Clinical Interventions in Aging

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Comprehensive Machine Learning-Based Prediction Model for Delirium Risk in Older Patients with Dementia: Risk Factors Identification

Qifan Xiao[1], Shirui Zhou[2], Bin Tang[1], Yuqing Zhu[1]

[1]General Practice, International Department, China-Japan Friendship Hospital, Beijing, People's Republic of China; [2]Graduate School, Beijing University of Chinese Medicine, Beijing, People's Republic of China

Correspondence: Yuqing Zhu, International Department of China-Japan Friendship Hospital, No. 2 Yinghua East Street, Chaoyang District, Beijing, 100029, People's Republic of China, Email circe_77@126.com

Background: Delirium superimposed on dementia (DSD) is a severe complication in older adults with dementia, marked by fluctuating cognition, inattention, and altered consciousness. Detection is challenging due to symptom overlap, yet it contributes to cognitive decline, prolonged hospitalization, and increased mortality. Identifying key risk factors and developing an accurate prediction model is crucial for timely intervention. This study aimed to establish a machine learning-based model to predict delirium risk, focusing on significant predictors to aid clinical decision-making.

Methods: We prospectively collected clinical data from 636 older dementia patients. Five machine learning algorithms—Extreme Gradient Boosting (XGB), Random Forest (RF), Multilayer Perceptron (MLP), Categorical Boosting (CB), and Logistic Regression (LR)—were used to construct prediction models. Feature importance was analyzed using SHapley Additive exPlanations (SHAP) to identify key risk factors. Data included demographic information, biochemical parameters, comorbidities, medication history, and Visual Analogue Scale (VAS) scores.

Results: The final analysis included 636 older dementia patients, with a mean age of 78.2 ± 6.3 years, of whom 187 (29.4%) developed delirium during hospitalization. The XGB model demonstrated the best performance, achieving the highest area under the receiver operating characteristic curve (0.930), accuracy (0.870), F1 score (0.892), and area under the precision-recall curve (0.989). The Brier score for the XGB model was 0.08. The SHAP method identified cerebrovascular disease, sedative drug use, hemoglobin levels, VAS score ≥4, superoxide dismutase, diabetes, hsCRP, hypertension, family presence, and hyperlipidemia as the most significant risk factors for delirium. The top 10 variables were used to construct a compact XGB model, which also exhibited good predictive performance.

Conclusion: This study developed a machine learning-based prediction model for delirium risk in older dementia patients, with the XGB model demonstrating the best performance. The identified key risk factors provide insights for early intervention, potentially improving delirium management in clinical practice.

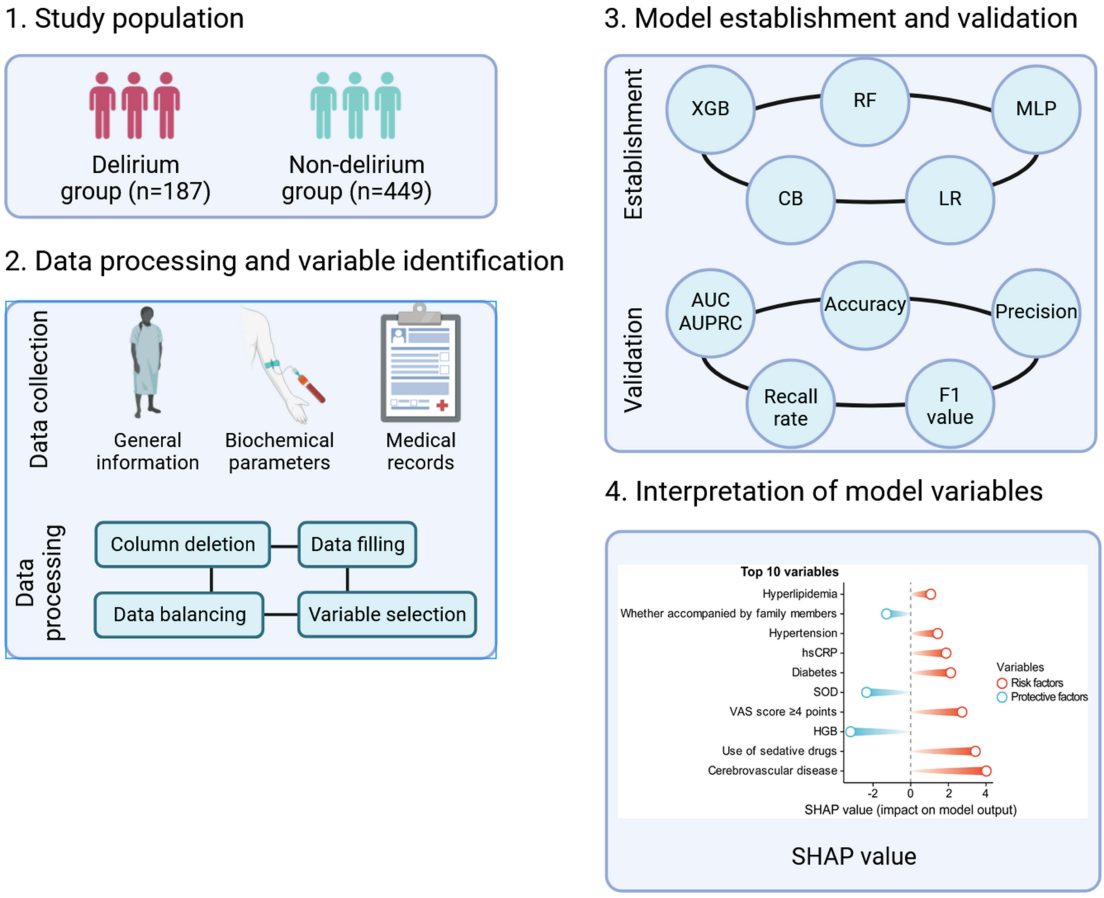
Keywords: delirium, machine learning, dementia, older, prediction, risk factors

