. Additionally, surgical trauma and prolonged anesthesia significantly increase PND and neuroinflammation risks (21, 22), suggesting that using a diverse set of variables could help uncover the intricate relationships between different factors (23).

On the other hand, the potential multicollinearity among feature variables could render the model parameter estimates unstable (24). Additionally, including more variables increases model complexity, which may improve data fitting but also heightens the risk of overfitting. To reduce overfitting and improve stability, we applied ten-fold cross-validation and adjusted parameters to ensure consistent model performance (25, 26).

Results showed the AdaBoost model achieved high AUC and accuracy in internal tests. This efficacy stems from AdaBoost's ability to handle small, imbalanced datasets effectively (27). However, concerns about the generalizability of models trained solely on internal data remain. Thus, collecting and evaluating external datasets is important for thoroughly assessing the model's predictive accuracy. Selecting features carefully is also vital for developing reliable predictive models.

In summary, despite challenges related to data imbalance and the diverse characteristics of older surgical patients, our findings demonstrate that machine learning can effectively build models to predict PND. By addressing class imbalances, incorporating sarcopenia-related indicators, and improving model accuracy and generalizability, we have developed a model with strong predictive power. This model provides a reliable risk assessment tool for older surgical patients, particularly those at risk due to sarcopenia.

Nevertheless, the present study has several limitations. Although the AdaBoost model achieved excellent performance (AUC = 0.95), potential overfitting cannot be ruled out due to the relatively small sample size and single-center data source. While 10-fold cross-validation and parameter tuning were employed to mitigate this risk, the model may still capture site-specific patterns, limiting generalizability. Additionally, sarcopenia-related features such as grip strength and muscle mass require standardized measurement protocols, which may not be uniformly available across institutions. Future multicenter and prospective validation is needed to confirm model robustness and clinical applicability.

Traditional machine learning algorithms are often critiqued for lacking transparency and interpretability (28). To address this issue, we applied SHAP values to explain our model's predictions, focusing on the AdaBoost model. SHAP values were used to determine how each feature influenced the model's output, highlighting important variables such as maximum grip strength, preoperative MMSE score, 6-meter walking speed, age, appendicular skeletal muscle mass, and sarcopenia assessment. These variables are crucial for tailoring perioperative strategies for older surgical patients.

Previous research indicates that grip strength correlates with cognitive declines in older adults (29). Our findings further suggest that stronger grip strength in elderly individuals is inversely associated with the risk of PND and cognitive impairment, whereas weaker grip strength is significantly linked to the onset of dementia and delirium. This association may stem from shared neural mechanisms between motor and cognitive functions (30). For instance, diminished grip strength may indicate systemic inflammation, a condition confirmed to correlate with cognitive decline, encompassing delirium and dementia. A decrease in preoperative grip strength not only reflects diminished muscle strength but may also signal dysfunction in related muscle groups, leading to impaired physical function and mobility, thereby increasing the risk of PND (10). Such patients are also found to have higher levels of postoperative inflammatory markers (S-100 β , IL-6) (31).

Multiple studies have demonstrated that appendicular skeletal muscle mass is a key indicator of sarcopenia and muscle health, showing strong correlations with PND incidence: (1) Liu et al. reported that patients with reduced skeletal muscle mass exhibited significantly poorer performance on postoperative cognitive function tests, suggesting that low skeletal muscle mass is a potential risk factor for PND (1). (2) The reduction in skeletal muscle mass is associated with insulin resistance (32), energy metabolism disorders, increased inflammatory responses, oxidative stress, and cerebrovascular disease. These factors contribute to amyloid-beta deposition, tau protein hyperphosphorylation, and reduced neural plasticity (33). (3) Sarcopenic patients may

experience changes in body composition, such as decreased lean body mass and increased fat, which can lead to dysfunction in organ systems, thereby altering the pharmacodynamics and pharmacokinetics of anesthetic drugs and resulting in lingering side effects that impair neural function.

