



OPEN

## Logistic regression technique is comparable to complex machine learning algorithms in predicting cognitive impairment related to post intensive care syndrome

TingTing Wu<sup>1,2,5</sup>, YueQing Wei<sup>2,3,5</sup>, JingBing Wu<sup>2,4</sup>, BiLan Yi<sup>2,3</sup> & Hong Li<sup>1,2</sup>✉

To evaluate the performance of machine learning (ML) models and to compare it with logistic regression (LR) technique in predicting cognitive impairment related to post intensive care syndrome (PICS-CI). We conducted a prospective observational study of ICU patients at two tertiary hospitals. A cohort of 2079 patients was screened, and finally 481 patients were included. Seven different ML models were considered, decision tree (DT), random forest (RF), XGBoost, neural network (NN), naïve bayes (NB), and support vector machine (SVM), and compared with logistic regression (LR). Discriminative ability was evaluated by area under the receiver operating characteristic curve (AUC), calibration belt plots, and Hosmer–Lemeshow test was used to assess calibration. Decision curve analysis was performed to quantify clinical utility. Duration of delirium, poor Richards–Campbell sleep questionnaire (RCSQ) score, advanced age, and sepsis were the most frequent and important candidates risk factors for PICS-CI. All ML models showed good performance (AUC range: 0.822–0.906). NN model had the highest AUC (0.906 [95% CI 0.857–0.955]), which was slightly higher than, but not significantly different from that of LR (0.898 [95% CI 0.847–0.949]) ( $P > 0.05$ , Delong test). Given the overfitting and complexity of some ML models, the LR model was then used to develop a web-based risk calculator to aid decision-making (<https://model871010.shinyapps.io/dynnomapp/>). In a low dimensional data, LR may yield as good performance as other complex ML models to predict cognitive impairment after ICU hospitalization.

### Abbreviations

AUC	Area under the receiver operating characteristic curve
CAM-ICU	Confusion assessment method for intensive care unit questionnaire
CCI	Charlson comorbidity index
DT	Decision tree
RF	Random forest
LR	Logistic regression
ML	Machine learning
MoCA	Montreal cognitive assessment
NN	Neural network
NB	Naïve bayes
NPV	Negative predictive value
PICS-CI	Cognitive impairment related to post intensive care syndrome
PPV	Positive predictive value
RCSQ	Richards–Campbell sleep questionnaire

<sup>1</sup>The School of Nursing, Fujian Medical University, Fujian, China. <sup>2</sup>Research Center for Nursing Theory and Practice, Fujian Provincial Hospital, No 134, East Street, Gulou District, Fuzhou 35001, Fujian, China. <sup>3</sup>Respiratory and Intensive Care Unit, Fujian Provincial Hospital, Fujian, China. <sup>4</sup>Medical Intensive Care Unit, Fujian Provincial Hospital, Fujian, China. <sup>5</sup>These authors contributed equally: TingTing Wu and YueQing Wei. ✉email: leehong99@126.com

SVM Support vector machine  
WBC White blood cell

Post-intensive care syndrome (PICS) is defined as development of new, or worsening of preexisting impairment in physical function, psychological disorder, or cognitive impairment that persists beyond ICU and hospital discharge<sup>1</sup>. Cognitive impairment related to PICS (PICS-CI) is one of the phenomena spanning from delirium to long-term cognitive impairment after discharge<sup>2</sup>, manifesting as memory loss, inattention, and/or executive dysfunction syndrome<sup>3</sup>. Neuroradiological and neuropathological studies in patients with PICS-CI have demonstrated diffuse brain damage, including global and local patterns of atrophy, as well as cortical and subcortical lesions, specifically in the corpus callosum, the hippocampus, and the basal ganglia<sup>4</sup>. PICS-CI is a common and severe complication of ICU admission. There is considerable variability in the reported incidence of PICS-CI after critical illness (19–78.2%)<sup>5,6</sup>. In a study, 23.9% patients showed persistence of cognitive deficit 6 years after discharge, and only 45.7% returned to the workforce<sup>7</sup>. PICS-CI after ICU admission can be greatly taxing to patients and their families<sup>8</sup>. The condition affects people of all ages<sup>8</sup> and is associated with increased mortality<sup>3</sup>. Moreover, it adversely affects the activities of daily living and the quality of life, and the recovery process is typically protracted and incomplete over several years after discharge, and carries enormous societal cost<sup>8</sup>.

The need to detect PICS-CI is increasingly recognized. Several studies have investigated the risk factors for PICS, particularly preventable ones, but have yielded discrepant results. Some studies have identified delirium as a strong risk factor<sup>6,9,10</sup>. In a few studies, delirium exhibited a dose–response relationship with PICS. Longer duration of delirium showed an independent association with poor cognitive function<sup>8</sup>, and repeated delirium exposure was associated with a greater risk of developing cognitive deficit compared to single delirium episode<sup>10</sup>. Older age has not yet been conclusively determined as an independent risk factor, as no age-dependent increase in the prevalence of cognitive deficit was observed after ICU admission<sup>8</sup>; however, other studies support the hypothesis of a higher risk at older ages<sup>11–13</sup>. Some ICU-related parameters such as longer ICU length of stay and mechanical ventilation showed a significant positive association with the occurrence of cognitive impairment<sup>13</sup>. However, a meta-analysis confirmed that delirium is the sole risk factor for cognitive impairment<sup>14</sup>. A recent meta-analysis revealed a paucity of rigorous cognitive function prediction models, and identified the need for further studies in this field<sup>15</sup>.

Machine learning (ML) has been regarded as an indispensable tool for revealing complex questions in the field of medicine<sup>16</sup>. In critical care settings, ML applications have been leveraged to create predictive and prognostic models using supervised learning algorithms<sup>17</sup>. Complex ML algorithms (e.g., RF, NN, XGboost) have been claimed to present superior predictive discrimination compared to logistic regression technique<sup>18–24</sup>; nevertheless, the advantage of state of the art ML models over the conventional generalized linear models remains controversial<sup>25–28</sup>. For instance, two systematic reviews found no significant difference in performance between logistic regression and ML in clinical prediction modelling<sup>25,28</sup>.

To the best of our knowledge, no studies have developed and compared performance of multiple ML models for prognostic modeling in the context of PICS-CI. Therefore, the purpose of this study was to (1) identify significant risk factors for PICS-CI; (2) develop multiple ML models and compare these with traditional LR; (3) generate a visualization web-based calculator risk prediction model with the greatest possible accuracy.

## Materials and methods

**Study setting and design.** We conducted a prospective observational study of ICU patients admitted between January and October 2019 at two tertiary hospitals in Fuzhou city, China, including general ICU, medical ICU, surgical ICU, cardiac ICU, and emergency ICU.

**Study population.** Patients who qualified the following criteria were included in the study: (1) age > 18 years and < 80 years; (2) ICU length of stay > 48 h. Patients who qualified any of the following criteria were excluded: (1) nervous system diseases such as cerebral infarction, cerebral hemorrhage, meningitis, intracranial infection, craniocerebral injury, Parkinson's, and brain tumor; (2) post-cardiopulmonary resuscitation or cardiac arrest; (3) pulmonary encephalopathy, hepatic encephalopathy, diabetic hyperosmolar coma, unexplained coma, and other diseases that may affect cognitive function; (4) patients with mental disorder, dementia/pre-dementia, or intellectual disability; (5) history of pesticide poisoning, drug poisoning, chronic alcohol abuse (consumption of alcohol for > 5 years, daily ethanol intake  $\geq$  80 g, calculation formula: ethanol volume (g) = consumed volume (mL)  $\times$  ethanol concentration  $\times$  0.8); (6) Abuse of drugs such as hypnotics and anesthetics; (7) hearing impairment or visual impairment and inability to complete the cognitive assessment; (8) accepted pharmacological or non-pharmacological cognitive interventions; unwilling or unable to cooperate with assessment (e.g., due to tracheotomy).

**Procedure and data collection.** After approval of the study protocol by the ethics institutional review board, which comply with Declaration of Helsinki, 3 researchers in each of the 5 ICUs underwent standardized training for data collection. The study nurse explained the purpose of the study to eligible patients and obtained written informed consent from those who agreed to participate, according to Declaration of Helsinki.

Two methods (on-spot assessment and review of medical records) were used for data collection. The collected data were classified into 5 categories (a total of 35 variables), including demographics, disease-related features, treatment-related features, laboratory test features, and sleep quality. Sleep quality was assessed daily during the period of stay in the ICU. Screening for cognitive impairment was conducted by researcher after patient was discharged and returned to the general ward. Other factors were measured and recorded in the electronic

medical record (EMR) by medical staff, and we extracted and confirmed the information with the patient. The data collection process is shown as follows.

**Demographics.** Demographic information was obtained from the EMR including age, sex (male or female), educational level (illiterate, primary school, junior high school, high school or secondary vocational school, college school or above), occupation (physical labor/nonphysical labor), body mass index, and history of alcohol consumption.