Introduction

China's rapidly aging population is projected to reach 487 million individuals aged ≥60 years by 2039.[1] This demographic shift will result in a substantial increase in the proportion of those aged 80 and above, highlighting the rising significance of older adults in the population. Delirium, a common neuropsychiatric syndrome characterized by acute cognitive decline, inattention, and altered consciousness, poses significant risks to older adults, particularly those with pre-existing dementia.[2–4] While delirium superimposed on dementia (DSD) represents a distinct clinical phenotype, the

Received: 2 February 2025
Accepted: 12 May 2025
Published: 15 May 2025

Graphical Abstract



broader issue of delirium in older patients with dementia remains understudied, with limited predictive tools for early intervention.

Delirium is a major contributor to adverse outcomes, including prolonged hospitalization, functional decline, and mortality.[5] However, its diagnosis remains challenging due to overlapping symptoms with dementia and underrecognition in clinical settings.[6–9] The ICD-code for the diagnosis of delirium is ICD-10 F05, which is commonly used in clinical practice. However, a study by Chuen et al[5] reported that the ICD-10 codes for delirium (eg, F05) exhibit a sensitivity of 46.3% and a specificity of 99.6%. This finding underscores the limitation of ICD-10 coding in fully capturing delirium, especially in populations with pre-existing cognitive impairment, where symptoms may be overlooked or misattributed to dementia. This diagnostic gap highlights the urgency to develop accurate prediction models that can identify high-risk patients before delirium onset.

Previous studies have identified risk factors for DSD, such as cerebrovascular disease, sedative use, and inflammation markers.[10] However, these investigations were limited by small sample sizes, retrospective designs, and narrow variable selection (eg, excluding socioeconomic factors). Additionally, most studies focused exclusively on DSD rather than general delirium in dementia patients, restricting their clinical applicability. To address these gaps, this study aimed to: 1) develop a machine learning-based prediction model for delirium risk in older dementia patients; 2) identify key risk factors using comprehensive clinical data; and 3) validate the model's performance across multiple algorithms.

We leveraged a large, prospectively collected dataset of 636 hospitalized dementia patients to construct and compare five machine learning models (Extreme Gradient Boosting [XGB], Random Forest [RF], Multilayer Perceptron [MLP],

Categorical Boosting [CB], and Logistic Regression [LR]). By incorporating demographic, biochemical, and clinical variables, we sought to improve predictive accuracy and identify modifiable risk factors. This study advances previous work by using a robust prospective design, diverse algorithm comparison, and interpretability tools (SHapley Additive exPlanations [SHAP]) to enhance clinical translation.

Methods

Study Population and Data Sources

This prospective study aimed to develop a machine learning-based risk prediction model for delirium in older patients with dementia. During the study period, a total of 671 patients were initially screened for eligibility. Among them, 35 patients were excluded according to the criteria: 15 cases transferred to other hospitals (primary reasons: need for specialized neurological care, n=8; family request due to personal reasons, n=7), 9 cases died during hospitalization (primary causes: advanced dementia-related complications, n=5; acute cardiovascular events, n=4), and 11 cases had ≥90% missing data in key variables. Finally, a total of 636 patients aged 65 years or older, diagnosed with dementia (The Tenth Revision of International Classification of Diseases, ICD-10 code G30), were recruited from the International Department of China-Japan Friendship Hospital between January 2020 and December 2023. All participants were hospitalized for the first time due to dementia. Comprehensive clinical data were collected for each patient, including demographic information (age, gender, body mass index [BMI], body temperature), biochemical parameters (C-reactive protein [CRP], SOD), comorbidities (hypertension, diabetes, hyperlipidemia, cardiovascular disease, cerebrovascular disease, renal insufficiency, anemia, and hypoalbuminemia), smoking and alcohol consumption history, presence of accompanying family members, surgical history during hospitalization, Visual Analogue Scale (VAS) scores, use of sedative medications, and placement of drainage tubes. For laboratory data, we used the results of the first tests conducted after the dementia diagnosis. To construct an accurate risk prediction model, we specifically focused on delirium onset during hospitalization, recording whether delirium occurred. The study was approved by the ethics committee of China-Japan Friendship Hospital and conducted in accordance with the principles of the Declaration of Helsinki. As the study involves elderly dementia patients, informed consent was obtained from their legal guardians or family members before participation.

