

Original Paper

A Supervised Explainable Machine Learning Model for Perioperative Neurocognitive Disorder in Liver-Transplantation Patients and External Validation on the Medical Information Mart for Intensive Care IV Database: Retrospective Study

Zhendong Ding¹, MD; Linan Zhang¹, MD; Yihan Zhang¹, MD; Jing Yang¹, MD, PhD; Yuheng Luo², MD; Mian Ge¹, MD, PhD; Weifeng Yao¹, MD, PhD; Ziqing Hei¹, MD, PhD; Chaojin Chen¹, MD, PhD

¹Department of Anesthesiology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

²Guangzhou AI & Data Cloud Technology Co., LTD, Guangzhou, China

Corresponding Author:

Chaojin Chen, MD, PhD

Department of Anesthesiology

The Third Affiliated Hospital of Sun Yat-sen University

No. 600 Tianhe Road

Guangzhou, 510630

China

Phone: 86 13430322182

Email: chenchj28@mail.sysu.edu.cn

Related Article:

This is a corrected version. See correction statement in: <https://www.jmir.org/2026/1/e88952>

Abstract

Background: Patients undergoing liver transplantation (LT) are at risk of perioperative neurocognitive dysfunction (PND), which significantly affects the patients' prognosis.

Objective: This study used machine learning (ML) algorithms with an aim to extract critical predictors and develop an ML model to predict PND among LT recipients.

Methods: In this retrospective study, data from 958 patients who underwent LT between January 2015 and January 2020 were extracted from the Third Affiliated Hospital of Sun Yat-sen University. Six ML algorithms were used to predict post-LT PND, and model performance was evaluated using area under the receiver operating curve (AUC), accuracy, sensitivity, specificity, and F_1 -scores. The best-performing model was additionally validated using a temporal external dataset including 309 LT cases from February 2020 to August 2022, and an independent external dataset extracted from the Medical Information Mart for Intensive Care IV (MIMIC-IV) database including 325 patients.

Results: In the development cohort, 201 out of 751 (33.5%) patients were diagnosed with PND. The logistic regression model achieved the highest AUC (0.799) in the internal validation set, with comparable AUC in the temporal external (0.826) and MIMIC-IV validation sets (0.72). The top 3 features contributing to post-LT PND diagnosis were the preoperative overt hepatic encephalopathy, platelet level, and postoperative sequential organ failure assessment score, as revealed by the Shapley additive explanations method.

Conclusions: A real-time logistic regression model-based online predictor of post-LT PND was developed, providing a highly interoperable tool for use across medical institutions to support early risk stratification and decision making for the LT recipients.

(*J Med Internet Res* 2025;27:e55046) doi: [10.2196/55046](https://doi.org/10.2196/55046)

KEYWORDS

machine learning; risk factors; liver transplantation; perioperative neurocognitive disorders; MIMIC-IV database; external validation

Introduction

Perioperative neurocognitive disorder (PND), encompassing various postsurgical cognitive impairments identified especially in the postoperative period, was first proposed in 2018 [1]. These cognitive changes are consistent with the clinical diagnostic criteria for neurocognitive disorders outlined in the *DSM-5 (Diagnostic and Statistical Manual of Mental Disorders [Fifth Edition])* [1-3]. In addition to postoperative delirium (POD) [4,5], other components of PND include emergence delirium, delayed neurocognitive recovery, and postoperative neurocognitive dysfunction [2,6]. POD or PND incidence is 2%-3% after general surgery [5,7] and 50%-70% in high-risk patients [8]. In addition, PND not only contributes to increased mortality rates but also extends hospitalization in patients undergoing liver transplantation (LT) [7,9], escalating health care costs and resource use. Preventative strategies and timely interventions for post-LT PND are crucial for enhancing patient outcomes and easing health care burdens [10].

Existing studies identify risk factors for post-LT PND, such as excessive alcohol consumption, Child-Turcotte-Pugh scores, and model for end-stage liver disease (MELD) scores [11,12]. Potential biomarkers for cognitive impairment prediction have also been proposed, including calcium binding protein β and neuron-specific enolase [13], yet their practical application is hindered by complex clinical scenarios and expense.

Machine learning (ML), a branch of artificial intelligence, offers a solution by distilling extensive clinical data into actionable

insights, identifying relative risk factors for PND [14,15]. However, there is a dearth of ML-based models predicting post-LT-related complications [16-22] and postoperative delirium during specific surgeries [4,23]. There are currently no appropriate models for predicting PND in LT recipients, with most current clinical prediction models often failing to maintain accuracy when applied to external datasets, resulting in significant limitations to their generalizability.

This study aimed to extract critical predictors and develop an efficient ML algorithm to predict PND in LT recipients using routinely collected clinical data and to validate its performance using the Medical Information Mart for Intensive Care IV (MIMIC-IV) database.

Material and Methods

Study Design and Patients

This retrospective, single-center study was conducted at our institution following the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis guidelines. We enrolled 1267 patients who underwent LT between January 2015 and August 2022. Records were extracted using the perioperative specialist database platform (PSDP) and electronic patient record (EPR) systems. The inclusion and exclusion criteria are shown in [Textbox 1](#).

All included recipients were formalized and registered in the China Organ Transplant Response System.

Textbox 1. Inclusion and exclusion criteria for the study.

Inclusion criteria

- Age >18 years.
- Allogeneic liver transplantation.

Exclusion criteria

- Simultaneous liver and kidney transplantation.
- Preoperative overt hepatic encephalopathy.
- Emergency reoperation.
- Persistent postoperative coma and inability to screen for cognitive function.
- Post-liver transplantation cerebral infarction or hemorrhage.
- Incomplete medical records.

Data Collection

The development and temporal validation cohort datasets were created by extracting original records from the Docare System (Medical system), Hospital Information System, and Laboratory Information System, and integrating them into the PSDP platform and EPR systems. To increase ML model accuracy and applicability, we included the following variables: (1) demographic characteristics; (2) liver donor characteristics; (3) preoperative comorbidities, complications, preoperative treatment, and LT etiology; (4) preoperative laboratory test results; (5) intraoperative surgery characteristics and

medications; (6) postoperative MELD scores, sequential organ failure assessment (SOFA) scores, and laboratory test results; and (7) complications and prognosis in LT recipients. All of the original data were made anonymous throughout the study.

Definitions of Outcomes

The primary outcome was postoperative PND occurrence from surgery until discharge from the hospital. A summary of perioperative neurocognitive impairments is shown in [Table S1 in Multimedia Appendix 1](#). The initial diagnosis criteria was the retrieval of any of the following terms from the medical records: “Delirium”, “Confusion”, “Confusional arousals”,

“Clouding of consciousness”, “Soma”, “Drowsiness”, “Changes in mental status”, “Hallucinations”, “Disorientation”, “Dyscalculia”, “Haziness of spirit-mind”, “Irritability”, “Agitation”, “Inattentiveness”, “Reactive confusion”, “Somatization disorder”, “Irritability”, and “Somatoform disorders”, or equivalent terms in Chinese [4,24,25]. Next, each patient was evaluated based on the *DSM-5* criteria by a designated neurologist without prior access to the patient’s records [3,26].

Variable Selection

A comprehensive set of 137 variables was extracted for the initial analysis (Table S2 in [Multimedia Appendix 1](#)). Table S3 in [Multimedia Appendix 1](#) provides a concise explanation of the main complications and relevant term definitions. Postoperative SOFA scores were calculated by intensive care unit (ICU) physicians immediately after surgery according to European Society of Intensive Care Medicine criteria [27] and submitted for statistical analysis.

