Infection and Drug Resistance

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

A Machine Learning Model Based on CT Imaging Metrics and Clinical Features to Predict the Risk of Hospital-Acquired Pneumonia After Traumatic Brain Injury

Shaojie Li[1,*], Qiangqiang Feng[1,*], Jiayin Wang[1], Baofang Wu[1], Weizhi Qiu[1], Yiming Zhuang[2], Yong Wang[3], Hongzhi Gao[1]

[1]Department of Neurosurgery, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, Fujian, 362000, People's Republic of China; [2]Internal Medicine, Quanzhou Quangang District Hillside Street Community Health Service Center, Quanzhou, Fujian, 362000, People's Republic of China; [3]Child and Adolescent Psychiatry, The Third Hospital of Quanzhou, Quanzhou, Fujian, 362000, People's Republic of China

*These authors contributed equally to this work

Correspondence: Yong Wang; Hongzhi Gao, Email 120432246@qq.com; gaohongzhi@fjmu.edu.cn

Objective: To develop a validated machine learning (ML) algorithm for predicting the risk of hospital-acquired pneumonia (HAP) in patients with traumatic brain injury (TBI).

Materials and Methods: We employed the Least Absolute Shrinkage and Selection Operator (LASSO) to identify critical features related to pneumonia. Five ML models—Logistic Regression (LR), Extreme Gradient Boosting (XGB), Random Forest (RF), Naive Bayes Classifier (NB), and Support Vector Machine (SVC)—were developed and assessed using the training and validation datasets. The optimal model was selected based on its performance metrics and used to create a dynamic web-based nomogram.

Results: In a cohort of 858 TBI patients, the HAP incidence was 41.02%. LR was determined to be the optimal model with superior performance metrics including AUC, accuracy, and F1-score. Key predictive factors included Age, Glasgow Coma Score, Rotterdam Score, D-dimer, and the Systemic Immune Response to Inflammation Index (SIRI). The nomogram developed based on these predictors demonstrated high predictive accuracy, with AUCs of 0.818 and 0.819 for the training and validation datasets, respectively. Decision curve analysis (DCA) and calibration curves validated the model's clinical utility and accuracy.

Conclusion: We successfully developed and validated a high-performance ML algorithm to assess the risk of HAP in TBI patients. The dynamic nomogram provides a practical tool for real-time risk assessment, potentially improving clinical outcomes by aiding in early intervention and personalized patient management.

Keywords: traumatic brain injury, machine learning, hospital-acquired pneumonia, dynamic nomogram, imaging metrics

Introduction

With the development of society and the rise of road traffic, the incidence of accidental injuries, especially traumatic brain injury (TBI), caused by frequent traffic accidents has been increasing year by year, thus substantially increasing the risk of death.[1] According to the US authorities, 64 to 74 million people worldwide suffer from TBI every year, of which about 55.9 million people are affected by mild TBI, while severe TBI affects about 5.48 million people.[2] For every 100,000 people, there are 262 cases of death or disability due to TBI-related causes. Especially in patients with severe TBI, further serious complications such as cerebral contusions, axonal injury, and delayed hemorrhage may develop, all of which can severely affect the cognitive and motor functions of patients.[3] It has been demonstrated that there is a complex and fine-grained interaction between the brain and other body systems, such as the central nervous system and the respiratory system, in a manner known as the "lung-brain axis".[4] On the one hand, the brain sends signals through the

Received: 28 June 2024
Accepted: 30 August 2024
Published: 5 September 2024

autonomic nervous system to directly regulate lung function, and the state of gas exchange in the lungs is markedly affected when in pathological conditions,[5] on the other hand, dysregulation of the lung microbiota significantly affects the CNS immune response through the blood-brain barrier.[6] One of the primary causes of death for TBI patients is severe pneumonia, which not only has a detrimental impact on the patients' prognosis but also raises the financial burden on families.[7] Therefore, therapeutic strategies for TBI should not only focus on cephalic conditions but also include interventions for complications such as pneumonia.

Earlier, hospital-acquired pneumonia (HAP) was a broad term covering ventilator-associated pneumonia (VAP) occurring after the establishment of an artificial airway and initiation of mechanical ventilation.[8] However, the recently updated HAP/VAP guidelines in the United States differentiate between HAP and VAP, pointing out that they differ significantly in terms of pathogenic microorganisms, treatment experience, and preventive strategies, and treating them as two separate groups.[9] Therefore, this study focused on patients with pneumonia unrelated to intubation and mechanical ventilation. For patients at high risk of HAP, HAP/VAP guidelines recommend immediate initiation of empirical antibiotic therapy.[10] However, the use of prophylactic antibiotics does not appear to be effective in preventing pneumonia in patients who have suffered TBI and stroke.[11–13] Nonetheless, certain studies have noted that prophylactic antibiotic use in specific groups can prevent pneumonia.[14,15] This suggests that preventive and therapeutic approaches for TBI-associated pneumonia require individualized strategies. Therefore, the development of early prediction models for TBI-associated pneumonia is particularly important to help healthcare professionals identify high-risk patients in time to implement the necessary preventive measures, rapid diagnosis, and precise treatment.