Data Preparation

To mitigate potential bias, all variables were blinded. Categorical variables were encoded as 0 and 1, while continuous variables were standardized using the Z-score method. Laboratory results were categorized as 1, 2, or 3, corresponding to values below the normal range, within the normal range, and above the normal range, respectively. Variables with missing data exceeding 90%, those with a single dominant value accounting for over 90%, or those with a coefficient of variation below 0.1 were excluded. Missing values were then imputed using a random forest (RF) algorithm, and Lasso regression was applied for variable selection.

Algorithms for Machine Learning

To develop the optimal prediction model, we employed five prominent machine learning algorithms: XGB, RF, Multilayer Perceptron (MLP), Categorical Boosting (CB), and Logistic Regression (LR). These algorithms were selected for their predictive capabilities in model development.

XGB

XGB is an ensemble classification method that leverages regression trees, designed to offer both fast training and superior predictive accuracy.[11] It enhances model performance by utilizing a gradient boosting technique, where weak learners are combined iteratively to improve predictions. XGB adjusts residuals from previous iterations to minimize errors, optimizing the model's overall precision. Additionally, it incorporates an internal regularization mechanism, which helps prevent overfitting, ensuring the model's generalizability. This combination of regularization and iterative optimization contributes to XGB's robustness and effectiveness in handling complex datasets.

RF

RF is a powerful ensemble learning technique that constructs a large collection of decision trees, often ranging from hundreds to thousands, to solve classification and regression problems.[12] Each tree in the forest is built using a random subset of the data and features, which promotes diversity among the trees and enhances the model's predictive power. The final prediction is made by taking the majority vote (in classification tasks) or averaging the outputs (in regression tasks) of all individual trees. One of the key strengths of the RF algorithm is its robustness and stability, which makes it highly effective in handling small amounts of noise and outliers in the data. Due to its ensemble nature, RF is less prone to overfitting compared to single decision trees, especially when the number of trees is sufficiently large.

MLP

The MLP is a type of machine learning algorithm that utilizes feedforward neural networks, which function similarly to the way human neurons process information.[13] In its architecture, the MLP is composed of several layers, including an input layer that receives the data, one or more hidden layers that process the information, and an output layer that generates the final result.[14] Each layer contains multiple neurons that work in unison to learn complex patterns from the input data. The layers are connected through weighted links, with the network adjusting these weights during the training process to improve accuracy.

CB

CB is a machine learning classification technique that utilizes oblivious trees, a specialized tree structure designed to handle classification tasks in an efficient and interpretable way.[15] Unlike traditional decision trees, oblivious trees ensure that the same feature is used for all branches at each level of the tree, thereby reducing the model's complexity. This approach effectively mitigates common challenges in machine learning models, such as gradient bias and prediction shift, which can contribute to overfitting. By addressing these issues, CB enhances the model's overall performance, leading to improved accuracy and stronger generalization capabilities.

LR

An LR model is a statistical tool used to examine the relationship between one or more independent variables and a binary dependent variable, often referred to as the outcome or response.[16] It is particularly useful for predicting the probability of a binary event occurring, such as the presence or absence of a disease, success or failure, or other dichotomous outcomes. LR models offer a straightforward interpretation, where the estimated coefficients represent the impact of each predictor on the odds of the outcome occurring.[17] Due to its simplicity and robustness, LR has been extensively applied across various medical research domains, including epidemiology, clinical studies, and public health, to model and understand complex relationships between risk factors and health outcomes.

Algorithm Comparison

LR offers the highest interpretability due to its linear coefficients, making it particularly well-suited for scenarios that require transparent risk factor analysis. RF and XGB provide a balance of interpretability through feature importance metrics (such as SHAP values) while maintaining strong generalization capabilities. CB exhibits robustness against gradient bias, though it may require more tuning for optimal performance. MLP, a deep learning model, has the lowest interpretability but excels in capturing complex non-linear relationships. However, it demands larger datasets and higher computational resources to prevent overfitting.

In terms of computational efficiency, LR and CB are the most cost-effective, followed by RF. XGB and MLP are more computationally intensive, particularly for large-scale datasets. Regarding sample size requirements, MLP and XGB perform best with datasets larger than 500 samples, while LR and RF remain stable even with smaller cohorts. These characteristics make LR ideal for initial hypothesis testing, XGB/RF preferable for complex clinical prediction tasks, and MLP a viable option when rich data and sufficient computational resources are available.

For readers interested in learning more about machine learning tools, I recommend visiting the website https://scikit-learn.org/stable/, which provides valuable resources and documentation on various algorithms and their applications.

Model Establishment

The prediction models were constructed using XGB, MLP, RF, LR, and CB algorithms. The data were first split into training and test sets using an 8:2 ratio. All preprocessing steps, including data cleaning, normalization, and missing data estimation, were performed separately on the training and test sets to prevent data leakage and ensure an independent evaluation of model performance. The training set was used to build the classification models, and the test set was used for performance evaluation. To address the class imbalance in the training set, the Borderline Synthetic Minority Oversampling Technique (SMOTE) was employed.[18] This improved version of SMOTE generates synthetic samples from borderline instances, enhancing the class distribution. The models were trained on the original data, and their predictive accuracy was evaluated on the test set.