**Disease-related features.** The disease-related parameters were extracted from the ICU discharge diagnosis from EMR, including diagnosis (sepsis, hypertension, diabetes, chronic cardiorespiratory disease, chronic renal disease, respiratory failure, etc.). We classified it as acute disease, acute exacerbation of chronic disease, cardiac surgery, and non-cardiac surgery. We also calculated the Charlson Comorbidity Index (CCI) to evaluate the severity of complications, and the Acute Physiology and Chronic Health Status Score (Acute Physiology and Chronic Health Evaluation II, APACHE II).

**Treatment-related features.** The treatment-related factors were summarized after discharge from the ICU, which were daily assessed and recorded in the EMR by medical staff, including duration of mechanical ventilation, cumulative doses of sedatives (propofol and midazolam), and cumulative doses of analgesics (sufentanil and remifentanyl), duration of ICU stay, delirium, and duration of delirium.

Delirium was assessed using the Confusion Assessment Method for Intensive Care Unit questionnaire (CAM-ICU)<sup>29</sup> during ICU hospitalization. Delirium assessments were conducted twice daily (before 10 AM and after 8 PM) by the charge nurse. Before each delirium assessment, the Richmond Agitation and Sedation Scale (RASS)<sup>3</sup> was used to assess the level of consciousness; patients with scores of  $-4$  or  $-5$  were not tested for delirium; delirium was assessed only when RASS score was below  $-3$ . CAM-ICU delirium was deemed positive when there was an acute change in mental status from baseline with difficulty in focusing attention in combination with altered level of consciousness or disorganized thinking. Duration of delirium referred to the number of days on which delirium was observed.

**Laboratory test features.** The laboratory data were derived from EMR after ICU discharge, including oxygenation index  $\text{PaO}_2/\text{FiO}_2$  (minimum), procalcitonin (maximum), glucose (minimum and maximum), WBC count (minimum and maximum), and lactate (maximum).

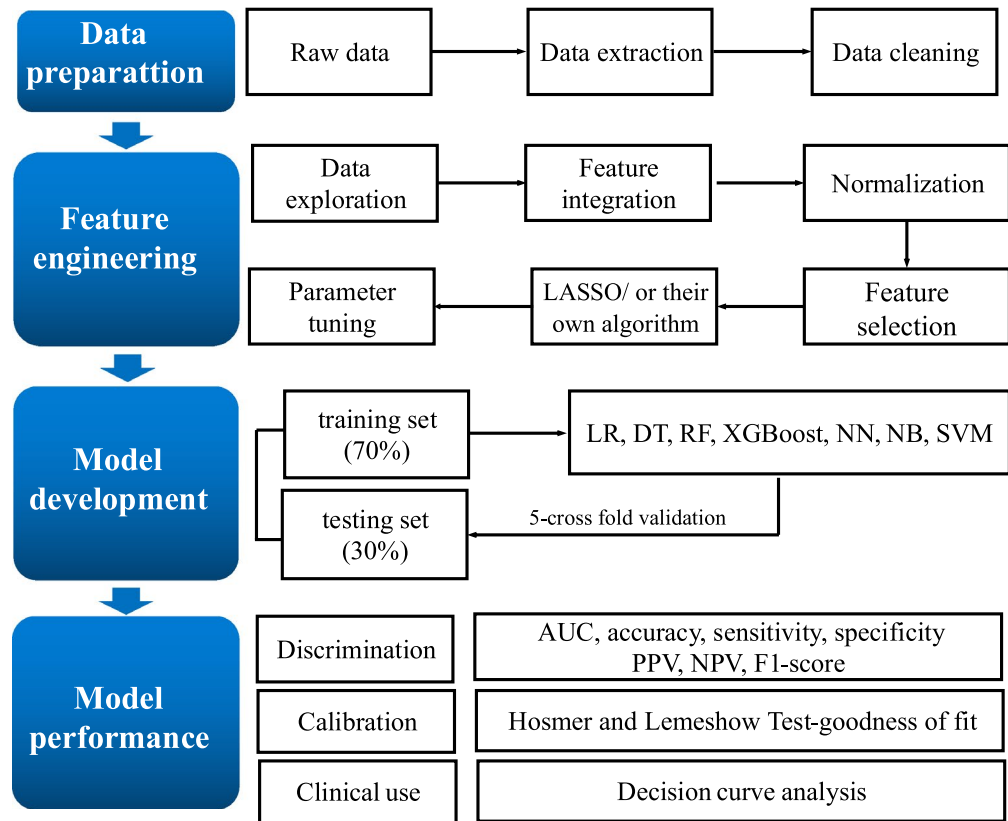
**Sleep quality.** Nocturnal sleep quality was evaluated daily by charge nurse at 9 AM using Richards-Campbell Sleep Questionnaire (RCSQ)<sup>30</sup> that comprises of five items: sleep depth, sleep latency, awakenings, returning to sleep, and sleep quality. Each item was rated on a 100-mm visual analog scale. Participants with a score of 1–25 were considered to have poor sleep quality; score of 26–50 indicated poor sleep; score of 51–75 indicated good sleep; and score of 76–100 indicated very good sleep.

**Outcome measurement.** Cognitive status was measured using Montreal Cognitive Assessment (MoCA) scores by researcher within 7 days after the transfer of the patient out of the ICU. MoCA is a brief but reliable test for evaluation of visuospatial/executive visuospatial/executive function, naming, memory, attention, language, abstraction, delayed recall and orientation, with a total score of 0–30 (the higher the score, the better is the function)<sup>31</sup>. Changsha version of MoCA-CS is a Chinese version of the original (English) version of MoCA that is adapted to Chinese culture, language, and demographic characteristics<sup>31</sup>. Use of 25/26 as the optimal cutoff value ( $\geq 26$  points: normal cognitive function;  $\leq 25$  points: cognitive impairment) was associated with sensitivity of 0.96 and specificity of 0.83 in Chinese older people. MoCA has been recommended for screening of PICS-related cognitive impairment by the Society of Critical Care Medicine consensus<sup>32</sup>.

**Statistical analysis.** All statistical analyses were conducted using R version 4.1.2 (R Foundation for Statistical Computing, Vienna, Austria). Categorical variables are presented as frequency (percentage). Normally distributed continuous variables are presented as mean  $\pm$  standard deviation and non-normally distributed continuous variables are presented as median (interquartile range). Comparisons between PICS and non-PICS group were performed using the *t* test, Mann–Whitney U-test, Fisher's exact test, or Chi squared test, as appropriate. *P* values  $< 0.05$  were considered indicative of statistical significance.

**Data preparation.** The flow chart of the probability analysis is shown in Fig. 1. We adopted the imputation and discretization methods to clean data and to deal with noise, missing values, and outliers. To ensure the facticity and reliability of the prediction model, we discarded one variable with 23.91% missing data, namely the maximum value of CRP; the missing values for 5 other variables were  $< 10\%$ , therefore, multiple imputation using the mice package was performed.

**Feature selection.** Feature selection is used to improve model performance and retrieve more clinically applicable compact models. We normalized the data and target values locked at 0–1. A total of 35 candidate features describing PICS-CI risk were collected in this study. We used least absolute shrinkage and selection operator (LASSO) regression as a variable selection mechanism for the algorithms that were unable to select features for themselves, such as XGBoost, neural network (NN), naïve Bayes (NB), and support vector machine (SVM);



**Figure 1.** Flow chart of the probability analysis. *LR* logistic regression, *DT* decision tree, *RF* random forest, *NN* neural network, *NB* naïve bayes, *SVM* support vector machine, *AUC* area under the receiver operating characteristic curve, *NPV* negative predictive value, *PPV* positive predictive value.

however, for models such as logistic regression (LR), decision tree (DT), random forest (RF), we used their own algorithms for feature selection.

**Model development.** The dataset was randomly split into two sets: the training set (70% participants) and the testing set (30% participants). Seven ensemble ML algorithms were employed to predict the probability of PICS-CI. These seven state-of-the-art algorithms were LR, DT, RF, XGBoost, neural network (NN), naïve Bayes (NB), and support vector machine (SVM). To select the optimal hyperparameters, the framework of the *caret* package was used. The conventional LR model was selected as a benchmark to compare with ML techniques. The LR model was conducted using significant variables identified by backward stepwise analysis with Chi-squared test. Then, we chose an entry probability of  $<0.05$  by the stepwise selection method, and the *glmnet* function from the *glmnet* package was used.

**Model validation.** To cope with the overfitting and inherent instability, a fivefold cross-validation procedure was applied in the training set. Subsequently, the performance of the final model was assessed, with the optimal cutoff identified by the maximum Youden index, both on the training set and the testing set.

**Model evaluation.** The performance was assessed in three domains. First, discrimination was quantified using the area under the receiver operating curve using the *ci.auc* function of the *pROC* package.  $AUC > 0.5$  indicated better predictive values; the closer the AUC value was to 1, the better the model performance. The differences between two ROC curves were assessed using the DeLong's test. More extensive metrics were generated for each model including accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1 score. Second, calibration curves were plotted to assess the calibration of the PICS-CI, accompanied with the Hosmer–Lemeshow test (a significant test statistic implies that the model does not calibrate perfectly), in order to examine the concordance between the calculated (using model) and observed probabilities of insufficient response. Third, decision curve analysis was performed to assess the benefits of clinical use. After comparing the seven models, the most accurate model was then used to construct a web-based risk calculator by *shiny* package in the R language.