To account for multicollinearity and confounding variables affecting the overall model fitting performance, variables that were statistically significant ($P < .05$) in the univariate test were subjected to stability selection (Table S4 in [Multimedia Appendix 1](#)) [28]. After 100 iterations of least absolute shrinkage and selection operator (LASSO) regression, the top 10 features with the highest selection frequencies were chosen to train the ML models. For each LASSO regression, 90% of the training set samples were randomly selected as subsamples.

Machine Learning Models

The following 6 ML models were developed, and their performances were further evaluated: logistic regression (LR), multilayer perceptron classifier (MLP), extreme gradient boosting with classification trees (XGB), light gradient boosting machine (LGB), support vector machine (SVM), and random forest classifier (RF). All models were constructed using the XGB, LGB, and Scikit-learn packages.

The primary cohort dataset was randomly divided into 80% development and 20% internal validation sets. The bootstrap method was implemented 1000 times on the internal validation set to determine a 95% CI for the discrimination assessment metrics for each model: the area under the receiver operating curve (AUC), accuracy, sensitivity, specificity, and F_1 -scores. Considering that ML models have multiple hyperparameters that are essential for model performance, a 5-fold cross-validation grid search method was used to optimize the parameters and AUCs (Table S5 in [Multimedia Appendix 1](#)). The Shapley additive explanations (SHAP) method was used to assess predictive feature importance and explain the ML algorithms’ predictions [29].

Model Performance Comparison and MIMIC-IV Dataset

Because the SOFA and MELD scores have been reported as potential predictors of various post-LT complications [16,30], our study also compared the ML model’s performance against SOFA and MELD scores.

An external validation set extracted from the MIMIC-IV (version 2.2) [31] database was used to evaluate the ML model’s performance, which was authorized by the review committee of Massachusetts Institute of Technology (agreement 1.5.0). Patients who underwent LT surgery and were diagnosed with PND according to the International Classification of Diseases (9th and 10th revisions) were enrolled. Data extraction and cleaning were performed using PostgreSQL (version 15.3) and Navicate Premium (version 16) with a Structured Query Language (Figure S1 in [Multimedia Appendix 1](#)).

Statistical Analysis

Data cleaning used Python (version 3.9.13) packages Pandas (version 1.4.4) and Numpy (version 1.23.5). Data analysis used the Python Scipy package (version 3.7), and SHAP (0.41.0) was used to visualize and analyze feature importance.

Data distribution was evaluated using the Kolmogorov Smirnov test. Normally distributed continuous variables are presented as mean (SD) and were compared by independent sample t tests. Non-normally distributed continuous data are presented as median (IQR) and were compared using the nonparametric equivalent (Mann Whitney test). Categorical variables are expressed as frequencies and percentages and were tested using the chi-square test or Fisher exact test. Long-term survival rates were estimated using the Kaplan Meier method. Group comparisons were conducted using the Gehan-Breslow Wilcoxon test and log-rank tests.

All tests were 2-tailed, with statistical significance set at 0.05. Before ML model training, continuous variables were normalized, dichotomous variables were coded as binary variables, and multicategory variables were coded as uniform numbers.

Variables with missing values exceeding 20% were excluded, and missing values below 20% were imputed with the median (for numeric variables) or mode (for categorical variables). The overall data distribution after imputation exhibited an acceptable level of variability.

Visualized Online Calculator

An online calculator with a visual interface was developed to facilitate the easy input of clinical variables and to generate clear and meaningful output indicating the absolute risk in percentages.

Ethical Considerations

The study protocol was approved by the Ethics Committee of the Third Affiliated Hospital of Sun Yat-sen University on July 27, 2022 (No. (2019)02-609-04) and was conducted in accordance with the Declaration of Helsinki. The requirement for informed patient consent was waived due to the study’s retrospective nature, and all data were anonymized before analysis.

Results

Patient Demographic Characteristics

The flowchart for patient recruitment is shown in [Figure 1](#). Of the 958 patients who underwent LT, 751 patients were enrolled

randomly into the development set (n=600) and internal validation set (n=151). Notably, PND occurred in 201 patients, accounting for 33.5% of the development cohort. Table 1 and Table S6 in Multimedia Appendix 1 summarizes the

development set's demographic characteristics, donor features, and perioperative variables of patients with or without post-LT PND.

Figure 1. Diagram of experimental procedure and flowchart, (A) brief diagram of the experimental procedure and (B) flowchart for patient enrollment, development and selection of machine learning model. LGB: light gradient boosting machine; LR: logistic regression; MIMIC-IV: the Medical Information Mart for Intensive Care IV; ML: machine learning; MLP: multilayer perceptron classifier; RF: random forest classifier; SVM: support vector machine; XGB: extreme gradient boosting with classification trees.

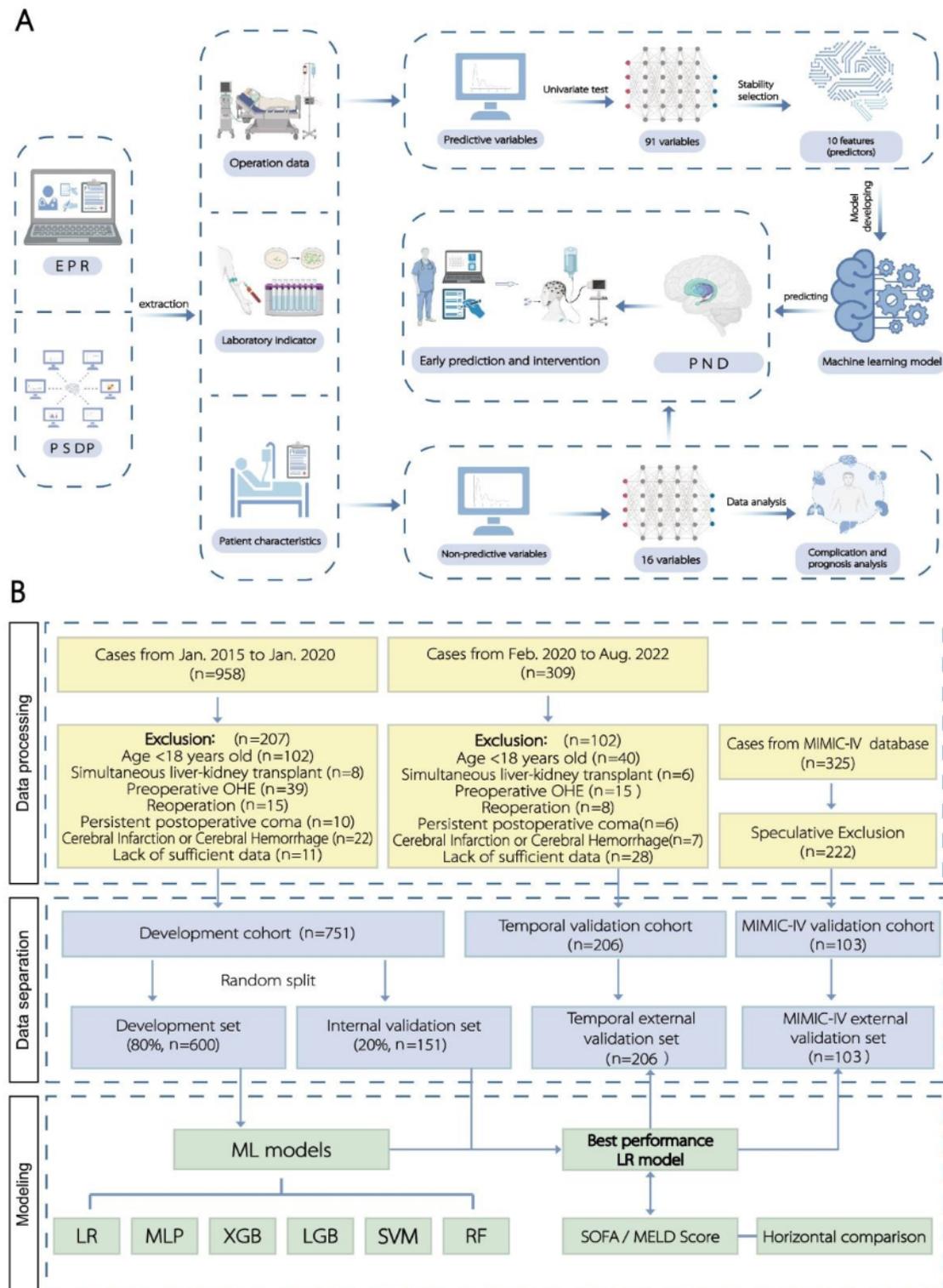


Table 1. Demographic characteristics and donor characteristics variables of patients with stratification by perioperative neurocognitive disorder.^a