Materials and Methods

Data Source and Extraction

This was a retrospective study. The study protocol was approved by the Ethics Committee of the Second Affiliated Hospital of Fujian Medical University (2023–287). Since the study was retrospective in nature, informed permission was not required. Clinical data were collected electronically through the hospital's electronic medical record system from 1 January 2018 to 1 June 2023 at the Second Affiliated Hospital of Fujian Medical University. The Ethics Committee was in charge of ensuring that the patients' private information was properly protected. The inclusion criteria for patients with TBI in this study were as follows: (1) patients with definite TBI by cranial CT; (2) patients aged 18 years and above. The exclusion criteria were as follows: (1) Incomplete or missing key clinical data (CT was not perfected within 48h of admission); (2) Patients with tracheotomy and use of ventilator; (3) Patients discharged from the hospital within 72h (4) Patients with multiple injuries of the whole body or accompanied by serious life-threatening organ or system diseases. (5) Patients diagnosed with pneumonia before admission. After admission, each patient was examined and treated by a professional neurologist according to the TBI treatment guidelines.

Definition of HAP

HAP was defined as pneumonia occurring in the hospital 48 hours after admission.[9] HAP was diagnosed by the presence of typical chest X-ray or computed tomography (CT) support, plus two or more of the following three clinical signs: (1) fever with a temperature of >38°C; (2) new onset of purulent sputum, change in sputum nature, increase in respiratory secretions, or increase in the need for suctioning; and (3) a peripheral blood leukocyte count of >10 × 10 9/L or < 4 × 109/L.[16] In this study, HAP was used as an outcome variable.

Covariates

Data on the patient's general status upon admission, neuroimaging features, and laboratory tests were gathered from the hospital's electronic medical record system at the time of initial admission. The general conditions on admission included the patient's sex, age, temperature on admission, heart rate, respiration, systolic blood pressure, diastolic blood pressure, cause of injury, time of onset, pupillary reflex, GCS, history of hypertension, diabetes mellitus, and coronary artery disease. Laboratory investigations on admission included leukocytes, centrocytes, lymphocytes, monocytes, hemoglobin, platelets, blood glucose, uric acid, creatinine, urea nitrogen, potassium, sodium, calcium, osmolality, APTT, PT, INR,

Fib, TT, and D-dimer, with the addition of novel combined inflammatory markers SII, SIRI, PLR, NLR, and LMR.In addition, a record of the Neuroimaging features of TBI, including frontal lobe injury, parietal lobe injury, temporal lobe injury, occipital lobe injury, brainstem injury, cerebellar injury, epidural hemorrhage, subdural hemorrhage, craniovertebral fracture, skull base fracture, type of brain herniation, and Rotterdam scores (including cerebral pool compression or disappearance, subarachnoid hemorrhage, ventricular hemorrhage, midline shift of >5 mm, and extradural space-occupying lesions).

Statistical Analysis

In this study, the data collected from TBI patients were analyzed using R version 4.3.1, Python 3.12.0 and SPSS version 25.0 for subsequent statistical analysis. For continuous variables, data were expressed as mean ± standard deviation (X ± SD). For categorical variables, frequencies or percentages (%) were used. Different data metrics were compared between the two groups (training and validation datasets). An independent samples *t*-test was utilized to compare groups for continuously distributed variables that were normally distributed, whereas non-parametric tests were employed for variables that were not normally distributed. For comparing categorical variables between groups, the chi-square test was employed. The significance level was set at $\alpha=0.05$, and $P<0.05$ was considered statistically significant. The LASSO was used to select significant features associated with pneumonia. Finally, seven features (glucose, type of brain hernia, age, GCS, Rotterdam, SIRI, D-dimer) were included in the established model. To obtain the highest predictive performance, five models were built, including a LR, a RF, a XGB classifier, a SVM classifier, and a NB classifier. The optimal model was evaluated based on the key performance metrics (Accuracy, Matthews correlation coefficient (MCC), recall, and F1 score) calculated from the confusion matrix. To further explain the optimal model, this study used multifactorial logistic regression to identify independent risk factors for HAP. In this model, the variance inflation factor (VIF) was used to estimate multicollinearity. A clinical prediction model was constructed based on the relative weights of the risk factors, and a dynamic nomogram was plotted. The discriminative power of the model was assessed using subject work characteristics (ROC) curves and AUC was calculated. Furthermore, calibration curves were constructed and the clinical prediction model was internally validated using the Bootstrap resampling approach with a repetition number of B=1000. Lastly, the Decision Clinical Curve (DCA) analysis was used to evaluate the clinical validity of the predictive models.

Results

Baseline Characteristics

This study comprised 998 participants who satisfied the primary diagnostic. After reviewing the medical records, we excluded 135 patients who were younger than 18 years of age and 5 patients with missing data. Finally, 858 eligible participants were included in this study for analysis. The flowchart of patient recruitment is shown below (Figure 1). The 858 included TBI patients were randomly split into a training group (n = 686) and a validation dataset (n = 172) in an 8:2 ratio to avoid overfitting in the impact factor analysis. Comparing the data between the two datasets (Table 1), we found no statistically significant differences in all covariates included ($P\geq0.05$), suggesting that the partitioning of our dataset was reasonable and comparable. Patients were categorized into HAP and Non-HAP groups based on whether pneumonia occurred after 2 days of admission (Table 1). Modelling was performed from the data collected in the validation dataset and the predictive models were validated in the validation dataset. Using 10-fold cross-validation, the optimal parameter (lambda) for the LASSO model was determined (Figure 2A). Vertical dashed lines were plotted at the optimal values using the minimum criterion and 1 SE of the minimum criterion.

A vertical line was plotted at the selected values using 10-fold cross-validation. where the optimal lambda yields 7 feature variables with non-zero coefficients (Figure 2B). We selected 7 non-zero feature variables in the LASSO regression results (Table 2), including age, type of brain herniation, admission GCS, Rotterdam score (Figure 3A–F), glucose, D-dimer, and SIRI.