Model Evaluation

To assess the predictive capability of the machine learning models, we computed several key performance metrics, including the area under the receiver operating characteristic curve (AUC), accuracy, precision, recall, and F1 score. Given the importance of precision and recall in clinical contexts, we also evaluated the area under the precision-recall curve (AUPRC). The calibration of the model was assessed using the Brier score and a calibration plot, with favorable calibration defined by a Brier score ≤ 0.25.[19] To understand the influence of each variable on the model, SHapley Additive exPlanations (SHAP) values were utilized.

Statistical Analysis

Categorical variables, presented as frequencies and percentages, were analyzed using chi-square tests. For continuous data, descriptive statistics included mean ± standard deviation or median with interquartile ranges (Q1–Q3). The *t*-test or Mann–Whitney *U*-test was employed to evaluate statistical differences between groups. A p-value below 0.05 indicated statistical significance. Data processing and statistical evaluations were carried out using SPSS software (Version 25, IBM SPSS Statistics, IBM Corp., Armonk, NY, USA). Additionally, statistical modeling was conducted with the stats and sklearn libraries in Python (Version 3.8).

Results

Study Population

A total of 636 cognitively impaired patients aged over 65 years, admitted to our hospital's comprehensive ward, were included in the study. Among them, 357 (56.1%) were men and 279 (43.9%) were women, with a mean age of 78.4 ± 6.5 years. Patients were categorized into two groups: 187 in the delirium group and 449 in the non-delirium group, yielding a delirium incidence rate of 29.4%. Baseline characteristics are summarized in Table 1.

Data Processing and Variable Identification

A total of 115 variables were initially collected, as detailed in Supplementary Table 1. These variables included three general information parameters (X1–X3), eight comorbidities (X4–X11), five medications (X12–X16), and 99 laboratory test results (X17–X115). Variables with missing data exceeding 90%, those with a single value comprising over 90% of the observations, and those with a coefficient of variation below 0.1 were excluded. This left 84 variables for further analysis (Supplementary Table 2). Lasso regression was then applied for variable selection, resulting in the final inclusion of 52 variables in the model (Supplementary Table 2).

Model Establishment and Validation

To predict delirium in older dementia patients, machine learning models, including XGB, MLP, RF, LR, and CB, were built using 52 features from the training set. Their predictive performance was assessed on a separate test set, with metrics such as AUC, accuracy, precision, recall, and F1 score reported in Table 2. Among the models, XGB outperformed the others, achieving the highest AUC (0.967), accuracy (0.907), and F1 score (0.918). To improve clinical feasibility, a streamlined version of the XGB model was developed, using the top 10 most important variables based on

Table 1 Baseline Characteristics of Delirium vs Non-Delirium Patients

Characteristics	Delirium Group (n=187)	Non-Delirium Group (n=449)	Z/χ^2	*P* value
Age, years	79 (68, 85)	78 (67, 84)	1.482	0.336
Male	97 (51.9%)	260 (57.9%)	1.953	0.162
BMI<18.5 kg/m^2	106 (56.7%)	273 (60.8%)	0.929	0.335
Hypertension	115 (61.5%)	218 (48.6%)	8.869	0.003
Diabetes	123 (65.8%)	197 (43.9%)	25.328	<0.001
Hyperlipidemia	98 (52.4%)	144 (32.1%)	23.159	<0.001
Cardiovascular disease	64 (34.2%)	126 (28.1%)	2.393	0.122
Cerebrovascular disease	57 (30.5%)	63 (14.0%)	23.337	<0.001
Renal insufficiency	26 (13.9%)	44 (9.8%)	2.270	0.132
Anemia	106 (56.7%)	238 (53.0%)	0.719	0.396
Hypoproteinemia	73 (39.0%)	109 (24.3%)	14.082	<0.001
Smoking	85 (45.5%)	221 (49.2%)	0.750	0.386
Drinking	91 (48.7%)	236 (52.6%)	0.803	0.370
Whether accompanied by family members	32 (17.1%)	284 (63.3%)	112.422	<0.001
Surgical history during hospitalization	55 (29.4%)	149 (33.2%)	0.863	0.353
VAS score ≥4 points	124 (66.3%)	136 (30.3%)	70.874	<0.001
Use of sedative drugs	118 (63.1%)	104 (23.2%)	92.680	<0.001
Whether a drainage tube was placed	77 (41.2%)	183 (40.8%)	0.010	0.922
Body temperature ≥37.3°C	34 (18.2%)	76 (16.9%)	0.145	0.703
GGT, U/L	25.3 (15.2, 43.4)	24.7 (15.9, 44.7)	0.394	0.118
ALB, g/L	37.8 (34.1, 41.5)	40.7 (36.5, 44.0)	−4.692	0.014
ALT, U/L	21.9 (14.5, 30.7)	18.1 (12.7, 27.3)	3.493	0.036
AST, U/L	26.9 (21.8, 34.4)	26.3 (21.4, 34.6)	0.280	0.495
TP, g/L	67.3 (61.3, 72.0)	68.8 (64.1, 73.8)	−2.874	0.062
TG, mmol/L	1.38 (0.97, 1.99)	1.20 (0.92, 1.78)	4.739	0.008
eGFR, mL/min	78.5 (55.1, 88.5)	85.7 (68.4, 94.1)	−6.724	<0.001
Urea, mmol/L	6.4 (5.0, 9.1)	5.9 (4.9, 7.6)	3.871	0.024
Creatinine, μmol/L	78.2 (63.1, 94.6)	66.4 (56.1, 84.0)	5.948	<0.001
Uric acid, μmol/L	343 (271.0, 429.4)	338 (261.6, 402.5)	2.016	0.082
Total bile acid, μmol/L	5.1 (3.4, 8.2)	4.9 (3.5, 8.0)	0.874	0.572
TC, mmol/L	3.7 (2.9, 4.3)	3.8 (3.0, 4.6)	−0.144	0.650
TBIL, μmol/L	13.4 (10.1, 17.4)	13.5 (10.3, 18.0)	−0.085	0.893
IBIL, μmol/L	8.9 (6.3, 12.6)	8.5 (5.1, 11.7)	2.145	0.102
WBC, *10^9/L	6.4 (5.0, 8.3)	6.2 (4.9, 8.2)	0.938	0.317
RBC, *10^{12}/L	4.2 (3.6, 4.5)	4.5 (3.8, 4.9)	−3.283	0.042
Lymphocyte, *10^9/L	1.3 (0.8, 1.6)	1.3 (0.9, 1.8)	−0.021	0.936
HGB, g/L	97.3 (89.2, 112.6)	106.5 (97.2, 121.3)	−6.749	<0.001
HCT, %	38.7 (33.4, 41.6)	39.2 (35.4, 42.7)	−0.843	0.568
Cholinesterase, KU/L	6.4 (5.0, 8.1)	7.3 (6.2, 8.8)	−5.382	<0.001
ALP, U/L	80.3 (62.5, 102.7)	82.7 (63.5, 117.4)	−3.845	0.021
HDLC, mmol/L	1.2 (0.9, 1.4)	1.3 (1.0, 1.5)	−2.344	0.097
LDLC, mmol/L	2.1 (1.5, 2.7)	2.0 (1.3, 2.6)	0.484	0.782
Globulin, g/L	27.2 (24.4, 31.1)	27.7 (24.5, 31.1)	−1.038	0.226
TSH, μ/L	2.0 (1.2, 3.3)	1.8 (1.1, 2.7)	2.948	0.052
Creatine kinase, U/L	79.3 (54.2, 119.6)	90.6 (63.2, 138.8)	−8.374	<0.001
hsCRP, mg/L	4.1 (1.1, 27.8)	1.6 (0.5, 11.1)	11.932	<0.001
SOD, U/mL	102.3 (54.4, 143.8)	138.4 (84.3, 179.5)	−8.484	<0.001