Characteristics	Total (n=600)	NonPND ^b (n=399)	PND ^b (n=201)	P value
Demographic characteristics				
Age (years), mean (SD)	49 (10.34)	49.24 (10.16)	48.53 (10.7)	.43
Sex				
				.06
Female, n (%)	74 (12.33%)	42 (10.63%)	32 (16%)	
Male, n (%)	521 (86.83%)	353 (89.37%)	168 (84%)	
Height (cm), median (IQR)	170 (172-165)	170 (172-165)	169 (170-163)	.17
Weight (kg), median (IQR)	64 (71-58)	64 (72-58.88)	64 (70-58)	.26
BMI, median (IQR)	22.84 (24.85-20.43)	22.78 (24.90-20.45)	22.86 (24.74-20.44)	.79
Blood group, n (%)				
				.76
A	233 (38.83)	156 (39.49)	77 (38.5)	
B	156 (26)	106 (26.84)	50 (25)	
O	167 (27.83)	110 (27.85)	57 (28.5)	
AB	39 (6.5)	23 (5.82)	16 (8)	
Donor characteristics				
Donor age (years), median (IQR)	40 (49-28)	40 (50-28)	40 (48-29.5)	.64
Donor BMI, median (IQR)	22.49 (24.22-20.7)	22.49 (24.22-20.76)	22.59 (24.52-20.38)	.35
Donor Type				
				.03
DBD ^c , n (%)	321 (53.5)	228 (62.3)	93 (50.54)	
DCD ^d , n (%)	225 (37.5)	136 (37.16)	89 (48.37)	
DBCD ^e , n (%)	4 (0.67)	2 (0.55)	2 (1.09)	
Steatosis of donor liver				
				.27
Steatosis grade 0, n (%)	390 (65)	265 (71.05)	125 (65.79)	
Steatosis grade 1, n (%)	147 (24.5)	94 (25.2)	53 (27.89)	
Steatosis grade 2, n (%)	26 (4.33)	14 (3.75)	12 (6.32)	

^aData presented in Table 1 are based on the original dataset prior to data imputation.

^bPND: perioperative neurocognitive dysfunction.

^cDBD: donation after brain death.

^dDCD: donation after circulatory death.

^eDBCD: donation after brain death followed by circulatory death.

Perioperative Characteristics

Among the preoperative characteristics, American Society of Anesthesiologists classification and preoperative comorbidities such as acute respiratory distress syndrome, and laboratory results including hemoglobin, white blood cell (WBC) count, liver function, coagulation function, and serum calcium were significantly different between patients with and without postoperative PND ($P < .01$, Table S6 in [Multimedia Appendix 1](#)). Specifically, individuals diagnosed with post-LT PND exhibited a notably elevated prevalence of preoperative cover hepatic encephalopathy (CHE; 45.27% vs 8.77%, $P < .001$) and hypercalcemia (7.57% vs 1.36%, $P < .001$). Furthermore, patients with post-LT PND had higher Child Pugh and MELD scores ($P < .001$), longer preoperative ICU stays, increased continuous blood purification, increased plasma exchange, longer

mechanical ventilation, and higher tracheal intubation (all $P < .001$, Table S6 in [Multimedia Appendix 1](#)).

Regarding intraoperative characteristics, patients with post-LT PND had longer anesthesia durations; increased sodium bicarbonate levels, red blood cell counts, plasma levels, and levels of cryoprecipitate transfusion; increased estimated blood loss (EBL); and reduced urine output (all $P < .001$, Table S6 in [Multimedia Appendix 1](#)). Differences in intraoperative medications between the 2 groups were not significant, except for recombinant activated factor VII ($P < .001$). Interestingly, our results showed no association between day or night surgery and the incidence of PND ($P = .44$, Table S6 in [Multimedia Appendix 1](#)).

For the postoperative characteristics, patients with post-LT PND showed significantly higher levels of aspartate aminotransferase

(AST), total bilirubin, blood urea nitrogen, prothrombin time (PT), international normalized ratio, hypersensitive C-reactive protein (hsCRP), procalcitonin, and serum calcium, as well as lower levels of hemoglobin, hematocrit, WBC, platelet (PLT), gamma-glutamyltransferase, albumin, and serum osmolality (all $P < .05$, Table S6 in [Multimedia Appendix 1](#)).

Feature Selection

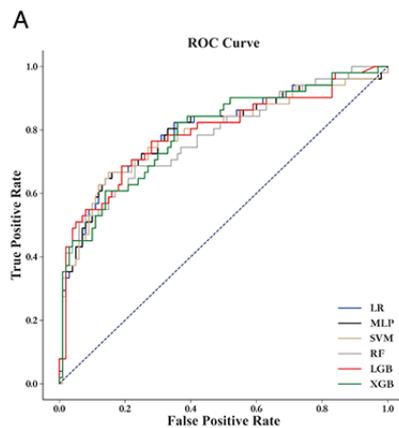
The frequency of LASSO algorithm selection for each variable is shown in detail in Figure S2 in [Multimedia Appendix 1](#). The top 10 features chosen as predictors for ML model development were preoperative CHE, PLT, PT, estimated glomerular filtration rate (eGFR), Ca^{2+} , MELD score, intraoperative EBL, postoperative SOFA score, hsCRP, and AST.

Model Performance and Horizontal Comparison

The performance of the 6 ML models is shown in [Figure 2](#). The LR model achieved the highest AUC (0.799, 95% CI 0.709-0.877) with acceptable accuracy (0.722, 95% CI 0.642-0.795), sensitivity (0.714, 95% CI 0.575-0.833), and specificity (0.73, 95% CI 0.639-0.811) compared with the other 5 models.