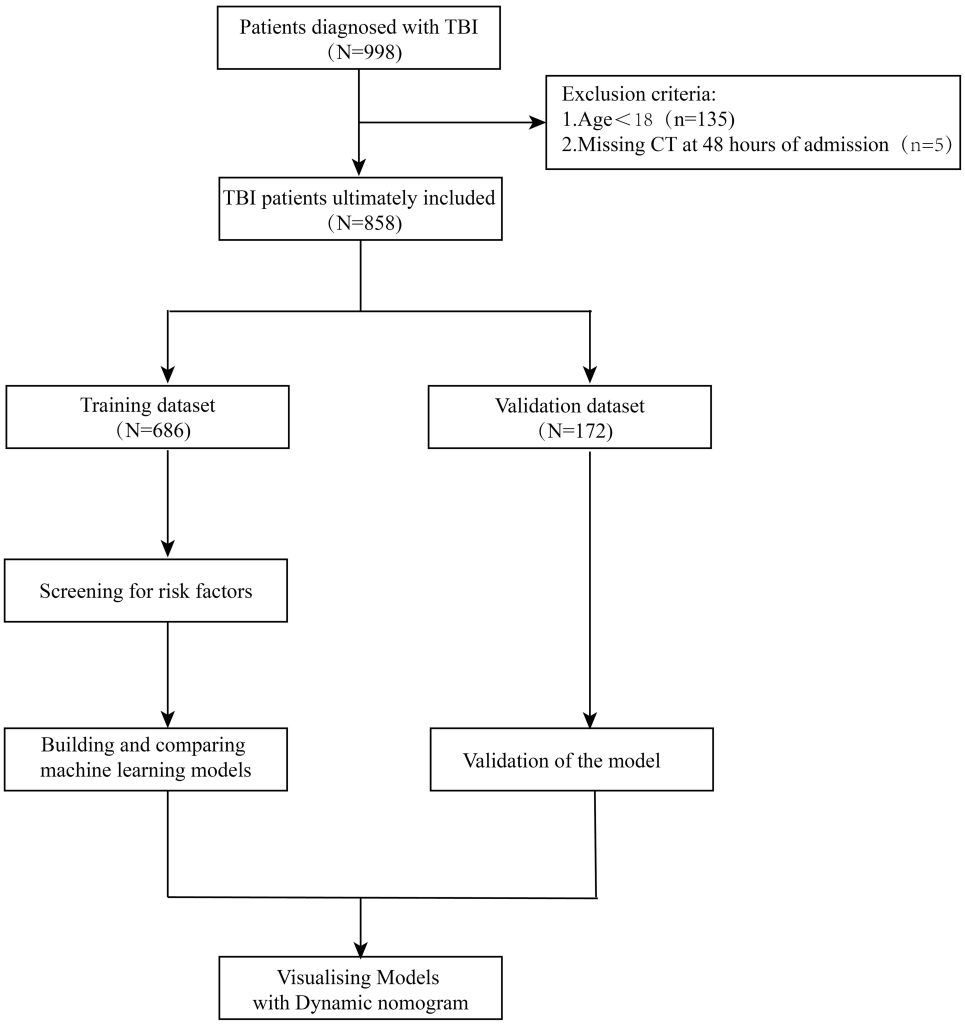

Figure 1 Flow chart of the study.

Development of Predictive Machine Learning Models and Model Evaluation

We used five algorithms to construct a prediction model for HAP and to visualize the strengths and weaknesses of the models, a confusion matrix was built in this study (Figure 4A–E), and several key performance metrics were calculated for each model: Accuracy, MCC, Recall, and F1-Score (Table 3). According to the results, the LR on Recall, Accuracy, and F1-Score all perform well and also have a high MCC value, indicating that it has a better balance and high-quality prediction performance in all of the pneumonia predictions. Although RF is close to LR in MCC, it has a significant shortfall in Recall. In addition, the analysis of the ROC curve reveals that the LR model exhibits significant performance on the training dataset, as evidenced by a higher area under the curve (AUC value of 0.818) (Figure 5A). While on the validation dataset, the LR model exhibited the highest area under the curve (AUC value of 0.819) compared to the other four models (Figure 5B), highlighting its superior generalization ability. The results indicate that LR is the optimal model, which performs more balanced and effective in identifying HAP.

Developing and Validating Nomogram Using Logistic Regression Models

Since the prediction of LR performed well, this study continued to use logistic regression to develop and test predictive models. In the training dataset, seven lasso regression-screened variables such as glucose, type of brain hernia, age, GCS at admission, Rotterdam, SIRI, and D-dimer were included in the multivariate logistic regression analysis. The results showed that age, GCS at admission, Rotterdam, SIRI, and D-dimer were the independent risk factors, while glucose and