**Ethics statement.** The study protocol was approved by the Fujian Provincial Hospital Institutional Review Board ([No:2016-07-001]), and the purpose of the study was explained to eligible patients and obtained written informed consent from those who agreed to participate.

## Results

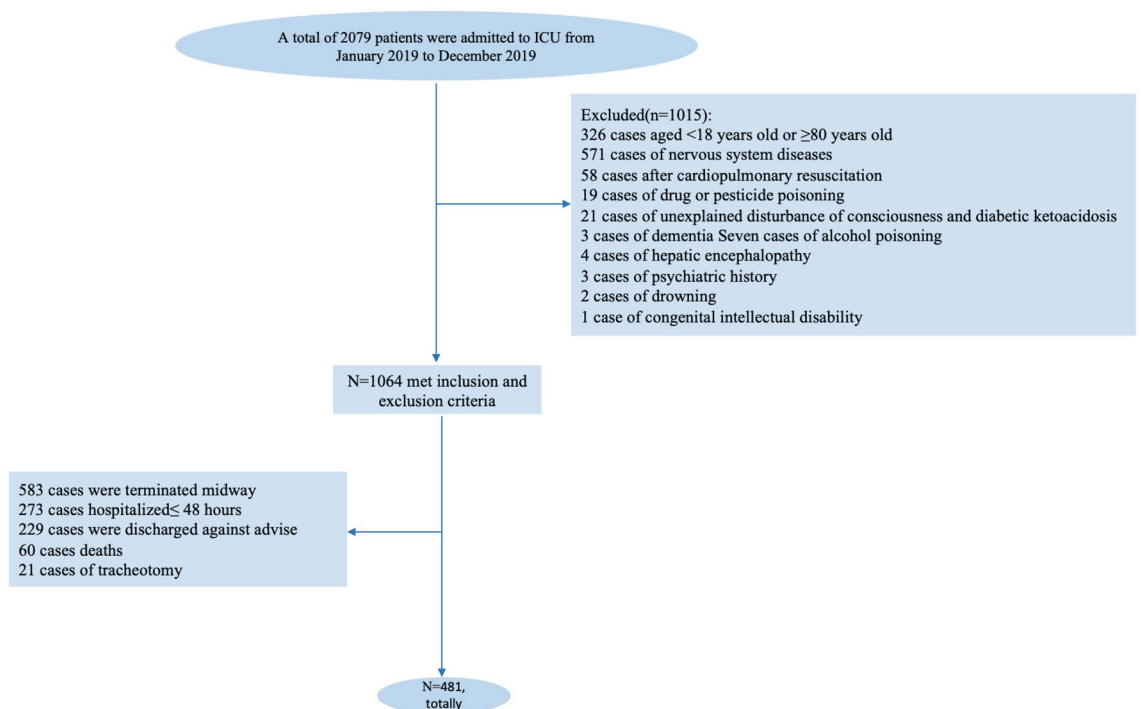
**Patient characteristics.** Among the 2079 patients admitted to ICUs at the two tertiary hospitals within the study reference period, 1064 patients qualified the study-selection criteria. Of these 583 patients were excluded due to various reasons (hospitalization < 48 h, discharge against medical advice, death, or inability to communicate due to tracheotomy), as shown in Fig. 2. Thus, a total of 481 patients were finally included; of these, 230 were diagnosed as PICS-CI while 251 had normal cognitive function. The patient characteristics are summarized in Table 1.

**Important variables selected for predicting PICS-CI.** Figure 3 depicts the four kinds of important features that were generated by LASSO, multivariate LR, DT, and RF, respectively. XGBoost, NN, NB, and SVM models were developed based on the different combination of the features that were selected by LASSO, while the LR, DT, and RF models were developed by candidate features selected by their own algorithms. Features that were important for all algorithms were duration of delirium, RCSQ, age, and sepsis.

**Discrimination of ML models for predicting PICS-CI.** In the testing set, ML models showed similar discrimination: NN (AUC: 0.906 [95% CI 0.857–0.955]), LR (AUC: 0.898 [95% CI 0.847–0.949]), SVM (AUC: 0.895 [95% CI 0.843–0.947]), NB (AUC: 0.877 [95% CI 0.822–0.932]), XGBoost (AUC: 0.866 [95% CI 0.807–0.925]), RF (AUC: 0.865 [95% CI 0.807–0.922]), DT (AUC: 0.822 [95% CI 0.752–0.892]), shown in Fig. 4. The Delong test showed that only DT model had significant difference (significantly lower AUC) when compared with LR technique ( $P < 0.05$ ), whereas, other ML models showed no significant difference in the discrimination ability ( $P > 0.05$ ). In terms of confusion matrix, LR showed the highest sensitivity of 0.899, but lowest specificity of 0.750; other ML models produced opposite results, i.e., poor sensitivity (0.594–0.783) but excellent specificity (0.895–1.00), shown in Table 2.

**Calibration of ML models for predicting PICS-CI.** Hosmer and Lemeshow Test-goodness of fit demonstrated that LR, RF, NN, and SVM models have good agreement between the predicted probability and observed incidence of PICS-CI ( $P > 0.05$ ); nevertheless, unacceptable calibration was observed for DT, XGBoost, and NB models ( $P < 0.05$ ) (Fig. 5).

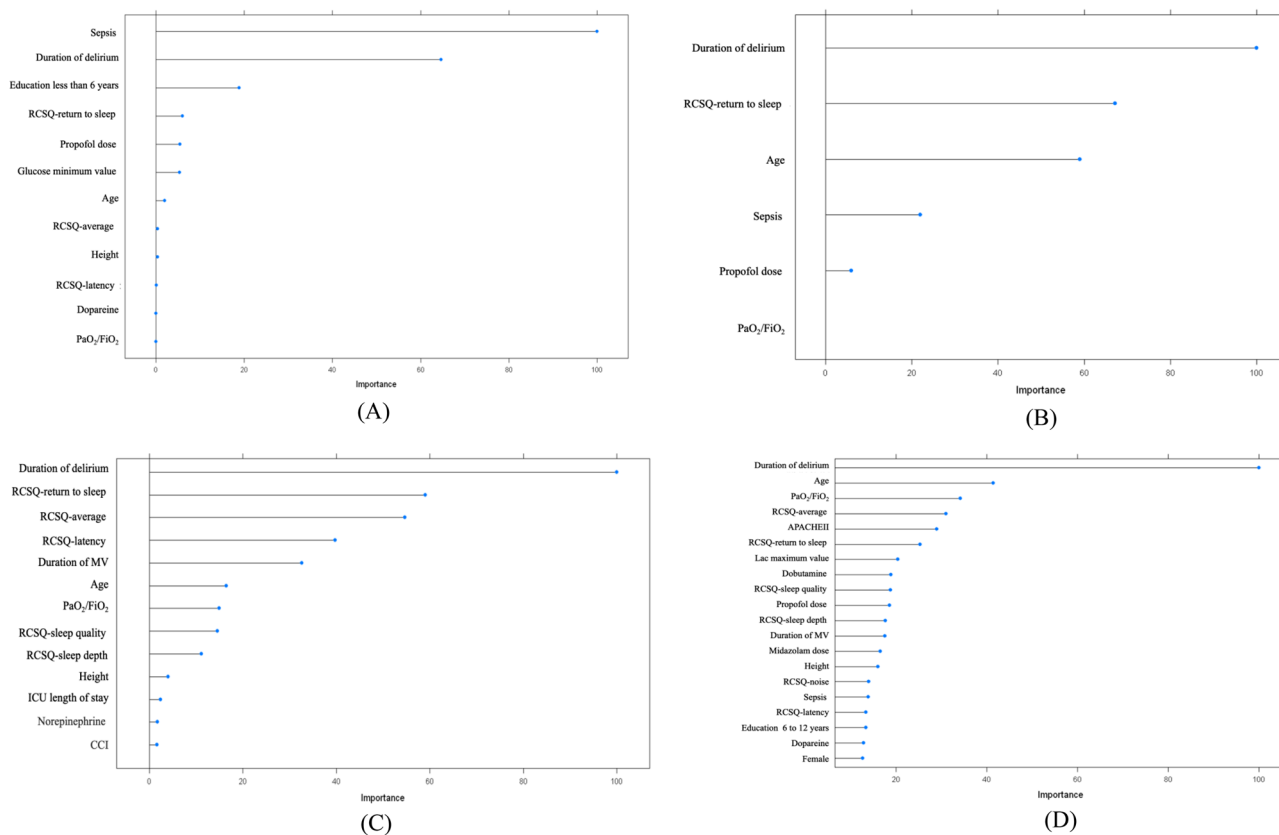
**Clinical use of ML models for predicting PICS-CI.** The decision curve analysis for predicting PICS-CI is presented in Fig. 6. Most of the ML models of LR, RF, XGBoost, NN, and SVM showed a higher net benefit



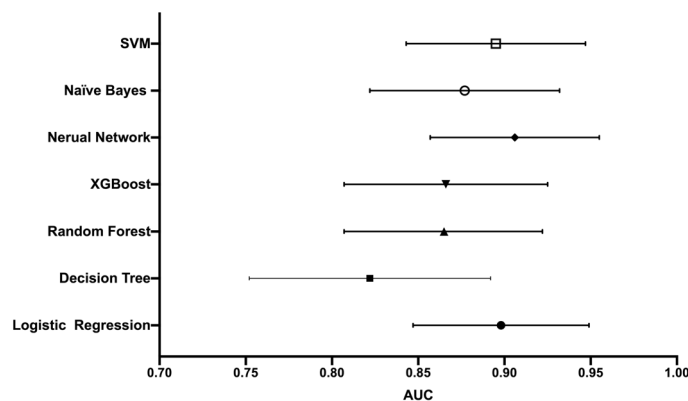
**Figure 2.** Flowchart of participants.