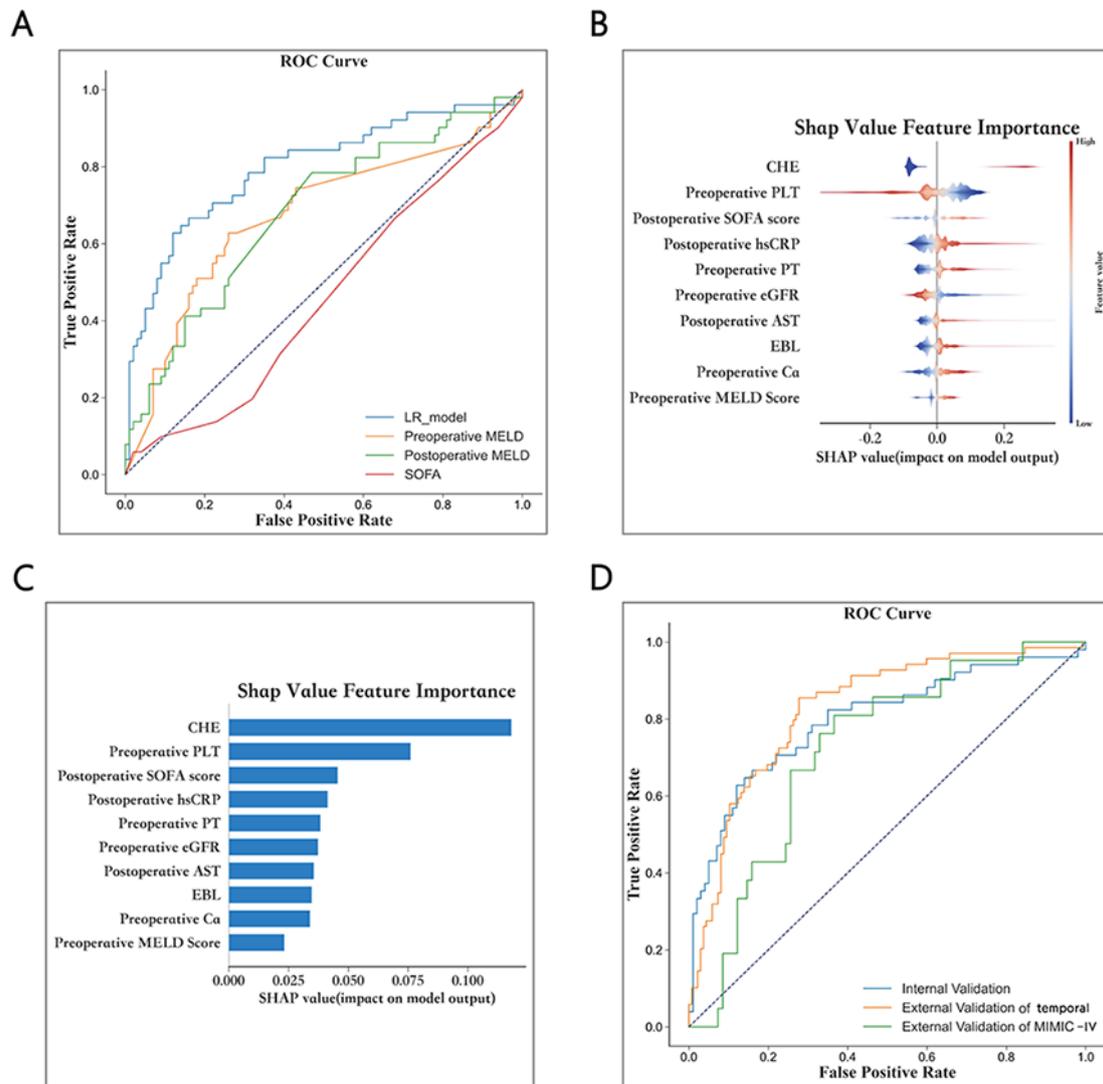
The SOFA (AUC=0.459, 95% CI 0.365-0.555), preoperative MELD (AUC=0.672, 95% CI 0.581-0.768), and postoperative MELD scores (AUC=0.679, 95% CI 0.587-0.772) had significantly lower AUCs than the LR model in the internal validation set ([Figure 3A](#)).

Figure 2. Performance metrics for six ML models. (A) ROC curves of six ML models. (B) Details of the model performance metrics. Accuracy=(TP+TN)/(TP+TN+FP+FN); AUC, the area under the receiver-operating curve; F1=2*Precision*Recall/(Precision + Recall); FN: false negative; FP: false positive; LGB: light gradient boosting machine; LR: logistic regression; MLP: multilayer perceptron classifier; RF: random forest classifier; Sensitivity=TP/(TP + FN); Specificity (Recall)=TN/(TN + FP); SVM: support vector machine; TN: true negative; TP: true positive; XGB: extreme gradient boosting with classification trees.



Model	AUC	Accuracy	Sensitivity	Specificity	F1
LR	0.799 (0.709, 0.877)	0.722 (0.642, 0.795)	0.714 (0.575, 0.833)	0.73 (0.639, 0.811)	0.632 (0.513, 0.734)
MLP	0.797 (0.708, 0.879)	0.788 (0.722, 0.854)	0.61 (0.463, 0.733)	0.882 (0.812, 0.942)	0.66 (0.537, 0.76)
SVM	0.787 (0.692, 0.862)	0.728 (0.656, 0.788)	0.725 (0.592, 0.837)	0.734 (0.645, 0.809)	0.643 (0.525, 0.735)
RF	0.791 (0.705, 0.871)	0.722 (0.656, 0.795)	0.628 (0.49, 0.767)	0.772 (0.686, 0.85)	0.603 (0.473, 0.708)
LGB	0.792 (0.709, 0.872)	0.748 (0.675, 0.821)	0.708 (0.583, 0.83)	0.769 (0.685, 0.851)	0.655 (0.547, 0.755)
XGB	0.795 (0.711, 0.863)	0.709 (0.636, 0.781)	0.688 (0.542, 0.809)	0.72 (0.636, 0.806)	0.611 (0.495, 0.712)

Figure 3. SHAP analysis of the LR model and model performance in horizontal comparison and external validation. (A) Horizontal comparison of predicting performance between the LR model and MELD/SOFA scores in the internal validation set. (B-C) The SHAP summary plot demonstrated the general importance of each feature in LR model. The color bar on the right indicates the relative value of a feature in each case, with red color representing higher value and blue color representing lower value. (D-E) ROC curves and model performance in the external validation. AST: aspartate aminotransferase; AUC: the area under the receiver-operating curve; CHE: cover hepatic encephalopathy; EBL: estimated blood loss; eGFR: estimated glomerular filtration rate; hsCRP: hypersensitive C-reactive protein; LR: logistic regression; MELD scores: model for end-stage liver disease score; MIMIC-IV: Medical Information Mart for Intensive Care IV; PLT: platelet; PT: prothrombin time; SHAP: Shapley additive explanations; SOFA scores: sequential organ failure assessment score.



Validation set	AUC	Accuracy	Sensitivity	Specificity	F1
Internal	0.799 (0.709, 0.877)	0.722 (0.642, 0.795)	0.714 (0.575, 0.833)	0.73 (0.639, 0.811)	0.632 (0.513, 0.734)
Temporal external	0.826 (0.765, 0.887)	0.748 (0.689, 0.811)	0.766 (0.662, 0.861)	0.743 (0.667, 0.821)	0.671 (0.589, 0.755)
MIMIC-IV external	0.72 (0.606, 0.829)	0.631 (0.534, 0.728)	0.812 (0.64, 0.95)	0.589 (0.481, 0.694)	0.472 (0.312, 0.602)

Feature Importance

The SHAP summary plot (Figures 3B and 3C) illustrates the correlation between the feature value magnitudes in the LR model. Both SHAP plots revealed that the presence of CHE, lower preoperative PLT, higher postoperative SOFA score, higher postoperative hsCRP, and higher preoperative PT were associated with a higher SHAP value output in the LR model, indicating a heightened likelihood of post-LT PND and forming the top 5 effective variables.

Three correctly classified examples (eg, patients 48, 80, and 122) are presented in Figure S3 in Multimedia Appendix 1, showing the SHAP decision and force plots.