Note: Data are expressed as counts (%) or as the median (Q1, Q3).

Abbreviations: GGT, γ-glutamyl transpeptidase; ALB, albumin; ALT, alanine aminotransferase; TG, triglycerides; eGFR, estimated glomerular filtration rate; AST, aspartate aminotransferase; TC, total cholesterol; TBIL, total bilirubin; TP, total protein; WBC, white blood cells; RBC, red blood cells; HCT, hematocrit; HGB: hemoglobin; LDLC, low-density lipoprotein cholesterol; HDLC, high-density lipoprotein cholesterol; IBIL, indirect bilirubin; ALP, alkaline phosphatase; TSH, thyroid stimulating hormone; hsCRP, high-sensitivity C-reactive protein.

Table 2 Performance of Machine Learning Models on the Test Set

Models	AUC	Accuracy	Precision	Recall	F1 value	Brier Score
MLP	0.872	0.791	0.790	0.874	0.887	0.36
LR	0.802	0.796	0.813	0.823	0.854	0.22
RF	0.959	0.904	0.963	0.748	0.906	0.09
CB	0.911	0.858	0.874	0.816	0.874	0.14
XGB	0.967	0.907	0.942	0.815	0.918	0.07
Compact XGB	0.930	0.870	0.908	0.872	0.892	0.11

Abbreviations: AUC, area under the receiver operating characteristic curve; MLP, multilayer perceptron; LR, logistic regression; RF, random forest; CB, categorical boosting; XGB, extreme gradient boosting.

their mean absolute SHAP values. The performance of this compact XGB model yielded an AUC of 0.930, accuracy of 0.870, precision of 0.908, recall of 0.872, and F1 score of 0.892 (Table 2). The corresponding ROC curves for these models are shown in Figure 1A.

For the original XGB model, trained on the test set, the AUC, accuracy, precision, recall, and F1 score were 0.905, 0.862, 0.873, 0.927, and 0.608, respectively (see Supplementary Table 3). Additional ROC curves for these models can be found in Supplementary Figure 1A.

Recall represents the proportion of delirium cases correctly identified, whereas precision reflects the accuracy of predicting delirium among the cases identified. Both metrics are critical in clinical practice for predicting delirium in patients and assisting healthcare providers in making informed decisions. To evaluate the performance of the machine learning models, we utilized precision-recall curves (Figure 1B). The XGB model achieved the highest AUPRC of 0.989, followed by the compact XGB model, which yielded an AUPRC of 0.966. The precision-recall curves for models trained on the original dataset are shown in Supplementary Figure 1B. In addition, calibration curves were employed to assess the models' calibration, with both the XGB and compact XGB models demonstrating strong calibration performance (Figure 2). The Brier score, which quantifies both calibration and discrimination, was 0.07 for the XGB model and 0.11 for the compact XGB model (Table 2).