Table 1 Baseline Data Comparison Between Training and Validation Datasets

Variable Names	Training (N=686)			Validation (N=172)			
	Non- HAP (N=393)	HAP (N=293)	*P*-value	Non- HAP (N=101)	HAP (N=71)	*P*-value	Total *P*-value
General features of admission							
Age (years)	51.651±17.256	56.795±16.459	<0.001	50.426±14.925	58.465±15.471	0.001	0.942
Temperature (°C)	36.635±0.334	36.702±0.413	0.02	36.662±0.322	36.821±0.543	0.017	0.05
Respiratory rate (per minute)	19.855±1.963	20.188±2.464	0.05	20.178±1.676	20.423±4.631	0.627	0.175
Heart rate (per minute)	82.003±14.176	84.768±18.917	0.029	82.218±12.85	84.592±19.457	0.337	0.992
Systolic pressure (mmHg)	137.379±24.844	145.983±30.229	<0.001	137.416±25.642	142.775±28.565	0.2	0.543
Diastolic pressure (mmHg)	81.153±13.698	81.863±16.336	0.536	80.901±14.018	81.746±14.937	0.705	0.87
GCS on admission	13.321±2.734	9.973±4.179	<0.001	13.475±2.46	9.958±3.743	<0.001	0.678
Gender (%)			0.072			0.812	0.565
Male	255 (64.89)	210 (71.67)		67 (66.34)	45 (63.38)		
Female	138 (35.11)	83 (28.33)		34 (33.66)	26 (36.62)		
History of hypertension (%)			0.23			0.527	0.18
No	314 (79.90)	222 (75.77)		86 (85.15)	57 (80.28)		
Yes	79 (20.10)	71 (24.23)		15 (14.85)	14 (19.72)		
History of diabetes (%)			0.022			0.556	1
No	366 (93.13)	257 (87.71)		90 (89.11)	66 (92.96)		
Yes	27 (6.87)	36 (12.29)		11 (10.89)	5 (7.04)		
Coronary heart disease (%)			0.311			0.984	0.76
No	379 (96.44)	277 (94.54)		98 (97.03)	68 (95.77)		
Yes	14 (3.56)	16 (5.46)		3 (2.97)	3 (4.23)		
Cause of injury (%)			0.216			0.494	0.381
Traffic accidents	49 (12.47)	48 (16.38)		10 (9.90)	11 (15.49)		
Fall from a height	21 (5.34)	21 (7.17)		11 (10.89)	4 (5.63)		
Fall down	63 (16.03)	36 (12.29)		11 (10.89)	8 (11.27)		
Other	260 (66.16)	188 (64.16)		69 (68.32)	48 (67.61)		
Pupillary reflex (%)			<0.001			<0.001	0.506
Sensitive	372 (94.66)	216 (73.72)		94 (93.07)	49 (69.01)		
Slow	7 (1.78)	26 (8.87)		2 (1.98)	10 (14.08)		
Hours	14 (3.56)	51 (17.41)		5 (4.95)	12 (16.90)		
Laboratory examinations			<0.001			0.47	0.946
SII	1823.816±1450.862	2377.313±2040.677		2153.657±1757.857	1952.098±1849.57		
SIRI	43.648±31.92	62.98±58.013	<0.001	55.748±48.906	54.821±53.297	0.906	0.387
NLR	8.154±5.974	10.98±8.829	<0.001	9.778±7.183	8.839±7.37	0.405	0.963
PLR	21.842±15.351	27.695±21.732	<0.001	25.271±18.7	23.284±19.668	0.503	0.946
LMR	2.889±2.049	2.453±2.143	0.007	2.497±1.844	2.988±2.612	0.15	0.988
Leucocyte (10^9/L)	13.128±5.45	15.528±6.504	<0.001	15.457±15.731	14.348±5.414	0.569	0.203
Centriole	77.407±12.949	79.659±12.735	0.024	79.445±11.478	76.344±13.915	0.112	0.852
Lymphocyte	15.919±11.512	13.574±11.43	0.008	13.915±10.102	16.713±13.27	0.118	0.877
Monocyte	5.883±2.309	5.984±2.212	0.562	5.984±2.221	6.162±1.993	0.591	0.491
Hemoglobin (g/L)	132.153±22.402	130.737±20.581	0.397	132.416±17.918	129.606±18.42	0.318	0.87
Urea nitrogen	5.514±5.769	5.716±4.945	0.63	5.035±1.73	5.305±1.83	0.325	0.28
Creatinine	70.483±42.606	73.031±51.193	0.478	63.962±13.636	65.344±17.665	0.564	0.05
Uric acid	341.925±111.392	351.121±129.254	0.319	331.624±110.484	326.366±108.416	0.757	0.102
Potassium	3.612±0.443	3.581±0.52	0.405	3.683±0.508	3.483±0.449	0.008	0.963
Sodium	140.014±4.01	139.633±5.218	0.28	140.401±3.424	140.269±4.08	0.819	0.188
Chloride	103.032±4.236	102.338±4.722	0.044	103.192±3.592	102.531±3.715	0.243	0.617
Calcium	2.239±0.128	2.219±0.153	0.059	2.202±0.133	2.216±0.12	0.501	0.05
Magnesium	0.884±1.354	0.869±1.232	0.881	0.814±0.091	0.801±0.124	0.448	0.487
Carbon dioxide	24.445±3.622	24.179±3.761	0.349	24.293±3.606	24.766±3.722	0.404	0.616
Anion gap	16.164±4.161	16.849±4.205	0.034	16.659±5.287	16.516±4.898	0.857	0.702
Osmotic pressure	276.573±15.418	277.018±8.877	0.658	277.868±6.532	277.947±7.886	0.943	0.27
INR	1.012±0.098	1.047±0.15	<0.001	1.043±0.167	1.029±0.095	0.511	0.356
PT	12.136±1.232	12.506±1.712	0.001	12.518±2.001	12.32±1.218	0.459	0.273
APTT	26.244±5.186	26.595±4.94	0.372	28.052±11.238	25.211±3.615	0.041	0.348

(*Continued*)

Table 1 (Continued).