Temporal External Validation and MIMIC-IV Dataset Validation

A comparison of the main demographic characteristics and key predictive variables between the development and validation sets is shown in Table S7 in Multimedia Appendix 1, and the incidence rates of post-LT PND in the temporal and MIMIC-IV external validation were 27.1%, and 20.3%, respectively. The LR model exhibited a comparable performance in the temporal external validation set (AUC=0.826, 95% CI 0.765-0.887) (Figure 3D). Surprisingly, the LR model also provided acceptable predictions for the MIMIC-IV dataset (Figure 3D,

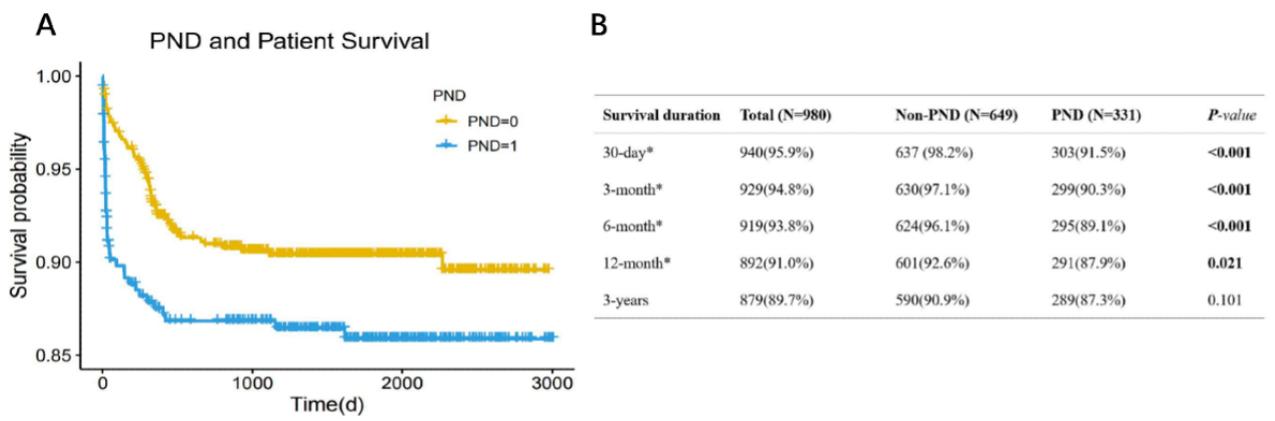
AUC=0.72, 95% CI 0.606-0.829). Figure 3E summarizes the main performance metrics of the LR model.

Effect of Perioperative Neurocognitive Dysfunction on Patients' Outcomes and Prognosis

Compared with patients without post-LT PND, patients with PND were more likely to experience perioperative complications (Table S8 in Multimedia Appendix 1), including higher incidences of sepsis (51.63% vs 21.55%, $P<.001$), pneumonia (75.56% vs 65.46%, $P<.05$), acute kidney injury (69.5% vs 39.75%, $P<.001$), and hemodialysis (51.35% vs 12.81%, $P<.001$). Furthermore, patients with post-LT PND had higher hospitalization costs (CNY 377,801.69 [US \$51,566.83], SD 177,855.53 [US \$24,275.82] vs CNY 277,018.95 [US \$37,810.82], SD 92,779.91 [US \$12,663.70]; $P<.001$), prolonged postoperative stays (25 {18} vs 21 {11} days, $P<.001$), longer postoperative ICU stay (113 {114} vs 65 {48.5} hours, $P<.001$), and a markedly higher in-hospital mortality rate (12.44% vs 2.51%, $P<.001$).

Further survival analysis (Figure 4) was conducted to assess patient prognosis. The PND group exhibited significantly lower survival rates at 30 days (91.5% vs 98.2%, $P<.001$), 3 months (90.3% vs 97.1%, $P<.001$), 6 months (89.1% vs 96.1%, $P<.001$), and 12 months (87.9% vs 92.6%, $P=.02$).

Figure 4. Post-liver transplantation survival associated with perioperative neurocognitive dysfunction. Patients with post-liver transplantation perioperative neurocognitive dysfunction showed a significantly lower survival rate. LT: liver transplantation; PND: perioperative neurocognitive dysfunction.

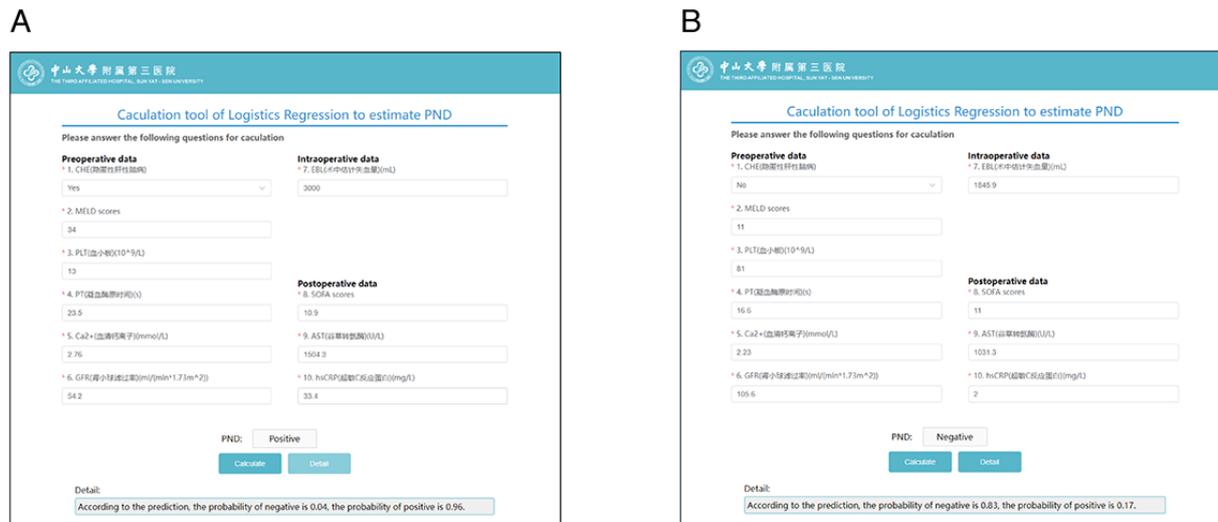


Clinical Availability of the Logistic Regression Model

Given the accessibility of the 10 predictive features, we constructed a visually oriented online calculator to facilitate clinical decision making. The perioperative information of 2 typical patients was entered into the online calculator: patient

48 had a positive final predicted probability of PND occurrence (probability: 96%), and patient 122 had a negative final predicted probability of PND occurrence (probability: 17%; Figure 5). The online calculator is freely accessible at the hospital website.

Figure 5. Online calculator for the clinical interface of the post-liver transplantation perioperative neurocognitive dysfunction risk prediction logistic regression model. (A) Patient No. 48 post-liver transplantation perioperative neurocognitive dysfunction will occur (probability of perioperative neurocognitive dysfunction: 94%); (B) Patient No. 122 post-liver transplantation perioperative neurocognitive dysfunction will not occur (probability of perioperative neurocognitive dysfunction: 17%).



Discussion

Principal Findings

Our retrospective study assessed 6 different ML algorithms to predict post-LT PND, using 10 readily available clinical parameters. We found that post-LT PND incidence was 33.5%. The 10 predictive features significantly associated with PND included preoperative CHE, PLT, PT, eGFR, Ca²⁺, MELD score, intraoperative EBL and postoperative SOFA score, hsCRP, and AST. The LR model demonstrated superior performance, with high AUC, accuracy, sensitivity, and specificity, surpassing traditional SOFA and MELD scores in predicting post-LT PND and performed acceptably in the rigorous temporal and MIMIC-IV external validations.

This study aids clinicians in detecting postoperative cognitive changes in LT recipients. Patients with PND typically faced more perioperative complications, higher hospitalization costs, and prolonged hospital and ICU stays, consistent with previous studies [4,23]. Hepatic encephalopathy has been reported as an independent risk factor for postoperative neurocognitive disorders [32]. To ensure cognitive assessment accuracy, we excluded patients with overt hepatic encephalopathy according to the spectrum of neurocognitive impairment in cirrhosis criteria [33]. CHE emerged as a significant predictor in our model analysis. Both oxidative stress and neuroinflammation have been implicated in POD pathophysiology [10,34]. A recent systematic review also links increased perioperative CRP levels to a high delirium risk [35], supporting our inclusion of hsCRP as a predictor. Calcium ions (Ca²⁺) are important cell signaling molecules, and previous studies reported a positive correlation between Ca²⁺ concentration and neuronal apoptosis extent in vitro [36], consistent with our results. Furthermore, the model identified PLT as an unconventional indicator of PND, showcasing ML's ability to highlight nontraditional risk factors.