Features		PICS-CI		$\chi^2/Z$	P value
		No n = 251	Yes n = 230		
Continuous variable, [N (%)]					
Sex	Male	170 (67.7)	153 (66.5)	0.079	0.778
Marital status	Married	230 (91.6)	215 (93.5)	2.432	0.272
	Widowed	4 (1.6)	6 (2.6)		
	Others	17 (6.8)	9 (3.9)		
Education	Less than 6 years	80 (31.9)	111 (48.3)	14.844	0.001
	6 to 12 years	149 (59.3)	109 (47.4)		
	More than 12 years	22 (8.8)	10 (4.3)		
Smoking		63 (25.1)	51 (77.8)	0.568	0.451
Sepsis		20 (8.0)	39 (17.0)	9.01	0.003
Diagnosis	Infectious disease	20 (8.0)	29 (12.6)	8.25	0.220
	Respiratory diseases	43 (17.1)	33 (14.3)		
	Kidney disease	7 (2.8)	5 (2.2)		
	Traumatic disease	20 (8.0)	13 (5.7)		
	Digestive system diseases	55 (21.9)	43 (18.7)		
	Cardiovascular diseases	89 (35.5)	98 (42.6)		
	Others	17 (6.7)	9 (3.9)		
T $\geq$ 38 °C		111 (44.2)	122 (53.0)	3.739	0.053
Categorical variable, [M(IQR)]					
Age, year		56.0 (22.0)	63.0 (18.0)	-5.11	0.00
Height		168.0 (13.0)	167.0 (11.3)	-1.83	0.07
Weight		64.0 (13.0)	63.0 (14.0)	-0.79	0.43
APACHEII		14.0 (8.0)	16.0 (10.0)	-2.65	0.01
SOFA		5.0 (5.0)	6.0 (5.0)	-2.73	0.01
CCI		2.0 (3.0)	2.0 (2.0)	-1.33	0.18
ICU length of stay, day		7.0 (8.0)	8.0 (10.3)	-2.60	0.01
Duration of mechanical ventilation, hours		0.0 (18.0)	45.0 (120.0)	-7.46	0.00
Sedative					
Propofol dose, g		0.0 (1.0)	1.4 (4.6)	-7.25	0.00
Midazolam dose, mg		0.0 (20.0)	20.00 (260.0)	-5.94	0.00
Analgesic					
Remifentanyl dose, mg		0.0 (2.0)	1.0 (9.0)	-4.36	0.00
Sufentanyl dose, mg		0.0 (300.0)	200.0 (600.0)	-4.69	0.00
Norepinephrine		2.0 (30.0)	8.0 (66.0)	-4.12	0.00
Dopamine		0.0 (20.0)	0.0 (20.0)	-3.36	0.00
Dobutamine		0.0 (0.0)	0.0 (0.0)	-4.18	0.00
Duration of delirium, days		0.0 (0.0)	3.0 (6.0)	-13.77	0.00
0		254 (69.4)	112 (30.6)	-7.48	0.00
<3		5 (21.7)	18 (78.3)		
3-7		4 (6.8)	55 (93.2)		
>7		3 (9.1)	30 (90.9)		
RCSQ-sleep depth		65.0 (20.0)	55.0 (20.0)	-7.48	0.00
RCSQ-sleep latency		66.0 (20.0)	55.0 (20.0)	-7.36	0.00
RCSQ-awakenings		70.0 (15.0)	60.0 (15.0)	-6.75	0.00
RCSQ-returning to sleep		70.0 (15.0)	60.0 (15.0)	-8.71	0.00
RCSQ-sleep quality		70.0 (15.0)	60.0 (15.0)	-8.19	0.00
RCSQ-average		69.0 (16.0)	56.1 (13.3)	-8.67	0.00
RCSQ-noise		60.0 (15.0)	55.0 (10.0)	-6.58	0.00
Glucose, mmol					
Minimum value		5.2 (1.7)	5.4 (1.7)	-1.33	0.18
Maximum value		14.1 (7.3)	15.9 (7.9)	-3.76	0.00
PCT maximum value, ng/ml		1.8 (8.7)	4.3 (15.7)	-3.48	0.00
WBC maximum value, count		15.3 (8.0)	16.9 (8.2)	-2.63	0.01
Lac, mmol		2.9 (5.8)	3.9 (8.5)	-2.50	0.01
PaO <sub>2</sub> /FiO <sub>2</sub>		170.0 (113.0)	152.2 (118.0)	-3.01	0.00

**Table 1.** Characteristics of the study population.



**Figure 3.** Importance of the candidate features. (A) Feature selection by LASSO, (B) feature selection by logistic regression, (C) feature selection by decision tree, (D) feature selection by random forest.



**Figure 4.** AUC of machine learning models predicting PICS-CI of testing set.

compared to both treat all (grey line) and treat none (black line) with a wide range of reasonable thresholds probability; however, the benefit of DT and NB models was comparatively lower, as depicted in Fig. 6.

**Web-based risk calculator of LR technique.** Considering the performance and complexity of most ML models, we regarded LR as an accurate and pragmatic model, and given the number of variables included in the final LR model, a simple web-based risk calculator (rather than a score chart or nomogram) was developed and housed at the website: <https://model871010.shinyapps.io/dynnomapp/>. The website offers a visual representation of the predicted probability of PICS-CI and its 95% prediction interval based on the variables' input (Fig. 7).

	Type	Cut off	Specificity	Sensitivity	Accuracy	NPV	PPV	Precision	Recall	F1 score	AUC (95%CI)
Training set	LR	0.453	0.880	0.752	0.819	0.794	0.852	0.852	0.752	0.799	0.896 (0.863–0.928)
	DT	0.550	0.926	0.758	0.845	0.806	0.904	0.904	0.758	0.824	0.878 (0.840–0.915)
	RF	0.391	0.903	0.969	0.936	0.969	0.902	0.902	0.969	0.934	0.979 (0.968–0.991)
	XGBoost	0.403	0.874	0.857	0.866	0.869	0.863	0.863	0.857	0.860	0.935 (0.910–0.960)
	NN	0.510	0.920	0.752	0.839	0.801	0.896	0.896	0.752	0.818	0.905 (0.874–0.936)
	NB	0.037	0.811	0.776	0.795	0.798	0.791	0.791	0.776	0.784	0.862 (0.823–0.901)
	SVM	0.370	0.817	0.820	0.819	0.831	0.805	0.805	0.820	0.812	0.901 (0.869–0.933)
Testing set	LR	0.369	0.750	0.899	0.821	0.891	0.765	0.765	0.899	0.827	0.898 (0.847–0.949)
	DT	0.926	1.000	0.594	0.807	0.731	1.000	1.000	0.594	0.746	0.822 (0.752–0.892)*
	RF	0.545	0.934	0.638	0.793	0.740	0.898	0.898	0.638	0.746	0.865 (0.807–0.922)
	XGBoost	0.587	0.961	0.638	0.807	0.745	0.936	0.936	0.638	0.759	0.866 (0.807–0.925)
	NN	0.551	0.895	0.783	0.841	0.819	0.871	0.871	0.783	0.824	0.906 (0.857–0.955)
	NB	0.307	0.934	0.681	0.814	0.763	0.904	0.904	0.681	0.777	0.877 (0.822–0.932)
	SVM	0.480	0.908	0.754	0.835	0.802	0.881	0.881	0.754	0.813	0.895 (0.843–0.947)

**Table 2.** The performance of machine learning models. *LR* logistic regression, *DT* decision tree, *RF* random forest, *NN* neural network, *NB* naïve bayes, *SVM* support vector machines, *NPV* negative predictive value, *PPV* positive predictive value. \*The AUC of DT model was significantly different from that of LR algorithm (Delong test), while others not.