Variable Names	Training (N=686)			Validation (N=172)			
	Non- HAP (N=393)	HAP (N=293)	*P*-value	Non- HAP (N=101)	HAP (N=71)	*P*-value	Total *P*-value
Fib	2.938±2.986	2.42±1.052	0.005	2.875±3.154	2.548±0.99	0.4	0.91
TT	17.493±2.541	17.865±4.13	0.146	17.74±1.958	17.387±1.56	0.209	0.826
D-dimer	12.546±11.662	19.151±12.734	<0.001	12.228±11.397	19.57±13.237	<0.001	0.92
Platelet (×10^9/L)	229.048±65.32	221.989±68.541	0.171	222.198±58.674	226.535±65.443	0.65	0.715
Glucose (mmol/L)	7.786±2.729	9.229±3.535	<0.001	7.77±2.634	8.788±2.61	0.013	0.421
CT imaging features							
Frontal lobe injury (%)			0.001			0.069	0.665
No	289 (73.54)	180 (61.43)		73 (72.28)	41 (57.75)		
Yes	104 (26.46)	113 (38.57)		28 (27.72)	30 (42.25)		
Temporal lobe injury (%)			0.007			0.004	0.329
No	285 (72.52)	183 (62.46)		74 (73.27)	36 (50.70)		
Yes	108 (27.48)	110 (37.54)		27 (26.73)	35 (49.30)		
Parietal lobe injury (%)			0.022			0.18	0.239
No	367 (93.38)	258 (88.05)		92 (91.09)	59 (83.10)		
Yes	26 (6.62)	35 (11.95)		9 (8.91)	12 (16.90)		
Occipital lobe injury (%)			0.004			1	0.13
No	383 (97.46)	271 (92.49)		99 (98.02)	70 (98.59)		
Yes	10 (2.54)	22 (7.51)		2 (1.98)	1 (1.41)		
Brainstem injury (%)			0.103			0.383	1
No	388 (98.73)	283 (96.59)		100 (99.01)	68 (95.77)		
Yes	5 (1.27)	10 (3.41)		1 (0.99)	3 (4.23)		
Cerebellar injury (%)			0.488			1	1
No	360 (91.60)	263 (89.76)		92 (91.09)	64 (90.14)		
Yes	33 (8.40)	30 (10.24)		9 (8.91)	7 (9.86)		
Epidural hemorrhage (%)			0.074			1	0.721
No	362 (92.11)	257 (87.71)		90 (89.11)	63 (88.73)		
Yes	31 (7.89)	36 (12.29)		11 (10.89)	8 (11.27)		
Parietal fracture of the skull (%)			0.004			0.082	0.253
No	305 (77.61)	198 (67.58)		75 (74.26)	43 (60.56)		
Yes	88 (22.39)	95 (32.42)		26 (25.74)	28 (39.44)		
Fracture of the base of the skull (%)			0.028			0.194	0.562
No	321 (81.68)	218 (74.40)		81 (80.20)	50 (70.42)		
Yes	72 (18.32)	75 (25.60)		20 (19.80)	21 (29.58)		
Subdural hemorrhage (%)			0.001			0.401	0.015
No	354 (90.08)	238 (81.23)		82 (81.19)	53 (74.65)		
Yes	39 (9.92)	55 (18.77)		19 (18.81)	18 (25.35)		
Centerline shift (%)			<0.001			0.002	0.418
No	376 (95.67)	220 (75.09)		97 (96.04)	57 (80.28)		
Yes	17 (4.33)	73 (24.91)		4 (3.96)	14 (19.72)		
Site of brain hernia (%)			<0.001			0.001	0.262
Absence of brain hernia	377 (95.93)	205 (69.97)		99 (98.02)	56 (78.87)		
Unilateral temporal sulcus hernia	12 (3.05)	71 (24.23)		2 (1.98)	13 (18.31)		
Bilateral temporal sulcus hernias	0 (0.00)	5 (1.71)		0 (0.00)	1 (1.41)		
Foramen magnum occipitalis hernia	4 (1.02)	12 (4.10)		0 (0.00)	1 (1.41)		
Rotterdam	1.964±0.92	3.078±1.361	<0.001	1.931±0.886	2.775±1.111	<0.001	0.122

type of brain hernia were not statistically significant in the multivariate model (Figure 6). Meanwhile, the variance inflation factors of the independent risk factors were all covariance diagnostic <2, indicating no multicollinearity between risk factors.

The validation dataset was used to evaluate the performance of the predictive model. The AUCs in the training (Figure 7A) and validation datasets were 0.818 and 0.819, respectively, and the optimal cut-off values were 0.41 and 0.402, respectively. The calibration curve results showed that the deviation of the ideal curve from the actual curve was