This discovery is partly supported by Eyer et al [37] suggesting a relationship between lower PLT and delirium tremens.

Our study used preoperative, intraoperative, and postoperative data (SOFA scores, hsCRP, and AST levels) to develop the LR model. Earlier studies have revealed that multiple postoperative factors were also risk factors for PND [11,12]. The postoperative variables included in this study were predominantly assessed upon initial admission to the ICU. Stability selection analysis revealed a positive correlation between elevated postoperative SOFA scores, hsCRP levels, and AST levels, and an increased likelihood of post-LT PND. This highlights the predictive value of these commonly observed postoperative variables for PND.

Our results suggest that LR outperforms other ML models in predicting post-LT PND, which is not surprising. A recent systematic review showed no performance superiority of other ML models over LR in predicting clinical complications [38]. Wiredu et al [35] also found that compared to ML algorithms, LR had the highest AUC when predicting sex-specific hip fractures. Song et al [4] developed an LR model to predict POD in older adult patients, achieving the highest AUC compared with other models. Given the evident linear relationships among the top 10 features, the LR may be more appropriate for capturing distribution patterns. In contrast to other algorithms, LR performs well on nonoversized and high-dimensional datasets, exhibits computational efficiency, and imposes lower dataset requirements.

As demonstrated by the example of prediction cases (Figure 5), we successfully developed a predictive model for post-LT PND, with its primary advantage in its reliable predictive performance, validated using 2 external datasets. The importance of early detection and prevention of PND in patients undergoing cardiac surgery or transplantation is clearly emphasized in current international guidelines [39]. However, the implementation of preventive measures is often challenged by limited resources [40], especially in cases where the shortage of liver donors persists. On accurate identification by the LR model, patients

at high risk for post-LT POD could be referred to enhanced LT perioperative management strategies, such as individualized pharmacological or nonpharmacological comprehensive multicomponent interventions, according to the 10 commonly accessible predictive parameters filtered by the ML algorithm.

Limitations

However, this study had several limitations. First, it was a single-center retrospective study, meaning the Confusion Assessment Method (CAM) or the associated CAM-ICU and 3D-CAM were inappropriate for our database. Instead, patients with PND were identified from medical records according to the *DSM-5* criteria [3,6,26]. Second, as a real-world study, researchers can only infer precise risk factors based on the data available, and inhomogeneous confounding among the datasets

could affect the study conclusions [41]. While our online decision tool has the potential to aid surgeons and anesthesiologists in clinical decision making, the causes and underlying mechanisms of PND remain subjects of intense debate, necessitating further research.

Conclusions

This study successfully develops a real-time and easily accessible parameter requiring LR-based PND prediction algorithm for post-LT settings. The LR model outperformed the other five models owing to its enhanced model performance and interpretability. The optimal use of our freely accessible online predictor would enable timely and convenient risk stratification, enhanced perioperative management strategies, and comprehensive multicomponent interventions.

Acknowledgments

We express our gratitude to Ms Li Jing from the Chinese MCC5 University for her valuable assistance in processing the figures. This study was supported partly by the Special Support Project of Guangdong Province (0720240209), the Natural Science Foundation of Guangdong Province (grant 2022A1515012603), the Joint Funds of the National Natural Science Foundation of China (U22A20276), Science and Technology Planning Project of Guangdong Province-Regional Innovation Capacity and Support System Construction (2023B110006), Provincial-enterprise Joint Funds of Guangdong Basic and Applied Basic Research Foundation (2021B1515230012), Science and Technology Program of Guangzhou, China (202201020429) and the “Five and five” project of the Third Affiliated Hospital of Sun Yat-Sen University (2023WW501).

Data Availability

All data generated or analyzed during this study are included in this published article and [Multimedia Appendix 1](#), and the codes used in this study are all common codes in Python packages mentioned in the “Methods” section of the manuscript.

Authors' Contributions

Authors ZD and CC had full access to all of the data in this study and take responsibility for the integrity of the data and the accuracy of the data analysis. ZD, ZH, and CC contributed to conceptualization and design. ZD, LZ, YZ, YL, and MG handled acquisition, analysis, or interpretation of data. ZD, CC, and WY managed drafting of the manuscript. All authors contributed to critical revision of the manuscript for important intellectual content. ZD, YL, and JY performed statistical analysis. ZH and CC obtained funding. JY and YL managed administrative, technical, or material support.

Address correspondence to CC (for access to raw data and statistical analysis): Department of Anesthesiology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, chenchj28@mail.sysu.edu.cn; and ZH (regarding study design and funding acquisition): Department of Anesthesiology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, heiziqing@sina.com; and WY (regarding the managed drafting and revision of the manuscript): Department of Anesthesiology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, yaowf3@mail.sysu.edu.cn.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Additional online content.

[\[DOCX File , 1000 KB-Multimedia Appendix 1\]](#)

References

1. Evered L, Silbert B, Knopman DS, Scott DA, DeKosky ST, Rasmussen LS, et al. Nomenclature Consensus Working Group. Recommendations for the nomenclature of cognitive change associated with anaesthesia and surgery-2018. *Anesthesiology*. 2018;129(5):872-879. [[FREE Full text](#)] [doi: [10.1097/ALN.0000000000002334](https://doi.org/10.1097/ALN.0000000000002334)] [Medline: [30325806](https://pubmed.ncbi.nlm.nih.gov/30325806/)]
2. Tasbihgou SR, Absalom AR. Postoperative neurocognitive disorders. *Korean J Anesthesiol*. 2021;74(1):15-22. [[FREE Full text](#)] [doi: [10.4097/kja.20294](https://doi.org/10.4097/kja.20294)] [Medline: [32623846](https://pubmed.ncbi.nlm.nih.gov/32623846/)]
3. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. Washington, DC London, England. American Psychiatric Association; 2013.