## Discussion

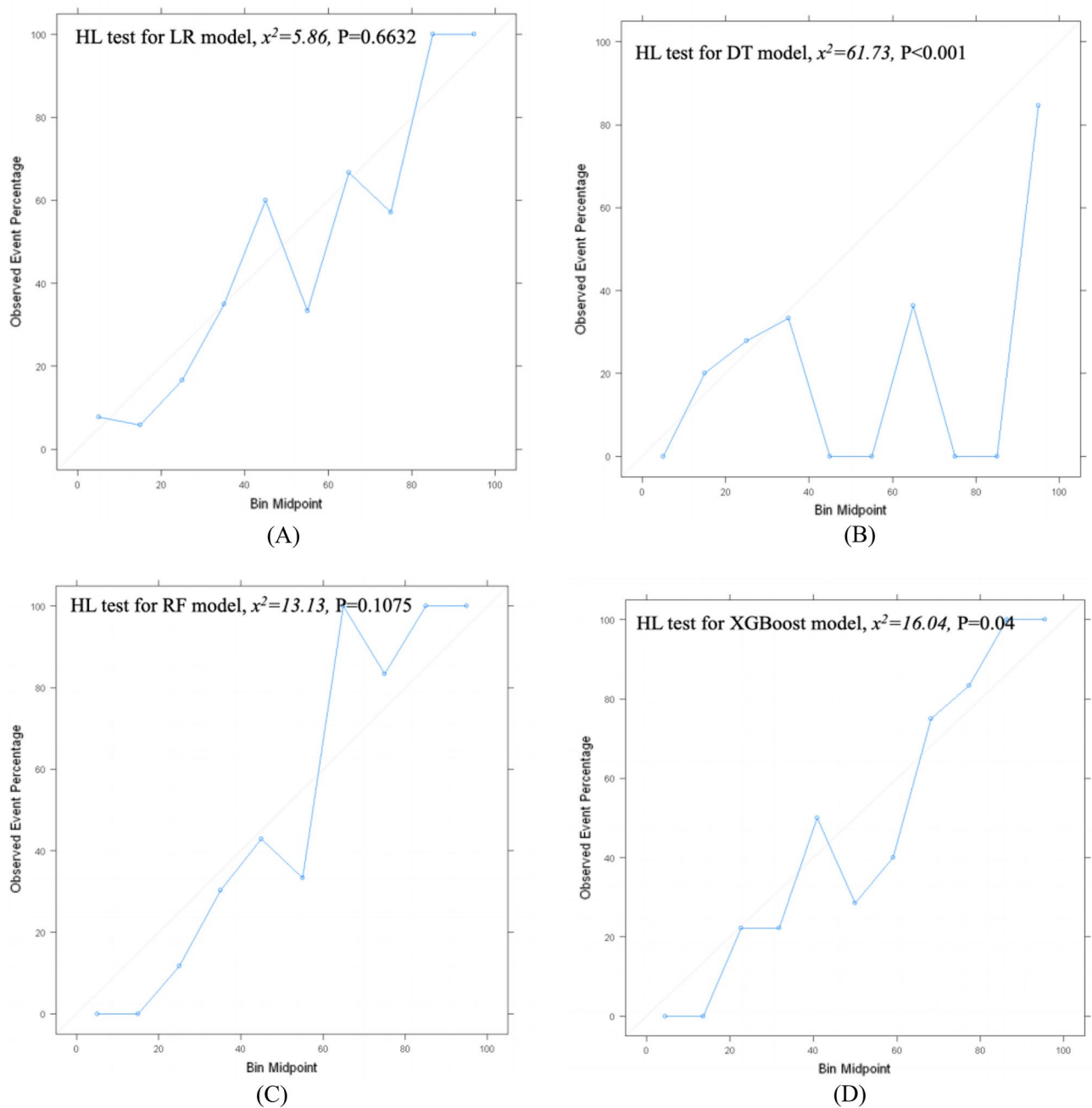
In the present study, we explored and compared the applicability of seven ML methods to predict the risk of PICS-CI in a low dimensional data. LR models exhibited accuracy and similar discrimination, good calibration and clinical utility among the seven ML models. Given the risk of overfitting and the lack of interpretability of some ML models, a simplified web risk calculator based on LR was proposed, which is convenient to implement in routine practice. Therefore, traditional regression models should continue to play a key role in disease risk prediction if and when low-dimensional data are used.

To the best of our knowledge, this was the first study to develop and compare ML models to predict the risk of PICS-CI. Our analysis indicated that the performance of complex ML models is not consistently better than traditional LR models in predicting PICS-CI events. Indeed, some models had lower performance, such as the DT model which significantly underperformed the LR model ( $P < 0.05$  for Delong test). One potential reason why ML classification algorithms did not outperform traditional LR models in our study may be that PICS-CI is an easily predictable behavior, and all the models were performing similarly well (AUC range: 0.822–0.906). It is also possible that we have collected sensitive variables predicting PICS-CI, and these variable domains can improve the prediction of appropriate PICS-CI. In fact, the added value of large datasets and the increase in algorithmic complexity may be negligible if the relevant features are not included<sup>33</sup>.

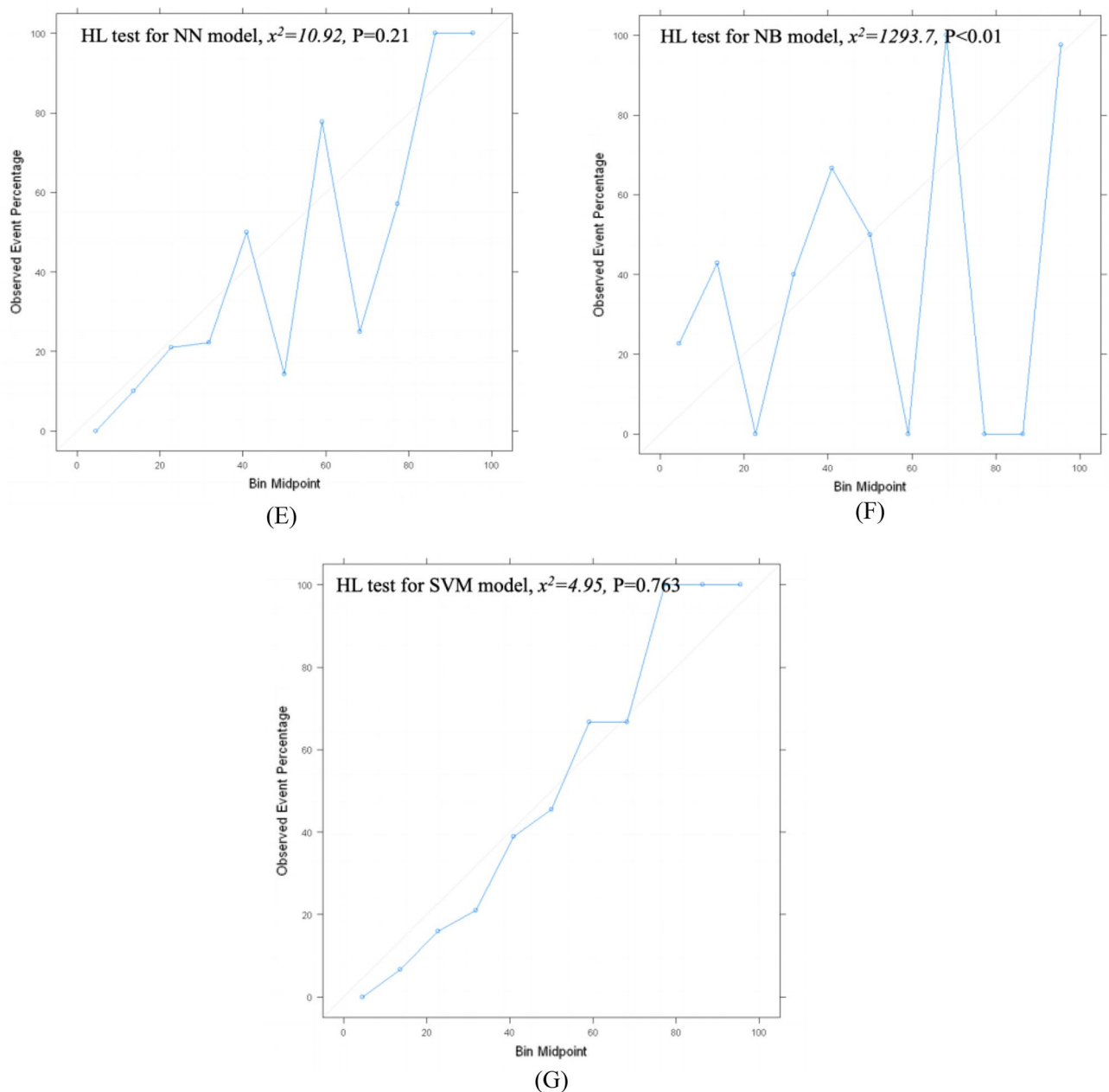
We identified a potentially modifiable or unchangeable risk factor for post-ICU cognitive impairment. We found that duration of delirium, poor RCSQ, advanced age, and sepsis were the most frequent and important candidates. In the present study, for each additional day of delirium, the risk of PICS-CI was increased by 2.64 times. Several prior studies have yielded consistent results. In a single-center prospective study of 409 ICU patients, compared with no delirium experience, the occurrence of PICS-CI increased by 36.849 times with a wide 95% CI when the duration of delirium was greater than 8 days<sup>34</sup>. A meta-analysis even suggested that delirium is likely to be the sole risk factor for PICS-CI<sup>14</sup>. We used RCSQ to measure patient perception of sleep quality, including sleep depth, sleep latency, awakenings, returning to sleep, and sleep quality<sup>35</sup>. We found that patient self-reported sleep quality was also associated with PICS-CI. One possibility is that sleep deprivation and delirium share many common symptoms; both these factors are likely to have a bidirectional effect and may create a potential vicious cycle in critically-ill patients<sup>36</sup>. Sleep fragmentation has also been shown to be associated with worse cognitive performance in hospital at 7 days of ICU discharge<sup>37</sup>. However, Yao et al. found that the RCSQ is not effective in predicting PICS-CI; this may be due to the fact that they just assessed sleep abnormalities occurring in critically-ill patients within 3 days after ICU discharge<sup>34</sup>. In our study, advanced age was found to be a risk factor for PICS-CI. This result should be interpreted with caution, because cognitive function is an age-dependent syndrome. In addition, we did not measure the cognitive function prior to ICU admission, so it is difficult to determine that it was acquired after critical illness. Sepsis-related encephalopathy is characterized by pathological behavior ranging from delirium to coma, and may lead to long-term cognitive impairment<sup>38</sup>, affecting up to 50% of patients during the course of sepsis<sup>39</sup>. The pathogenesis is due to a dysregulated host response and the absence of direct central nervous system infection, resulting in residual diffuse brain dysfunction<sup>40</sup>.