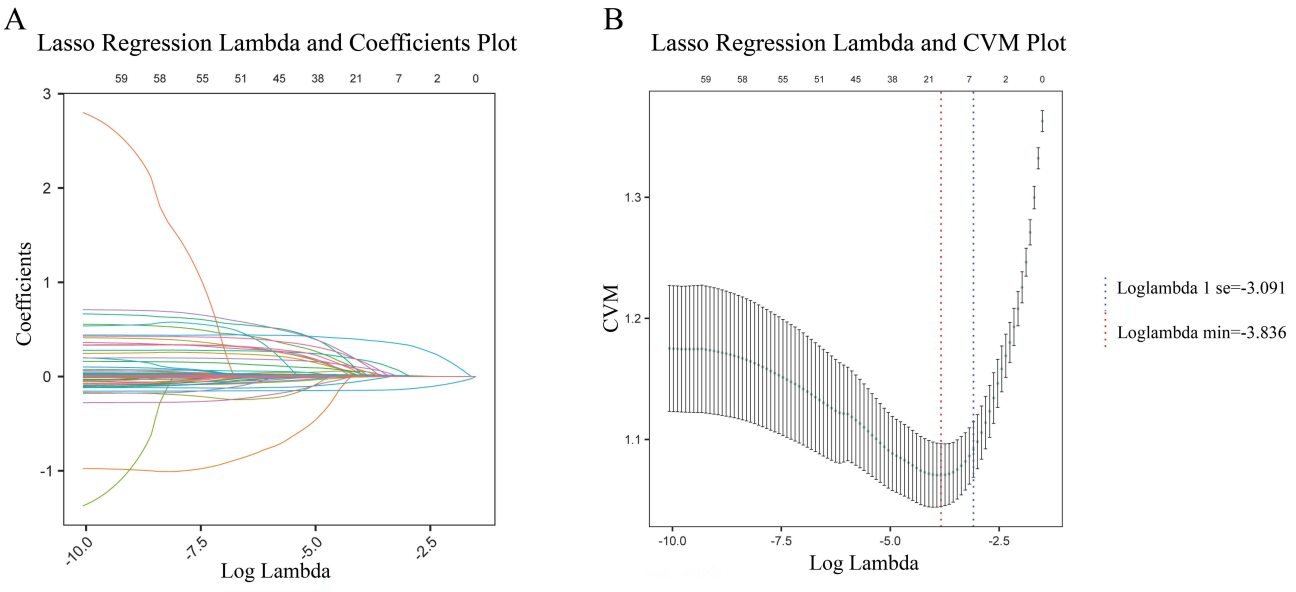

Figure 2 Lasso regression results for admission clinical characteristics and imaging characteristics variables. (**A**) LASSO regression path diagram. (**B**) LASSO Cross validation diagram.

small. The actual prediction curves were consistent with the ideal curves (B=1000 repetitions), indicating that the prediction model had good accuracy (mean absolute error (train) = 0.005 and mean absolute error (validation) = 0.031). The predictive model was internally validated to have some predictive value (Figure 7). In addition, both the training dataset and validation dataset calibration curves showed good performance according to the Hosmer-Lemeshow test results (p (train) = 0.4307; p(validation) = 0.3387) (Figure 8). The DCA for the training dataset showed that patients could achieve satisfactory net benefit from the predictive model, with a wide range of high-risk thresholds for the DCA (10–85%), which also performed well in the validation dataset (10–80%) (Figure 9).

All independent predictors were considered in the construction of the dynamic nomogram. Compared to the complex logistic regression formulae, the dynamic nomogram are simple straightforward, and more clinically useful (Figure 10). As an example, the first participant in this database (age 83 years, GCS score of 15, Rotterdam score of 3, SIRI of 42.058, and D-dimer of 11.97) had a risk of 49.8% (95% CI=0.4–0.597) of developing HAP. Thus, the dynamic nomogram allows rapid screening of patients at high risk of HAP based on individual patient admissions.

Table 2 Table of Regression Coefficients for Important Variables Related to HAP (Training Dataset)

Coef Name	Coeff_se_Lamda
(Intercept)	−0.35068
Age	0.009833
Site of brain hernia	0.039852
GCS	−0.13516
Rotterdam	0.341734
Glucose	0.004465
D-dimer	0.009684
SIRI	0.001027