The 6-meter walking speed is a reliable indicator of functional capacity and endurance in older individuals, correlating reductions with cognitive decline and increased depression, highlighting its role in diagnosing and managing sarcopenia (34). Studies have shown that a walking speed of <0.8 meters per second in the 6-meter walking speed suggests the presence of sarcopenia (34). These measures serve not only as diagnostic criteria for sarcopenia but also reflect the overall health and functional status of patients. The connection between sarcopenia and PND indicates that patients with sarcopenia may be at greater risk for poor cognitive and physical recovery after surgery (35). Therefore, effective preoperative management of sarcopenia could improve outcomes and reduce the incidence of PND.

Age significantly influences the risk of PND, with a study of 1,064 patients showing higher prevalence in the elderly (36). Animal research also confirms that older mice are more prone to cognitive decline and Alzheimer's disease-related changes post-surgery, linked to β -amyloid accumulation (37). The risk of PND in older surgical patients is positively correlated with age, and this risk is further exacerbated by sarcopenia, which affects muscle mass and strength. By identifying these variables, we can improve the precision of PND risk prediction. To mitigate this risk, targeted preoperative interventions such as strength training and nutritional support could be implemented, potentially reducing the risk of PND and enhancing both cognitive and physical recovery after surgery.

5 Conclusion

In this study, we leveraged key variables, particularly those indicative of sarcopenia, to develop a suite of machine learning models that effectively predict the incidence of PND among the elderly. Specifically, the AdaBoost model demonstrated superior performance. Additionally, we applied the SHAP method to address the “black box” nature of machine learning, enabling us to determine the significance of each feature within our models. These models not only enable clinicians to identify older surgical patients more accurately at high risk for PND, particularly those with sarcopenia-related characteristics, but also aid in devising personalized perioperative management strategies that could significantly enhance patients' postoperative quality of life.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The data that support the findings of this study are available from the first author upon reasonable request. Requests to access these datasets should be directed to Zhengyu Qian, 18384970837@163.com.

Ethics statement

The studies involving humans were approved by Medical Ethics Committee of the First Affiliated Hospital of Chengdu Medical College (Approval No: 2021CYFYIRB-BA-67-01) on November 28, 2021. 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

ZQ: Writing – original draft, Writing – review & editing. XW: Writing – review & editing. KH: Writing – review & editing. KL: Writing – review & editing. XL: Writing – review & editing. TZ: Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research received funding support from: High-level talent introduction project of the First Affiliated Hospital of Chengdu Medical College (CYFY-GQ51); Scientific research project of Sichuan Medical Hygiene and Health Promotion Association (No. KY2022QN0289); Foundation of Chengdu Medical College (24LHLNYX1-13 and 23LHPDZYB27).

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.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2025.1604333/full#supplementary-material>