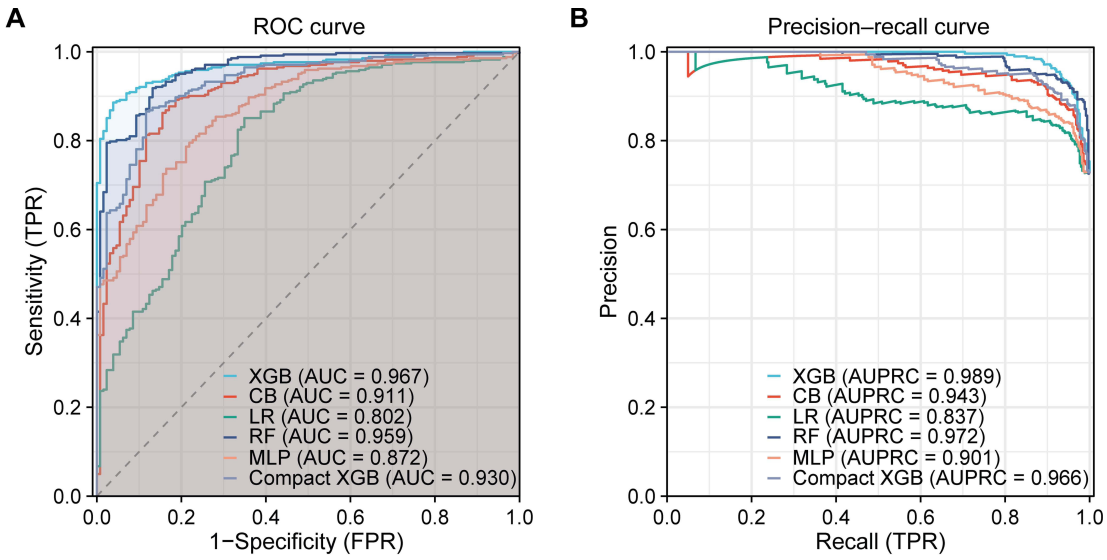


Figure 1 Model establishment. (**A**) Receiver operating characteristic (ROC) curves and (**B**) precision-recall curves for the six machine learning models on the test set.
Abbreviations: AUC, area under the receiver operating characteristic curve; MLP, multilayer perceptron; LR, logistic regression; RF, random forest; CB, categorical boosting; SHAP, SHapley Additive exPlanations; XGB, extreme gradient boosting.

https://doi.org/10.2147/CIA.S519366

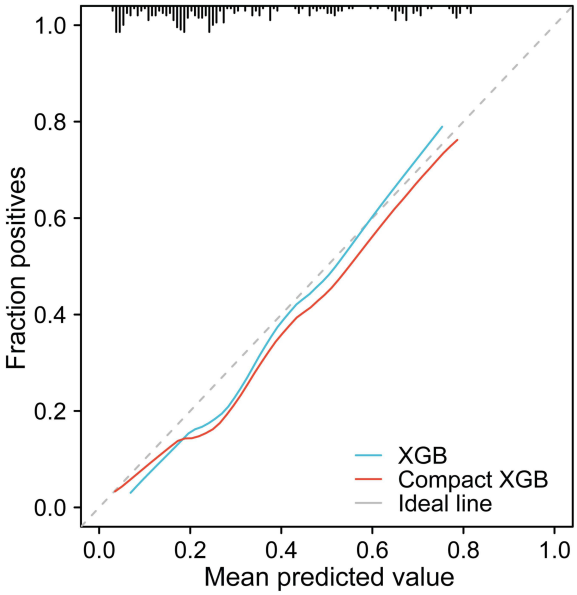


Figure 2 Calibration plot for both the XGB and compact XGB models. XGB: extreme gradient boosting.

Interpretation of Model Variables

Based on SHAP value analysis, the 10 most influential variables in the XGB model were identified, including cerebrovascular disease, use of sedative drugs, hemoglobin, VAS score ≥4 points, SOD, diabetes, hsCRP, hypertension, whether accompanied by family members, and hyperlipidemia. These factors were subsequently employed to develop a streamlined XGB model.

Discussion

This study aimed to develop a robust machine learning-based prediction model to assess the risk of delirium in older patients with dementia, a group highly susceptible to this severe and often underrecognized condition. Our analysis demonstrated that machine learning algorithms, particularly XGB, can effectively predict the onset of delirium in this population, with the model showing excellent performance in terms of both accuracy and predictive power. We utilized XGB, RF, MLP, CB, and LR algorithms, combined with a comprehensive set of clinical data, to predict delirium risk in older dementia patients. Among these, XGB emerged as the most effective model, achieving the highest AUC, accuracy, and F1 score. While MLP exhibited a slightly higher recall than XGB, its lower precision and accuracy limited its clinical applicability.[2,20] Moreover, the XGB model excelled in calibration, achieving the highest AUPRC and the best Brier score. To enhance clinical applicability, we refined the XGB model by focusing on the top 10 most significant risk factors, resulting in a more efficient model that maintained strong predictive power.