4. Song YX, Yang XD, Luo YG, Ouyang CL, Yu Y, Ma YL, et al. Comparison of logistic regression and machine learning methods for predicting postoperative delirium in elderly patients: a retrospective study. *CNS Neurosci Ther*. 2023;29(1):158-167. [[FREE Full text](#)] [doi: [10.1111/cns.13991](https://doi.org/10.1111/cns.13991)] [Medline: [36217732](#)]
5. Marcantonio ER. Delirium in hospitalized older adults. *N Engl J Med*. 2017;377(15):1456-1466. [[FREE Full text](#)] [doi: [10.1056/NEJMc1605501](https://doi.org/10.1056/NEJMc1605501)] [Medline: [29020579](#)]
6. Kong H, Xu LM, Wang DX. Perioperative neurocognitive disorders: a narrative review focusing on diagnosis, prevention, and treatment. *CNS Neurosci Ther*. 2022;28(8):1147-1167. [[FREE Full text](#)] [doi: [10.1111/cns.13873](https://doi.org/10.1111/cns.13873)] [Medline: [35652170](#)]
7. Gleason LJ, Schmitt EM, Kosar CM, Tabloski P, Saczynski JS, Robinson T, et al. Effect of delirium and other major complications on outcomes after elective surgery in older adults. *JAMA Surg*. 2015;150(12):1134-1140. [[FREE Full text](#)] [doi: [10.1001/jamasurg.2015.2606](https://doi.org/10.1001/jamasurg.2015.2606)] [Medline: [26352694](#)]
8. Jin Z, Hu J, Ma D. Postoperative delirium: perioperative assessment, risk reduction, and management. *Br J Anaesth*. 2020;125(4):492-504. [[FREE Full text](#)] [doi: [10.1016/j.bja.2020.06.063](https://doi.org/10.1016/j.bja.2020.06.063)] [Medline: [32798069](#)]
9. Mottaghi S, Nikoupour H, Firoozifar M, Jalali SS, Jamshidzadeh A, Vazin A, et al. The effect of taurine supplementation on delirium post liver transplantation: a randomized controlled trial. *Clin Nutr*. 2022;41(10):2211-2218. [doi: [10.1016/j.clnu.2022.07.042](https://doi.org/10.1016/j.clnu.2022.07.042)] [Medline: [36081295](#)]
10. Oh ES, Fong TG, Hshieh TT, Inouye SK. Delirium in older persons: advances in diagnosis and treatment. *JAMA*. 2017;318(12):1161-1174. [[FREE Full text](#)] [doi: [10.1001/jama.2017.12067](https://doi.org/10.1001/jama.2017.12067)] [Medline: [28973626](#)]
11. Zhou S, Deng F, Zhang J, Chen G. Incidence and risk factors for postoperative delirium after liver transplantation: a systematic review and meta-analysis. *Eur Rev Med Pharmacol Sci*. 2021;25(8):3246-3253. [[FREE Full text](#)] [doi: [10.26355/eurrev_202104_25733](https://doi.org/10.26355/eurrev_202104_25733)] [Medline: [33928610](#)]
12. Zhou J, Xu X, Liang Y, Zhang X, Tu H, Chu H. Risk factors of postoperative delirium after liver transplantation: a systematic review and meta-analysis. *Minerva Anestesiol*. 2021;87(6):684-694. [[FREE Full text](#)] [doi: [10.23736/S0375-9393.21.15163-6](https://doi.org/10.23736/S0375-9393.21.15163-6)] [Medline: [33594873](#)]
13. Wang CM, Chen WC, Zhang Y, Lin S, He HF. Update on the mechanism and treatment of sevoflurane-induced postoperative cognitive dysfunction. *Front Aging Neurosci*. 2021;13:702231. [[FREE Full text](#)] [doi: [10.3389/fnagi.2021.702231](https://doi.org/10.3389/fnagi.2021.702231)] [Medline: [34305576](#)]
14. Arora A. Artificial intelligence: a new frontier for anaesthesiology training. *Br J Anaesth*. 2020;125(5):e407-e408. [[FREE Full text](#)] [doi: [10.1016/j.bja.2020.06.049](https://doi.org/10.1016/j.bja.2020.06.049)] [Medline: [32682557](#)]
15. Mathis MR, Kheterpal S, Najarian K. Artificial intelligence for anesthesia: what the practicing clinician needs to know: more than black magic for the art of the dark. *Anesthesiology*. 2018;129(4):619-622. [[FREE Full text](#)] [doi: [10.1097/ALN.0000000000002384](https://doi.org/10.1097/ALN.0000000000002384)] [Medline: [30080689](#)]
16. Chen C, Chen B, Yang J, Li X, Peng X, Feng Y, et al. Development and validation of a practical machine learning model to predict sepsis after liver transplantation. *Ann Med*. 2023;55(1):624-633. [[FREE Full text](#)] [doi: [10.1080/07853890.2023.2179104](https://doi.org/10.1080/07853890.2023.2179104)] [Medline: [36790357](#)]
17. Tran J, Sharma D, Gotlieb N, Xu W, Bhat M. Application of machine learning in liver transplantation: a review. *Hepatol Int*. 2022;16(3):495-508. [doi: [10.1007/s12072-021-10291-7](https://doi.org/10.1007/s12072-021-10291-7)] [Medline: [35020154](#)]
18. Zhang Y, Yang D, Liu Z, Chen C, Ge M, Li X, et al. An explainable supervised machine learning predictor of acute kidney injury after adult deceased donor liver transplantation. *J Transl Med*. 2021;19(1):321. [[FREE Full text](#)] [doi: [10.1186/s12967-021-02990-4](https://doi.org/10.1186/s12967-021-02990-4)] [Medline: [34321016](#)]
19. Chen C, Yang D, Gao S, Zhang Y, Chen L, Wang B, et al. Development and performance assessment of novel machine learning models to predict pneumonia after liver transplantation. *Respir Res*. 2021;22(1):94. [[FREE Full text](#)] [doi: [10.1186/s12931-021-01690-3](https://doi.org/10.1186/s12931-021-01690-3)] [Medline: [33789673](#)]
20. Spann A, Yasodhara A, Kang J, Watt K, Wang B, Goldenberg A, et al. Applying machine learning in liver disease and transplantation: a comprehensive review. *Hepatology*. 2020;71(3):1093-1105. [doi: [10.1002/hep.31103](https://doi.org/10.1002/hep.31103)] [Medline: [31907954](#)]
21. Lee BP, Vittinghoff E, Hsu C, Han H, Therapondos G, Fix OK, et al. Predicting low risk for sustained alcohol use after early liver transplant for acute alcoholic hepatitis: the sustained alcohol use post-liver transplant score. *Hepatology*. 2019;69(4):1477-1487. [[FREE Full text](#)] [doi: [10.1002/hep.30478](https://doi.org/10.1002/hep.30478)] [Medline: [30561766](#)]
22. Ayllón MD, Ciria R, Cruz-Ramírez M, Pérez-Ortiz M, Gómez I, Valente R, et al. Validation of artificial neural networks as a methodology for donor-recipient matching for liver transplantation. *Liver Transpl*. 2018;24(2):192-203. [doi: [10.1002/lt.24870](https://doi.org/10.1002/lt.24870)] [Medline: [28921876](#)]
23. Zhang Y, Wan DH, Chen M, Li YL, Ying H, Yao GL, et al. Automated machine learning-based model for the prediction of delirium in patients after surgery for degenerative spinal disease. *CNS Neurosci Ther*. 2023;29(1):282-295. [doi: [10.1111/cns.14002](https://doi.org/10.1111/cns.14002)] [Medline: [36258311](#)]
24. 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):950-958. [[FREE Full text](#)] [doi: [10.1001/jamasurg.2020.2853](https://doi.org/10.1001/jamasurg.2020.2853)] [Medline: [32822459](#)]
25. Hornor MA, Ma M, Zhou L, Cohen ME, Rosenthal RA, Russell MM, et al. Enhancing the American college of surgeons NSQIP surgical risk calculator to predict geriatric outcomes. *J Am Coll Surg*. 2020;230(1):88-100.e1. [doi: [10.1016/j.jamcollsurg.2019.09.017](https://doi.org/10.1016/j.jamcollsurg.2019.09.017)] [Medline: [31672676](#)]