References

- Liu Y, Fu H, Wang T. Neuroinflammation in perioperative neurocognitive disorders: from bench to the bedside. *CNS Neurosci Ther.* (2022) 28:484–96. doi: 10.1111/cns.13794
- Sun Y, Wang K, Zhao W. Gut microbiota in perioperative neurocognitive disorders: current evidence and future directions. *Front Immunol.* (2023) 14:1178691. doi: 10.3389/fimmu.2023.1178691
- Likhvantsev VV, Landoni G, Berikashvili LB, Ermokhina NV, Yadgarov MY, Kotani Y, et al. Effects of early postoperative neurocognitive disorders on clinically relevant outcomes: a meta-analysis. *Korean J Anesthesiol.* (2023) 76:490–500. doi: 10.4097/kja.202126
- Han C, Ji H, Guo Y, Fei Y, Wang C, Yuan Y, et al. Effect of subanesthetic dose of esketamine on perioperative neurocognitive disorders in elderly undergoing gastrointestinal surgery: a randomized controlled trial. *Drug Des Devel Ther.* (2023) 17:863–73. doi: 10.2147/DDDT.S401161
- Gómez-Soria I, Ferreira C, Oliván Blázquez B, Magallón Botaya RM, Calatayud E. Short-term memory, attention, and temporal orientation as predictors of the cognitive impairment in older adults: a cross-sectional observational study. *PLoS One.* (2021) 16:e0261313. doi: 10.1371/journal.pone.0261313
- Hadjistavropoulos T, Herr K, Prkachin KM, Craig KD, Gibson SJ, Lukas A, et al. Pain assessment in elderly adults with dementia. *Lancet Neurol.* (2014) 13:1216–27. doi: 10.1016/S1474-4422(14)70103-6
- Lu W, Tang S, Li A, Huang Q, Dou M, Zhang Y, et al. The role of PKC/PKR in aging, Alzheimer's disease, and perioperative neurocognitive disorders. *Front Aging Neurosci.* (2022) 14:973068. doi: 10.3389/fnagi.2022.973068
- Zhao Q, Wan H, Pan H, Xu Y. Postoperative cognitive dysfunction-current research progress. *Front Behav Neurosci.* (2024) 18:1328790. doi: 10.3389/fnbeh.2024.1328790
- Petermann-Rocha F, Balntzi V, Gray SR, Lara J, Ho FK, Pell JP, et al. Global prevalence of sarcopenia and severe sarcopenia: a systematic review and meta-analysis. *J Cachexia Sarcopenia Muscle.* (2022) 13:86–99. doi: 10.1002/jcsm.12783
- Nemoto A, Goyagi T, Nemoto W, Nakagawasa O, Tan-No K, Niiyama Y. Low skeletal muscle mass is associated with perioperative neurocognitive disorder due to decreased neurogenesis in rats. *Anesth Analg.* (2022) 134:194–203. doi: 10.1213/ANE.0000000000005681
- Mosk CA, van Vugt JLA, de Jonge H, Witjes CD, Buettner S, Ijzermans JN, et al. Low skeletal muscle mass as a risk factor for postoperative delirium in elderly patients undergoing colorectal cancer surgery. *Clin Interv Aging.* (2018) 13:2097–106. doi: 10.2147/CIA.S175945
- Soma D, Kawamura YI, Yamashita S, Wake H, Nohara K, Yamada K, et al. Sarcopenia, the depletion of muscle mass, an independent predictor of respiratory complications after oncological esophagectomy. *Dis Esophagus.* (2019) 32:doy092. doi: 10.1093/dote/doy092
- Choi KM. Sarcopenia and sarcopenic obesity. *Korean J Intern Med.* (2016) 31:1054–60. doi: 10.3904/kjim.2016.193
- Xie X, Li J, Zhong Y, Fang Z, Feng Y, Chen C, et al. A risk prediction model based on machine learning for postoperative cognitive dysfunction in elderly patients with non-cardiac surgery. *Aging Clin Exp Res.* (2023) 35:2951–60. doi: 10.1007/s40520-023-02573-x
- Xie N, Yan S, Sun X, Liu H. Establish a nomogram of cardiac postoperative cognitive dysfunction. *Heart Surg Forum.* (2021) 24:E320–E6. doi: 10.1532/hfs.3551
- Borchers F, Spies CD, Feinkohl I, Brockhaus WR, Kraft A, Kozma P, et al. Methodology of measuring postoperative cognitive dysfunction: a systematic review. *Br J Anaesth.* (2021) 126:1119–27. doi: 10.1016/j.bja.2021.01.035
- Chen LK, Woo J, Assantachai P, Auyeung TW, Chou MY, Iijima K, et al. Asian Working Group for Sarcopenia: 2019 consensus update on Sarcopenia diagnosis and treatment. *J Am Med Dir Assoc.* (2020) 21:300–7.e2. doi: 10.1016/j.jamda.2019.12.012
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* (1988) 44:837–45. doi: 10.2307/2531595