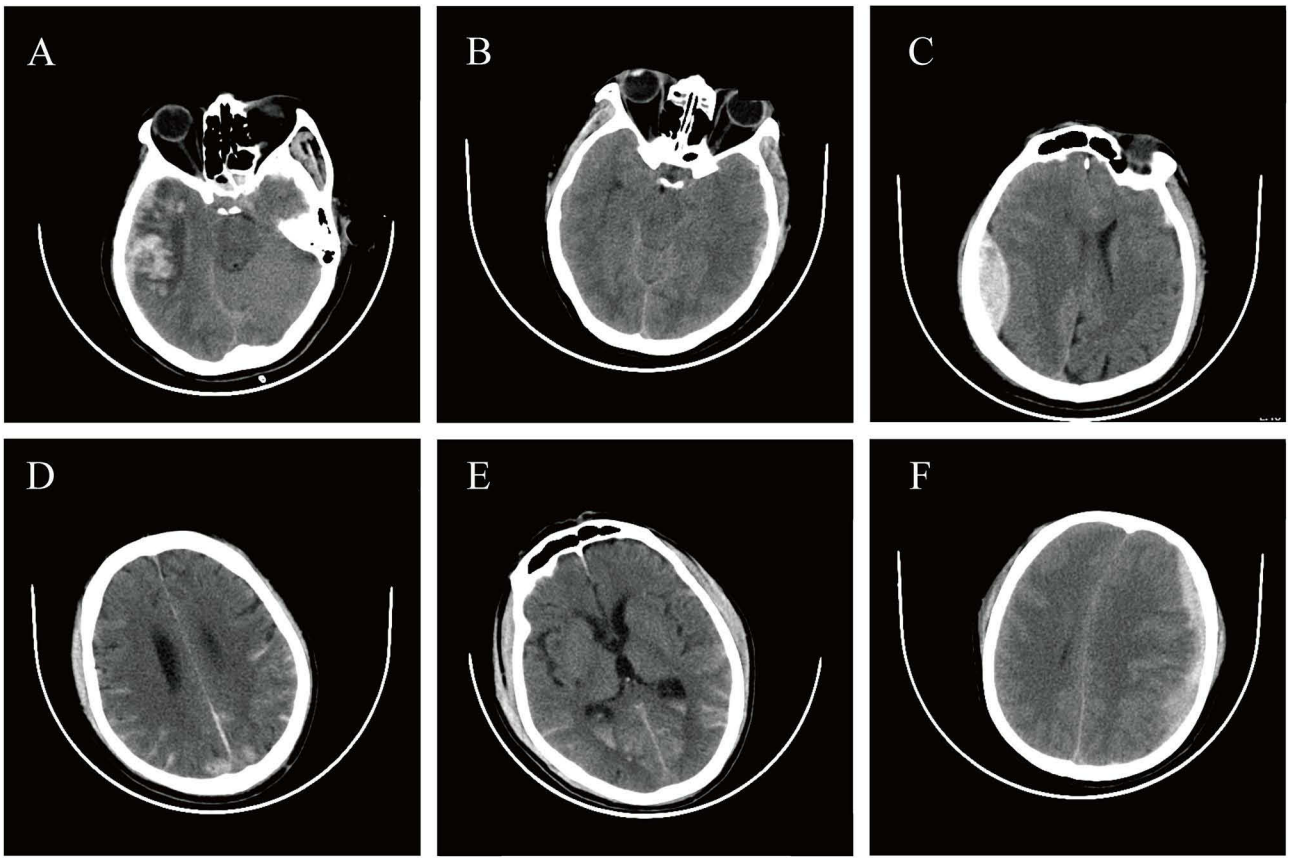

Figure 3 Rotterdam CT score corresponding to CT signs. (**A**) Compression of the cerebral pool. (**B**) Loss of the cerebral pool. (**C**) Epidural hemorrhage. (**D**) Subarachnoid hemorrhage. (**E**) Ventricular hemorrhage. (**F**) midline shift ≥5 mm.

Discussion

This is a study using machine algorithm models to predict the incidence of HAP in patients with TBI. In this retrospective study, we used five machine-learning algorithms for the assessment of outcome events, and LR performed well on Recall, Accuracy, and F1-Score, along with high MCC values. In addition, LR showed a higher area under the curve in both the training and validation datasets (train-AUC: 0.818; validation-AUC: 0.819). The results indicate that LR has a good balance of high-quality predictive performance coupled with superior generalization ability. To further explain the model, based on the optimal model, we identified age, GCS on admission, Rotterdam, SIRI, and D-dimer as independent risk factors for HAP in TBI patients. In the validation of the model, the calibration curve and DCA of LR performed well. Finally, a simple and accurate web-based dynamic nomogram was developed in this study (https://shaojieli.shinyapps.io/DynNomapp/).

HAP after TBI is a complication following brain injury due to brain-pulmonary axis interactions.[17,18] Pneumonia not only increases the economic burden of patients with traumatic brain injury but is also strongly associated with poor disease prognosis. Predictive models for predicting the prevalence of pneumonia in stroke patients have been explored in several previous studies.[19–21] In a population-based study of 1333 patients with spontaneous cerebral hemorrhage, older age, multilobar involvement, extension of ICH to the ventricles, dysphagia, impaired consciousness, and poor muscle strength in male patients were included as independent risk factors for pneumonia after ICH, and the predictive accuracy was shown to be 0.8116 by a ROC curve.[19] In a prospective study conducted by Wang et al cerebral hemorrhage, NLR, SII and SIRI at admission had high predictive accuracy for pneumonia and were strongly associated with poorer neurological outcomes at discharge.[20] Similarly, in patients with ischaemic stroke (IS), the prevalence of pneumonia is of interest. A retrospective study combining five large medical centers demonstrated that the severity of malnutrition was associated with the prevalence of pneumonia in patients with IS and that the malnutrition index was an independent