Our findings are consistent with existing research highlighting the increased risk of delirium in hospitalized older patients, particularly those with cognitive impairment.[21] With a delirium incidence rate of 29.4% in our cohort, our study emphasizes the clinical relevance of this issue. The identification of key risk factors, including cerebrovascular disease, sedative drug use, and elevated VAS scores, aligns with prior studies that have associated these variables with a higher likelihood of delirium.[2,22] These factors provide critical insights into the pathophysiology and risk stratification of delirium, which are essential for timely intervention.

The XGB model demonstrated superior performance compared to other machine learning algorithms in this study, achieving the highest AUC of 0.967, accuracy of 0.907, and F1 score of 0.918. These results underscore the efficacy of XGB in clinical prediction tasks. By reducing the model to the top 10 most important features based on SHAP values, we developed a streamlined version that maintained high predictive performance, with an AUC of 0.930 and an F1 score of 0.892. These findings emphasize the potential of machine learning as a robust and clinically feasible tool for predicting delirium, which could be implemented in high-paced hospital environments to enhance patient outcomes.

Our study also provided valuable insights into the key risk factors for delirium in patients with dementia. Factors such as cerebrovascular disease, sedative use, and elevated hsCRP levels were identified as the most significant predictors. These findings align with existing literature, which highlights the critical roles of cerebrovascular insults and inflammation in the onset of delirium.[8,23] The identification of cerebrovascular disease, sedative use, and elevated hsCRP levels as significant predictors is biologically plausible. Cerebrovascular disease may disrupt brain perfusion and increase vulnerability to delirium by impairing neurovascular coupling and promoting neuroinflammation.[24] Sedative medications, such as benzodiazepines, enhance GABAergic neurotransmission, which can exacerbate cognitive dysfunction and increase delirium risk through sedation and altered arousal regulation.[25] Elevated hsCRP levels reflect systemic inflammation, which contributes to blood-brain barrier permeability and microglial activation, driving neuroinflammation and cognitive impairment.[26] These mechanisms underscore the multifactorial nature of delirium in dementia patients and highlight potential targets for intervention. Additionally, the use of sedative medications remains a well-established risk factor, underscoring the importance of careful medication management in this vulnerable population.[9,27] Notably, family support emerged as a significant variable, suggesting that the presence of family members may reduce the risk of delirium, potentially by fostering cognitive stimulation and providing emotional support.

One of the strengths of this study lies in the use of a large, well-characterized cohort of older dementia patients, with comprehensive clinical data collected prospectively. This methodology facilitated the identification of a broad range of potential risk factors, providing a solid foundation for the development of our prediction model. However, several limitations should be acknowledged. First, while the model demonstrated strong performance on the validation set, its generalizability to other populations—such as those with different dementia subtypes or individuals in community settings—requires further validation. Additionally, despite the robustness of the dataset, some potential risk factors, such as education level or socioeconomic status, were not included, and their omission may influence the model's performance and broader applicability. Future research should examine these variables and assess their impact on delirium risk.

In conclusion, our study highlights the feasibility and effectiveness of machine learning in predicting delirium risk in older patients with dementia. By identifying key risk factors and developing a streamlined, clinically applicable model, we provide a valuable tool for clinicians in managing and preventing delirium. Future research should focus on validating the model across diverse clinical settings, exploring additional risk factors, and further refining the prediction algorithms to enhance their clinical utility.

Data Sharing Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on request.

Ethics Approval and Consent to Participate

The study was approved by the ethics committee of China-Japan Friendship Hospital and conducted in accordance with the principles of the Declaration of Helsinki. As the study involves elderly dementia patients, informed consent was obtained from their legal guardians or family members before participation.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This work was supported by National High Level Hospital Clinical Research Funding. (2023-NHLHCRF-YYPPLC-ZR-23).

Disclosure

The authors declare no competing or financial interests.