- Hsu WH, Ko AT, Weng CS, Chang CL, Jan YT, Lin JB, et al. Explainable machine learning model for predicting skeletal muscle loss during surgery and adjuvant chemotherapy in ovarian cancer. *J Cachexia Sarcopenia Muscle.* (2023) 14:2044–53. doi: 10.1002/jcsm.13282
- Xiong X, Chen D, Shi J. Is Perioperative dexmedetomidine associated with a reduced risk of perioperative neurocognitive disorders following cardiac surgery? A systematic review and meta-analysis with trial sequential analysis of randomized controlled trials. *Front Med.* (2021) 8:645975. doi: 10.3389/fmed.2021.645975
- Lai Z, Min J, Li J, Shan W, Yu W, Zuo Z. Surgery trauma severity but not anesthesia length contributes to postoperative cognitive dysfunction in mice. *J Alzheimers Dis.* (2021) 80:245–57. doi: 10.3233/JAD-201232
- Fan W, Mai L, Wu Z, Wu Q, Yang X, Gu W, et al. The effect of anesthesia and surgery on postoperative changes in plasma biomarkers of neuronal injury, Alzheimer's disease, and inflammation in healthy subjects. *Res Square.* (2021). doi: 10.21203/rs.3.rs-317062/v1
- Williamson BD, Gilbert PB, Simon NR, Carone M. A general framework for inference on algorithm-agnostic variable importance. *J Am Stat Assoc.* (2023) 118:1645–58. doi: 10.1080/01621459.2021.2003200
- Han Y, Wang S. Disability risk prediction model based on machine learning among Chinese healthy older adults: results from the China Health and Retirement Longitudinal Study. *Front Public Health.* (2023) 11:1271595. doi: 10.3389/fpubh.2023.1271595
- Roosbeh M. Optimal QR-based estimation in partially linear regression models with correlated errors using GCV criterion. *Comput Stat Data Anal.* (2018) 117:45–61. doi: 10.1016/j.csda.2017.08.002
- Valier A, editor. “The Cross Validation in Automated Valuation Models: A Proposal for Use”. In: *Computational Science and Its Applications – ICCSA*. Cham: Springer International Publishing (2020). doi: 10.1007/978-3-030-58814-4_45
- Jiang Z, Walker K, Shi D. Applying AdaBoost to improve diagnostic accuracy. *Methodology.* (2019) 15:77–87. doi: 10.1027/1614-2241/a000166
- Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface.* (2018) 15:20170387. doi: 10.1098/rsif.2017.0387
- Sprague BN, Phillips CB, Ross LA. Age-varying relationships between physical function and cognition in older adulthood. *J Gerontol B Psychol Sci Soc Sci.* (2019) 74:772–84. doi: 10.1093/geronb/gbx126
- Qian J, Shen X, Gao X, Sun Q. Grip strength is a predictor for subsyndromal delirium among older adults following joint replacement. *Clin Interv Aging.* (2023) 18:1675–85. doi: 10.2147/CIA.S423727
- Peng L, Xu L, Ouyang W. Role of peripheral inflammatory markers in postoperative cognitive dysfunction (POCD): a meta-analysis. *PLoS ONE.* (2013) 8:e79624. doi: 10.1371/journal.pone.0079624
- Stump CS, Henriksen EJ, Wei Y, Sowers JR. The metabolic syndrome: role of skeletal muscle metabolism. *Ann Med.* (2006) 38:389–402. doi: 10.1080/07853890600888413
- Gomez-Rios MA, Abad-Gurumeta A. Anesthesia in the elderly patient. Resilience in frailty time. *Med Clin.* (2022) 159:486–8. doi: 10.1016/j.medcle.2022.05.023
- Morley JE, Abbatecola AM, Argiles JM, Baracos V, Bauer J, Bhasin S, et al. Sarcopenia with limited mobility: an international consensus. *J Am Med Dir Assoc.* (2011) 12:403–9. doi: 10.1016/j.jamda.2011.04.014
- Wang SL, Zhuang CL, Huang DD, Pang WY, Lou N, Chen FF, et al. Sarcopenia adversely impacts postoperative clinical outcomes following gastrectomy in patients with gastric cancer: a prospective study. *Ann Surg Oncol.* (2016) 23:556–64. doi: 10.1245/s10434-015-4887-3
- Monk TG, Weldon BC, Garvan CW, Dede DE, van der Aa MT, Heilman KM, et al. Predictors of cognitive dysfunction after major noncardiac surgery. *Anesthesiology.* (2008) 108:18–30. doi: 10.1097/01.anes.0000296071.19434.1e
- Xu Z, Dong Y, Wang H, Culley DJ, Marcantonio ER, Crosby G, et al. Age-dependent postoperative cognitive impairment and Alzheimer-related neuropathology in mice. *Sci Rep.* (2014) 4:3766. doi: 10.1038/srep03766
- Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. *Ann Transl Med.* (2016) 4:30. doi: 10.3978/j.issn.2305-5839.2015.12.63