26. Kuhn E, Du X, McGrath K, Coveney S, O'Regan N, Richardson S, et al. Validation of a consensus method for identifying delirium from hospital records. *PLoS One*. 2014;9(11):e111823. [FREE Full text] [doi: [10.1371/journal.pone.0111823](https://doi.org/10.1371/journal.pone.0111823)] [Medline: [25369057](https://pubmed.ncbi.nlm.nih.gov/25369057/)]
27. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the working group on sepsis-related problems of the European society of intensive care medicine. *Intensive Care Med*. 1996;22(7):707-710. [doi: [10.1007/BF01709751](https://doi.org/10.1007/BF01709751)] [Medline: [8844239](https://pubmed.ncbi.nlm.nih.gov/8844239/)]
28. Meinshausen N, Bühlmann P. Stability selection. *J R Stat Soc*. 2010;72(4):417-473. [doi: [10.1111/j.1467-9868.2010.00740.x](https://doi.org/10.1111/j.1467-9868.2010.00740.x)]
29. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1):56-67. [FREE Full text] [doi: [10.1038/s42256-019-0138-9](https://doi.org/10.1038/s42256-019-0138-9)] [Medline: [32607472](https://pubmed.ncbi.nlm.nih.gov/32607472/)]
30. Yao L, Li Y, Yin R, Yang L, Ding N, Li B, et al. Incidence and influencing factors of post-intensive care cognitive impairment. *Intensive Crit Care Nurs*. 2021;67:103106. [doi: [10.1016/j.iccn.2021.103106](https://doi.org/10.1016/j.iccn.2021.103106)] [Medline: [34246526](https://pubmed.ncbi.nlm.nih.gov/34246526/)]
31. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*. 2023;10(1):1. [FREE Full text] [doi: [10.1038/s41597-022-01899-x](https://doi.org/10.1038/s41597-022-01899-x)] [Medline: [36596836](https://pubmed.ncbi.nlm.nih.gov/36596836/)]
32. Garcia-Martinez R, Rovira A, Alonso J, Jacas C, Simón-Talero M, Chavarria L, et al. Hepatic encephalopathy is associated with posttransplant cognitive function and brain volume. *Liver Transpl*. 2011;17(1):38-46. [FREE Full text] [doi: [10.1002/lt.22197](https://doi.org/10.1002/lt.22197)] [Medline: [21254343](https://pubmed.ncbi.nlm.nih.gov/21254343/)]
33. Bajaj JS, Cordoba J, Mullen KD, Amodio P, Shawcross DL, Butterworth RF, et al. International Society for Hepatic Encephalopathy Nitrogen Metabolism (ISHEN). Review article: the design of clinical trials in hepatic encephalopathy--an international society for hepatic encephalopathy and nitrogen metabolism (ISHEN) consensus statement. *Aliment Pharmacol Ther*. 2011;33(7):739-747. [FREE Full text] [doi: [10.1111/j.1365-2036.2011.04590.x](https://doi.org/10.1111/j.1365-2036.2011.04590.x)] [Medline: [21306407](https://pubmed.ncbi.nlm.nih.gov/21306407/)]
34. Inouye SK, Westendorp RGJ, Saczynski JS. Delirium in elderly people. *Lancet*. 2014;383(9920):911-922. [FREE Full text] [doi: [10.1016/S0140-6736\(13\)60688-1](https://doi.org/10.1016/S0140-6736(13)60688-1)] [Medline: [23992774](https://pubmed.ncbi.nlm.nih.gov/23992774/)]
35. Wiredu K, Aduse-Poku E, Shaefi S, Gerber SA. Proteomics for the discovery of clinical delirium biomarkers: a systematic review of major studies. *Anesth Analg*. 2023;136(3):422-432. [FREE Full text] [doi: [10.1213/ANE.0000000000006246](https://doi.org/10.1213/ANE.0000000000006246)] [Medline: [36580411](https://pubmed.ncbi.nlm.nih.gov/36580411/)]
36. Kahraman S, Zup S, McCarthy M, Fiskum G. GABAergic mechanism of propofol toxicity in immature neurons. *J Neurosurg Anesthesiol*. 2008;20(4):233-240. [FREE Full text] [doi: [10.1097/ANA.0b013e31817ec34d](https://doi.org/10.1097/ANA.0b013e31817ec34d)] [Medline: [18812886](https://pubmed.ncbi.nlm.nih.gov/18812886/)]
37. Eyer F, Schuster T, Felgenhauer N, Pfab R, Strubel T, Saugel B, et al. Risk assessment of moderate to severe alcohol withdrawal--predictors for seizures and delirium tremens in the course of withdrawal. *Alcohol Alcohol*. 2011;46(4):427-433. [doi: [10.1093/alcalc/agr053](https://doi.org/10.1093/alcalc/agr053)] [Medline: [21593124](https://pubmed.ncbi.nlm.nih.gov/21593124/)]
38. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12-22. [doi: [10.1016/j.jclinepi.2019.02.004](https://doi.org/10.1016/j.jclinepi.2019.02.004)] [Medline: [30763612](https://pubmed.ncbi.nlm.nih.gov/30763612/)]
39. Foley KA, Djaiani G. Update of the European society of anaesthesiology and intensive care medicine evidence-based and consensus-based guideline on postoperative delirium in adult patients. *Eur J Anaesthesiol*. 2025;42(1):86-87. [doi: [10.1097/EJA.0000000000002043](https://doi.org/10.1097/EJA.0000000000002043)] [Medline: [39628423](https://pubmed.ncbi.nlm.nih.gov/39628423/)]
40. Hughes CG, Boncyk CS, Culley DJ, Fleisher LA, Leung JM, McDonagh DL, et al. Perioperative Quality Initiative (POQI) 6 Workgroup. American society for enhanced recovery and perioperative quality initiative joint consensus statement on postoperative delirium prevention. *Anesth Analg*. 2020;130(6):1572-1590. [FREE Full text] [doi: [10.1213/ANE.0000000000004641](https://doi.org/10.1213/ANE.0000000000004641)] [Medline: [32022748](https://pubmed.ncbi.nlm.nih.gov/32022748/)]
41. Concato J, Corrigan-Curay J. Real-world evidence - where are we now? *N Engl J Med*. 2022;386(18):1680-1682. [doi: [10.1056/NEJMp2200089](https://doi.org/10.1056/NEJMp2200089)] [Medline: [35485775](https://pubmed.ncbi.nlm.nih.gov/35485775/)]

Abbreviations

- AST:** aspartate aminotransferase
- AUC:** area under the receiver operating curve
- CAM:** Confusion Assessment Method
- CHE:** cover hepatic encephalopathy
- DSM-5:** Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
- EBL:** estimated blood loss
- eGFR:** estimated glomerular filtration rate
- EPR:** electronic patient record
- hsCRP:** hypersensitive C-reactive protein
- ICU:** intensive care unit
- LASSO:** least absolute shrinkage and selection operator
- LGB:** light gradient boosting machine

LR: logistic regression
LT: liver transplantation
MELD: model for end-stage liver disease
MIMIC-IV: Medical Information Mart for Intensive Care IV
ML: machine learning
MLP: multilayer perceptron classifier
PLT: platelet
PND: perioperative neurocognitive disorder
POD: postoperative delirium
PSDP: perioperative specialist database platform
PT: prothrombin time
RF: random forest classifier
SHAP: Shapley additive explanations
SOFA: sequential organ failure assessment
SVM: support vector machine
WBC: white blood cell
XGB: extreme gradient boosting with classification trees

Edited by T de Azevedo Cardoso; submitted 30.Nov.2023; peer-reviewed by K Xie, V Chauhan; comments to author 07.Feb.2024; revised version received 12.Apr.2024; accepted 30.Oct.2024; published 15.Jan.2025

Please cite as:

Ding Z, Zhang L, Zhang Y, Yang J, Luo Y, Ge M, Yao W, Hei Z, Chen C

A Supervised Explainable Machine Learning Model for Perioperative Neurocognitive Disorder in Liver-Transplantation Patients and External Validation on the Medical Information Mart for Intensive Care IV Database: Retrospective Study

J Med Internet Res 2025;27:e55046

URL: <https://www.jmir.org/2025/1/e55046>

doi: [10.2196/55046](https://doi.org/10.2196/55046)

PMID: [39813086](https://pubmed.ncbi.nlm.nih.gov/39813086/)

©Zhendong Ding, Linan Zhang, Yihan Zhang, Jing Yang, Yuheng Luo, Mian Ge, Weifeng Yao, Ziqing Hei, Chaojin Chen. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 15.Jan.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.