Second, the inherent flexibility and scope for automation make ML well suited to handling complex high-dimensional data (i.e., with many variables or features, perhaps more than 50 predictors) that would be challenging for conventional approaches<sup>26,33</sup>. The flexibility of ML algorithms is manifested in their ability to capture complex nonlinear and interactive effects<sup>41</sup>, which may be a common feature of high-dimensional data. In the present study, we used a limited set of clinical predictors ( $n = 35$ ) which may have contributed to the comparable performance of ML and traditional method. In a study using large data set ( $n = 11,022$ ) with low dimensional





**Figure 5.** Calibration curve and HL-test result of machine learning models of testing set. (A) for logistic regression, (B) for decision tree, (C) for random forest, (D) for XGBoost, (E) for neural network, (F) for naïve bayes, (G) for support vector machine.

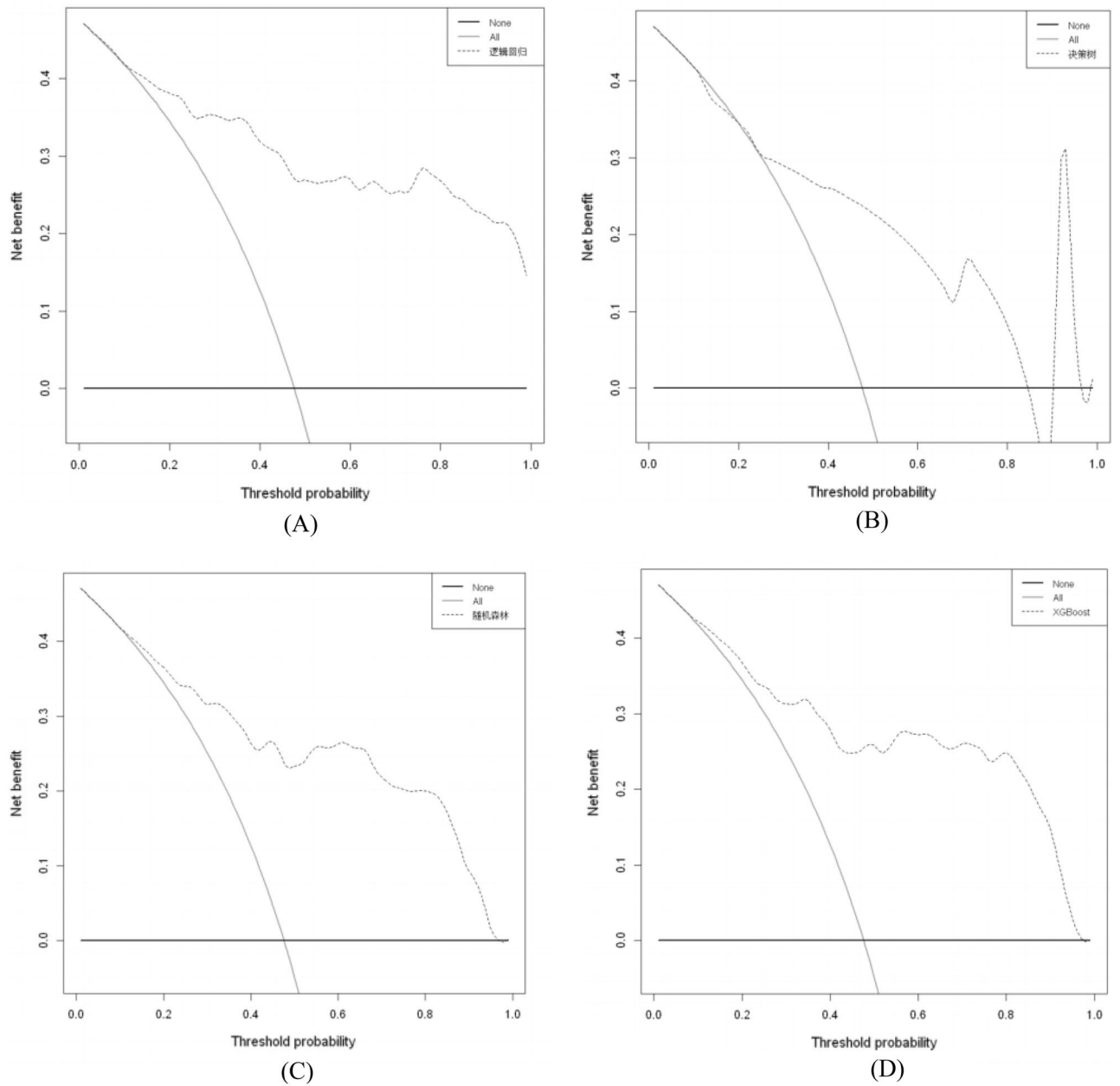


**Figure 5.** (continued)

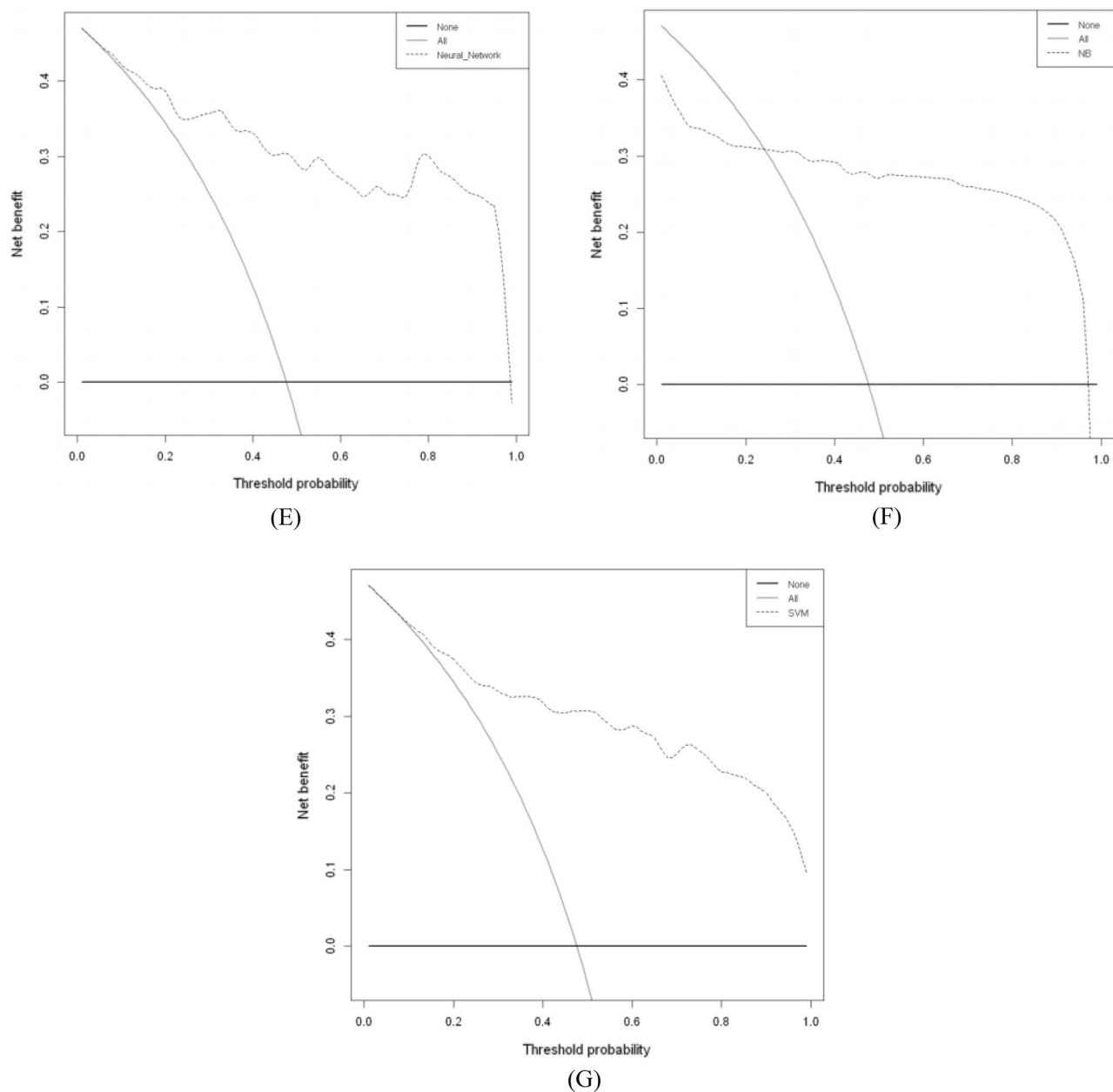
data (only 11 variables), ML methods were not found to improve prognostic prediction as well<sup>26</sup>. As suggested by previous studies, when the clinical data is low-dimensional data (often with limited data set, a small number of clinical predictors, few nonlinear and interactive effects), LR is an easy-to-use and appropriate model for disease risk prediction<sup>28</sup>. The ability of ML to process high-dimensional data should not distract from the often-greater benefits of traditional LR.