References

1. Bugiani O. Delirium and dementia in the elderly: sometimes associated or always together? *Eur Neurol*. 2023;86(3):161–165. doi:10.1159/000530226
2. Xiao Q, Zhang S, Li C, Zhu Y. Risk factors for delirium superimposed on dementia in elderly patients in comprehensive ward. *Am J Alzheimers Dis Other Demen*. 2023;38:15333175231206023. doi:10.1177/15333175231206023
3. Cheung ECL, Leung MTY, Chen K, et al. Risk of adverse events and delirium after COVID-19 vaccination in patients living with dementia. *J Am Med Directors Assoc*. 2023;24(6):892–900.e812. doi:10.1016/j.jamda.2023.04.003
4. Sillner AY, Berish D, Mailhot T, Sweeder L, Fick DM, Kolanowski AM. Delirium superimposed on dementia in post-acute care: nurse documentation of symptoms and interventions. *Geriatric Nurs*. 2023;49:122–126. doi:10.1016/j.gerinurse.2022.11.015
5. Chuen VL, Chan ACH, Ma J, Alibhai SMH, Chau V. Assessing the accuracy of international classification of diseases (ICD) coding for delirium. *J Appl Gerontol*. 2022;41(5):1485–1490. doi:10.1177/07334648211067526
6. Hshieh TT, Inouye SK, Oh ES. Delirium in the elderly. *Clin Geriatric Med*. 2020;36(2):183–199. doi:10.1016/j.cger.2019.11.001
7. Ryan SL, Kimchi EY. Evaluation and management of delirium. *Seminars Neurol*. 2021;41(5):572–587. doi:10.1055/s-0041-1733791
8. Shahabi Raberi V, Solati Kooshk Qazi M, Zolfi Gol A, GhorbaniNia R, Kahourian O, Faramarz Zadeh R. Postoperative delirium and dementia in patients undergoing cardiac surgery: a review of randomized controlled trials. *Galen Med J*. 2023;12:1–9.
9. Cai Y, Wang J, Wang X, Yuan Q, Xu Y, Xu P. Causal relationship between dementia and delirium: insights from a bidirectional two-sample Mendelian randomization analysis. *J Affective Disorders*. 2024;349:69–76. doi:10.1016/j.jad.2024.01.064
10. Boltz M, Kuzmik A, Resnick B. Delirium and behavioral symptoms in persons with dementia at hospital admission: mediating factors. *Alzheimer Dis Associated Disord*. 2023;37(2):120–127. doi:10.1097/WAD.0000000000000552
11. Yuan G, Lv B, Du X. Prediction model for missed abortion of patients treated with IVF-ET based on XGBoost: a retrospective study. *Peer J*. 2023;11:e14762. doi:10.7717/peerj.14762
12. Chen X, Ishwaran H. Random forests for genomic data analysis. *Genomics*. 2012;99(6):323–329. doi:10.1016/j.ygeno.2012.04.003
13. Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H. State-of-the-art in artificial neural network applications: a survey. *Heliyon*. 2018;4(11):e00938. doi:10.1016/j.heliyon.2018.e00938
14. Mohammadi F, Dehbozorgi L, Akbari-Hasanjani HR. Evaluation of effective features in the diagnosis of Covid-19 infection from routine blood tests with multilayer perceptron neural network: a cross-sectional study. *Health Sci Rep*. 2023;6(1):e1048. doi:10.1002/hsr2.1048
15. Lyu Y, Wu H-M, Yan H-X. Classification of coronary artery disease using radial artery pulse wave analysis via machine learning. *BMC Med Inf Decis Making*. 2024;24(1):256. doi:10.1186/s12911-024-02666-1
16. Zabor EC, Reddy CA, Tendulkar RD, Patil S. Logistic regression in clinical studies. *Int J Radiat Oncol Biol Phys*. 2022;112(2):271–277. doi:10.1016/j.ijrobp.2021.08.007
17. Abe D, Inaji M, Hase T. A prehospital triage system to detect traumatic intracranial hemorrhage using machine learning algorithms. *JAMA Network Open*. 2022;5(6):e2216393. doi:10.1001/jamanetworkopen.2022.16393
18. Bunkhumpornpat C, Boonchieng E, Chouvatut V, Lipsky D. FLEX-SMOTE: synthetic over-sampling technique that flexibly adjusts to different minority class distributions. *Patterns*. 2024;5(11):101073. doi:10.1016/j.patter.2024.101073
19. Lin J, Yin M, Liu L, et al. The development of a prediction model based on random survival forest for the postoperative prognosis of pancreatic cancer: a SEER-based study. *Cancers*. 2022;14(19).
20. Korkatti-Puoskari N, Tiihonen M, Caballero-Mora MA, et al. Therapeutic dilemma's: antipsychotics use for neuropsychiatric symptoms of dementia, delirium and insomnia and risk of falling in older adults, a clinical review. *Eur Geriatric Med*. 2023;14(4):709–720. doi:10.1007/s41999-023-00837-3
21. Planche V. Delirium, blood biomarkers for neurodegeneration, and dementia. *Lancet Healthy Longevity*. 2023;4(8):e362–e363. doi:10.1016/S2666-7568(23)00104-6
22. Jaqua EE, Nguyen VTN, Chin E. Delirium in older persons: prevention, evaluation, and management. *Am Family Phys*. 2023;108(3):278–287.
23. Umoh ME, Fitzgerald D, Vasunilashorn SM, Oh ES, Fong TG. The relationship between delirium and dementia. *Seminars Neurol*. 2024;44(6):732–751. doi:10.1055/s-0044-1791543
24. Fang X, Border JJ, Zhang H. A soluble epoxide hydrolase inhibitor improves cerebrovascular dysfunction, neuroinflammation, amyloid burden, and cognitive impairments in the hAPP/PS1 TgF344-AD rat model of alzheimer's disease. *Int J Mol Sci*. 2025;26(6):2433. doi:10.3390/ijms26062433
25. Zhong S, Zhao Z, Xie W. GABAergic interneuron and neurotransmission are mTOR-dependently disturbed in experimental focal cortical dysplasia. *Mol Neurobiol*. 2021;58(1):156–169. doi:10.1007/s12035-020-02086-y
26. Xie W, Kong C, Luo W, Zheng J, Zhou Y. C-reactive protein and cognitive impairment: a bidirectional Mendelian randomization study. *Arch Gerontol Geriatr*. 2024;121:105359. doi:10.1016/j.archger.2024.105359
27. Gordon EH, Ward DD, Xiong H, Berkovsky S, Hubbard RE. Delirium and incident dementia in hospital patients in New South Wales, Australia: retrospective cohort study. *BMJ*. 2024;384:e077634. doi:10.1136/bmj-2023-077634