Finally, despite the good discriminatory power of some ML models, for instance, DT (AUC: 0.822 [95% CI 0.752–0.892]), XGBoost (AUC: 0.866 [95% CI 0.807–0.925]), and NB (AUC: 0.877 [95% CI 0.822–0.932]), these models may not end up being used, because these clearly underperformed in terms of calibration ( $P < 0.05$ ). This indicates that overfitting is a common problem arising in small data sets<sup>42</sup>. In contrast, DT, XGBoost, and NB methods have exhibited better performance in other studies<sup>43–45</sup>, which however did not assess calibration. A recent systematic review<sup>28</sup> did not find an incremental value of flexible ML techniques over traditional statistical methods in relatively small data sets (median sample size: 1250); moreover, calibration was not addressed in 79% of ML studies. Calibration evaluation is a critical step before implementing predictive models in clinical practice. Moreover, reporting guidelines for building predictive models using ML methods recommend reporting calibration<sup>46</sup>.

Given the lack of interpretability of some ML models, we therefore presented a web-based risk calculator based on LR model, which could allow for a more precise predicted probability compared to a points-based



**Figure 6.** Decision curve analysis of machine learning models of testing set. The y-axis measures the net benefit. DCA shows the clinical usefulness of the machine learning models, according to a continuum of potential thresholds for PICS-CI risk (x-axis) and the net benefit of using the prediction model to stratify patients according to risk (y-axis). Transverse line represents the assumption that no patients have PICS-CI; oblique line represents the assumption that all patients have PICS-CI; dotted line represents the PICS-CI model). (A) For logistic regression, (B) for decision tree, (C) for random forest, (D) for XGBoost, (E) for neural network, (F) for naïve bayes, (G) for support vector machine.

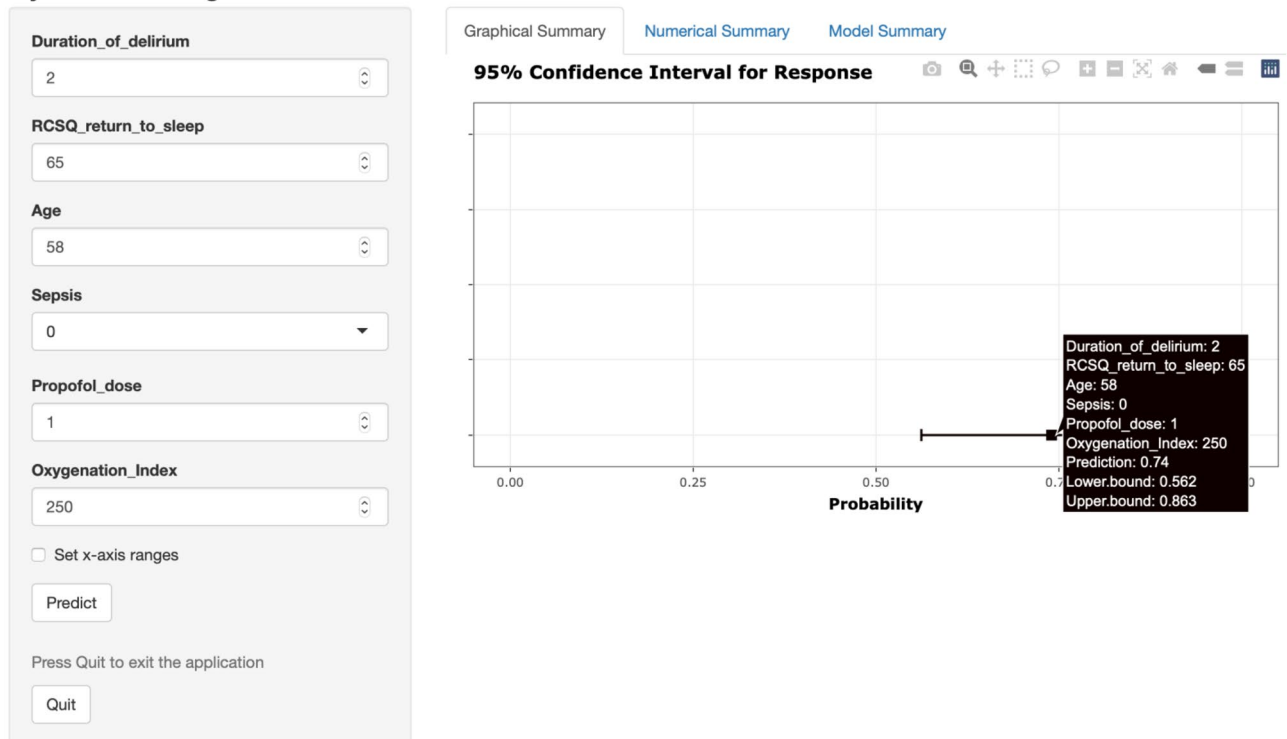


**Figure 6.** (continued)

scoring system. For example, a 72-year-old critically ill patient complicated with delirium for 3 days and having fair sleep quality (RCSQ-returning to sleep = 60), no sepsis, no propofol, and minimum value of  $PO_2/FiO_2$  in daily assessment, would have an 87% risk of PICS-CI. However, in another patient with similar characteristics, except having no delirium, the observed probability would reduce to 49%.

Some limitations of our study should be acknowledged. First, we considered only a limited number of variables for predicting PICS-CI. We restricted our analyses to predictive modeling with known or potential risk factors. This may limit the generalizability of our conclusion to data sets with more variables. Second, though the several ML models demonstrated perfect performance in internal validity, further research should use a prospective external validation to check for web-based risk calculator for routine clinical use. Third, the present study was unable to measure cognitive function prior to ICU admission which may have possibly contributed to overestimation of the incidence of ICU-acquired cognitive impairment. Of note, we excluded patients aged  $\geq 80$  years and those with nervous system disease as in these patients, the cognitive impairment may be due to physiology or the disease itself rather than ICU stay. Finally, we only investigated short-term cognitive impairment during hospitalization and did not conduct a long-term follow-up after discharge.

## Dynamic Nomogram



**Figure 7.** Web-based risk calculator of PICS-CI (Example of a 58-year-old critically ill patient complicated with delirium for 2 days and having fair sleep quality (RCSQ-returning to sleep = 65), no sepsis, use propofol, and  $PO_2/FiO_2$  value of 250 in daily assessment; this patient would have a 74% risk of PICS-CI).

## Conclusion

Our results indicate that the risk of early cognitive impairment post ICU hospitalization can be parsimoniously assessed by several ML prediction models. Indeed, in a low-dimensional and simple limited data setting, complex ML algorithms performed no better than traditional LR model for prediction of PICS-CI. We hope that LR model with its accompanying web-based risk calculator will facilitate prevention and early recognition of PICS-CI.

## Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 31 August 2022; Accepted: 18 January 2023

Published online: 11 February 2023

## References

1. Yuan, C., Timmins, F. & Thompson, D. R. Post-intensive care syndrome: A concept analysis. *Int. J. Nurs. Stud.* **114**, 9 (2021).
2. Karnatovskaia, L. V., Johnson, M. M., Benzo, R. P. & Gajic, O. The spectrum of psychocognitive morbidity in the critically ill: A review of the literature and call for improvement. *J. Crit. Care* **30**(1), 130–137 (2015).
3. Barr, J. *et al.* Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. *Crit. Care Med.* **41**(1), 263–306 (2013).
4. Hopkins, R. O. & Girard, T. D. Medical and economic implications of cognitive and psychiatric disability of survivorship. *Semin. Respir. Crit. Care Med.* **33**(4), 348–356 (2012).
5. Sakuramoto, H., Subrina, J., Unoki, T., Mizutani, T. & Komatsu, H. Severity of delirium in the ICU is associated with short term cognitive impairment. A prospective cohort study. *Intens. Crit. Care Nurs.* **31**(4), 250–257 (2015).
6. Girard, T. D. *et al.* Delirium as a predictor of long-term cognitive impairment in survivors of critical illness. *Crit. Care Med.* **38**(7), 1513–1520 (2010).
7. Rothenhäusler, H. B., Ehrentraut, S., Stoll, C., Schelling, G. & Kapfhammer, H. P. The relationship between cognitive performance and employment and health status in long-term survivors of the acute respiratory distress syndrome: Results of an exploratory study. *Gen. Hosp. Psychiatry* **23**(2), 90–96 (2001).
8. Pandharipande, P. P. *et al.* Long-term cognitive impairment after critical illness. *N. Engl. J. Med.* **369**(14), 1306–1316 (2013).
9. Wilcox, M. E., Girard, T. D. & Hough, C. L. Delirium and long term cognition in critically ill patients. *BMJ* **373**, n1007 (2021).
10. Meeks, J. R. *et al.* Long-term cognitive impairment associated with delirium in acute neurological injury. *Crit. Care Explor.* **2**(6), e0130 (2020).
11. Jackson, J. C. *et al.* Depression, post-traumatic stress disorder, and functional disability in survivors of critical illness in the BRAIN-ICU study: A longitudinal cohort study. *Lancet Respir. Med.* **2**(5), 369–379 (2014).
12. Sakusic, A. & Rabinstein, A. A. Cognitive outcomes after critical illness. *Curr. Opin. Crit. Care* **24**(5), 410–414 (2018).
13. Kosilek, R. P., Schmidt, K., Baumeister, S. E., Gensichen, J. & Group, S. S. Frequency and risk factors of post-intensive care syndrome components in a multicenter randomized controlled trial of German sepsis survivors. *J. Crit. Care* **65**, 268–273 (2021).

14. Lee, M., Kang, J. & Jeong, Y. J. Risk factors for post-intensive care syndrome: A systematic review and meta-analysis. *Aust. Crit. Care* **33**(3), 287–294 (2020).
15. Haines, K. J. *et al.* Prediction models for physical, cognitive, and mental health impairments after critical illness: A systematic review and critical appraisal. *Crit. Care Med.* **48**(12), 1871–1880 (2020).
16. Obermeyer, Z. & Emanuel, E. J. Predicting the future—Big data, machine learning, and clinical medicine. *N. Engl. J. Med.* **375**(13), 1216–1219 (2016).
17. Sanchez-Pinto, L. N., Luo, Y. & Churpek, M. M. Big data and data science in critical care. *Chest* **154**(5), 1239–1248 (2018).
18. Zhang, Z., Ho, K. M. & Hong, Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. *Crit. Care* **23**(1), 112 (2019).
19. Fluren, L. M. *et al.* Machine learning for the prediction of sepsis: A systematic review and meta-analysis of diagnostic test accuracy. *Intens. Care Med.* **46**(3), 383–400 (2020).
20. Wu, T. T., Lin, X. Q., Mu, Y., Li, H. & Guo, Y. S. Machine learning for early prediction of in-hospital cardiac arrest in patients with acute coronary syndromes. *Clin. Cardiol.* **44**(3), 349–356 (2021).
21. Lee, H. C. *et al.* Prediction of acute kidney injury after liver transplantation: Machine learning approaches vs logistic regression model. *J. Clin. Med.* **7**(11), 428 (2018).
22. Racine, A. M. *et al.* Machine learning to develop and internally validate a predictive model for post-operative delirium in a prospective, observational clinical cohort study of older surgical patients. *J. Gen. Intern. Med.* **36**(2), 265–273 (2021).
23. Kanwal, F. *et al.* Development, validation, and evaluation of a simple machine learning model to predict cirrhosis mortality. *JAMA Netw. Open* **3**(11), e2023780 (2020).
24. Mamlook, R. E. A., Wells, L. J. & Sawyer, R. Machine-learning models for predicting surgical site infections using patient pre-operative risk and surgical procedure factors. *Am. J. Infect. Control.* <https://doi.org/10.1016/j.ajic.2022.08.013> (2022).
25. Sun, Z. *et al.* Comparing machine learning models and statistical models for predicting heart failure events: A systematic review and meta-analysis. *Front. Cardiovasc. Med.* **9**, 812276 (2022).
26. Gravestijn, B. Y. *et al.* Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. *J. Clin. Epidemiol.* **122**, 95–107 (2020).
27. Cote, M. *et al.* Are machine learning algorithms more accurate in predicting vegetable and fruit consumption than traditional statistical models? An exploratory analysis. *Front. Nutr.* **9**, 740898 (2022).
28. Christodoulou, E. *et al.* A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J. Clin. Epidemiol.* **110**, 12–22 (2019).
29. Ely, E. W. *et al.* Delirium in mechanically ventilated patients: Validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU). *JAMA* **286**(21), 2703–2710 (2001).
30. Richards, K. C., O'Sullivan, P. S. & Phillips, R. L. Measurement of sleep in critically ill patients. *J. Nurs. Meas.* **8**(2), 131–144 (2000).
31. Cai, Y. J. Z. *et al.* Study on multi-center investigation of the Chinese version of national Institutes of health stroke scale. *J. Tradit. Chin. Med.* **31**, 494–498 (2008).
32. Mikkelsen, M. E. *et al.* Society of Critical Care Medicine's International Consensus Conference on prediction and identification of long-term impairments after critical illness. *Crit. Care Med.* **48**(11), 1670–1679 (2020).
33. Wilkinson, J. *et al.* Time to reality check the promises of machine learning-powered precision medicine. *Lancet Dig. Health* **2**(12), e677–e680 (2020).
34. Yao, L. *et al.* Incidence and influencing factors of post-intensive care cognitive impairment. *Intens. Crit. Care Nurs.* **67**, 103106 (2021).
35. Chen, L. X. *et al.* Richards–Campbell sleep questionnaire: Psychometric properties of Chinese critically ill patients. *Nurs. Crit. Care* **24**(6), 362–368 (2019).
36. Pisani, M. A. & D'Ambrosio, C. Sleep and delirium in adults who are critically ill: A contemporary review. *Chest* **157**(4), 977–984 (2020).
37. Wilcox, M. E. *et al.* Sleep fragmentation and cognitive trajectories after critical illness. *Chest* **159**(1), 366–381 (2021).
38. Heming, N. *et al.* Neuroanatomy of sepsis-associated encephalopathy. *Crit. Care* **21**(1), 65 (2017).
39. Sonnevile, R. *et al.* Potentially modifiable factors contributing to sepsis-associated encephalopathy. *Intens. Care Med.* **43**(8), 1075–1084 (2017).
40. Chung, H. Y., Wickel, J., Brunkhorst, F. M. & Geis, C. Sepsis-associated encephalopathy: From delirium to dementia? *J. Clin. Med.* **9**(3), 703 (2020).
41. Nusinovi, S. *et al.* Logistic regression was as good as machine learning for predicting major chronic diseases. *J. Clin. Epidemiol.* **122**, 56–69 (2020).
42. Steyerberg, E. W. *et al.* Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology* **21**(1), 128–138 (2010).
43. Hou, N. *et al.* Predicting 30-days mortality for MIMIC-III patients with sepsis-3: A machine learning approach using XGboost. *J. Transl. Med.* **18**(1), 462 (2020).
44. Harada, D., Asanoi, H., Noto, T. & Takagawa, J. Naive bayes prediction of the development of cardiac events in heart failure with preserved ejection fraction in an outpatient clinic-beyond B-type natriuretic peptide. *Circ. J.* **86**(1), 37–46 (2021).
45. Rodrigo, H., Beukes, E. W., Andersson, G. & Manchaiah, V. Exploratory data mining techniques (decision tree models) for examining the impact of internet-based cognitive behavioral therapy for tinnitus: Machine learning approach. *J. Med. Internet Res.* **23**(11), e28999 (2021).
46. Luo, W. *et al.* Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. *J. Med. Internet Res.* **18**(12), e323 (2016).

## Author contributions

H.L. substantively conceived, designed, and led the study, in addition to being a main contributor in writing the manuscript. T.T.W. and Y.Q.W. substantively contributed to the methodology of the study, reviewed data analysis by R software, further analyzed and interpreted the collected data using factor analysis, and was a major contributor in writing the manuscript. Y.Q.W., J.B.W., B.L.Y. substantively contributed to the acquisition of data, oversight of the data collection. H.L. substantively revised the manuscript. All authors have read and approved the final manuscript.

## Funding

This study was supported by Grants from the Guided Project of Fujian Provincial Department of Science and Technology [2021Y0051].

## Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to H.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023