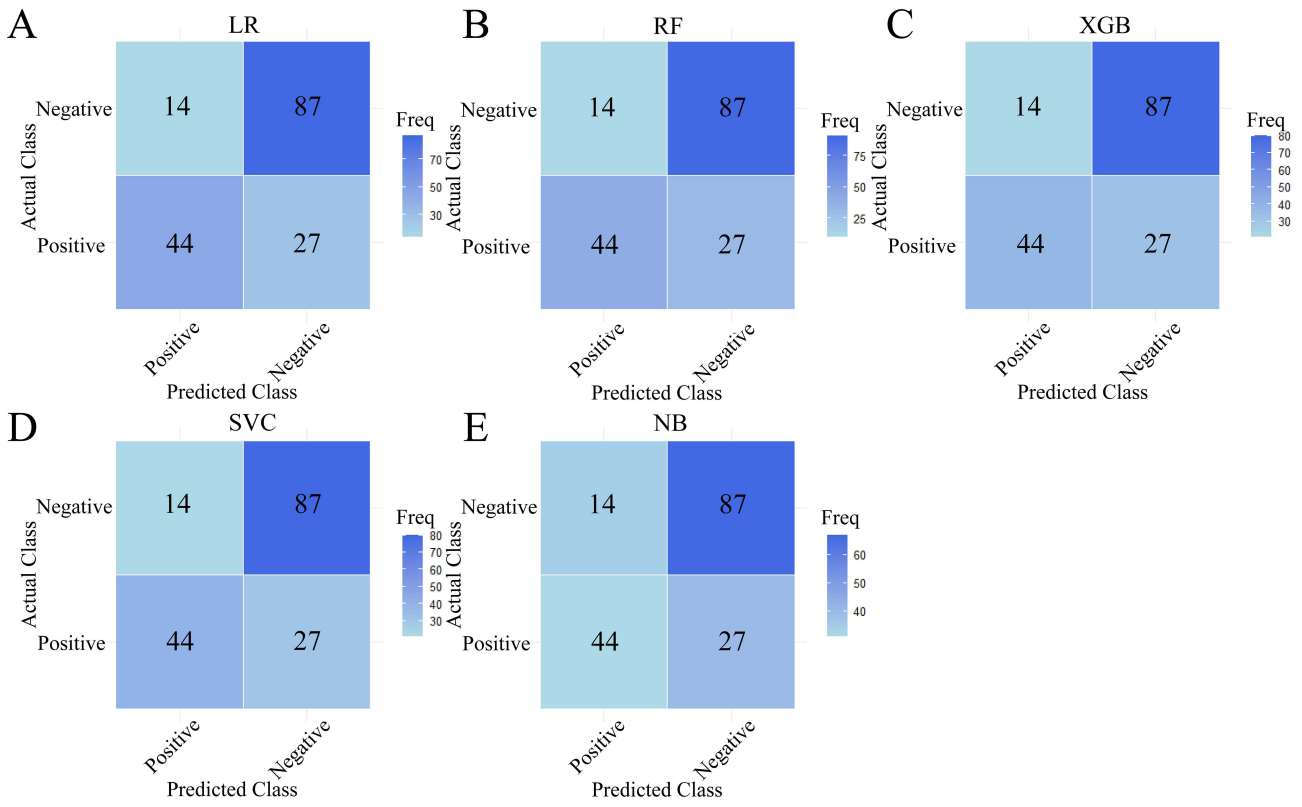

Figure 4 (**A–E**) represents the confusion matrix for each of the five models in the validation dataset. Every classification model's performance is detailed in the confusion matrix. For example, the LR model has a balanced prediction of 27 false negatives and 14 false positives.

predictor of pneumonia and improved the predictive efficacy of pneumonia after stroke after adjusting for confounders.[22] However, to the best of our knowledge, no studies are reporting early predictive tools that can identify the occurrence of HAP in patients with TBI. Rotterdam CT score is an objective representation of the imaging characteristics of patients with traumatic brain injury and is commonly used in the assessment of the severity of TBI.[23] Previous studies have mainly used it for intracranial hypertension monitoring[24] and mortality studies.[25] In this study, it was used for the first time to predict pneumonia in patients with TBI. The results showed that, in addition to the above data on basic clinical features, combining some of the cranial imaging examinations better improves the diagnostic efficacy of pneumonia.

In this study, we analyzed in depth the risk factors associated with the development of HAP in patients with TBI. The findings revealed that elderly patients and TBI patients with lower GCS were more likely to develop pneumonia, a finding that is consistent with previous studies.[19,26,27] In addition, we paid special attention to the use of the Rotterdam CT score in the management of TBI. The Rotterdam CT score is based on several metrics on cranial CT, including midline shift of more than

Table 3 Performance Metrics of the Five Machine Learning Models

Model Name	Train					Validation				
	AUC	Recall	Accuracy	F1-Score	MCC	AUC	Recall	Accuracy	F1-Score	MCC
LR	0.818	0.631	0.761	0.693	0.507	0.819	0.620	0.762	0.682	0.501
RF	0.846	0.601	0.771	0.691	0.531	0.811	0.3	0.762	0.661	0.503
XGB	1	1	0.999	0.998	0.997	0.794	0.563	0.698	0.606	0.366
SVC	0.802	0.532	0.742	0.638	0.472	0.752	0.535	0.686	0.584	0.339
NB	0.683	0.553	0.659	0.581	0.295	0.609	0.437	0.570	0.456	0.102

Notes: Specificity = TN/(TN + FP); Recall = TP/(TP + FN); Accuracy = (TP + TN)/(TP + TN + FP + FN); F1 score = 2/([1/Recall] + [1/Precision]); MCC= $(TP\times TN)-(FP\times FN)/\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}$.

Abbreviations: FN, false negatives; FP, false positives; TN, true negatives; TP, true positives; AUC, Area Under Curve.

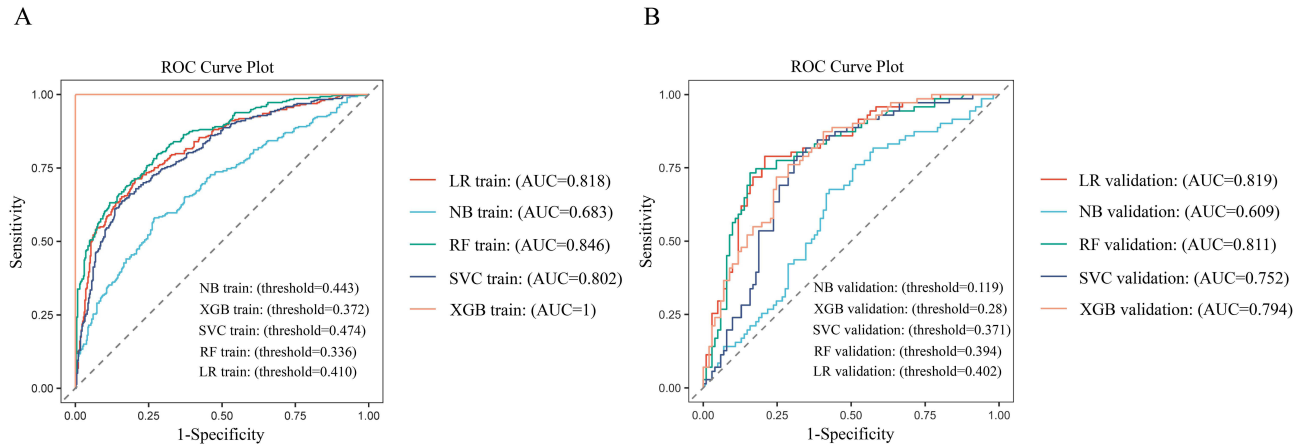

Figure 5 (**A** and **B**) represents the ROC curves of the five models in the training and validation datasets, respectively.

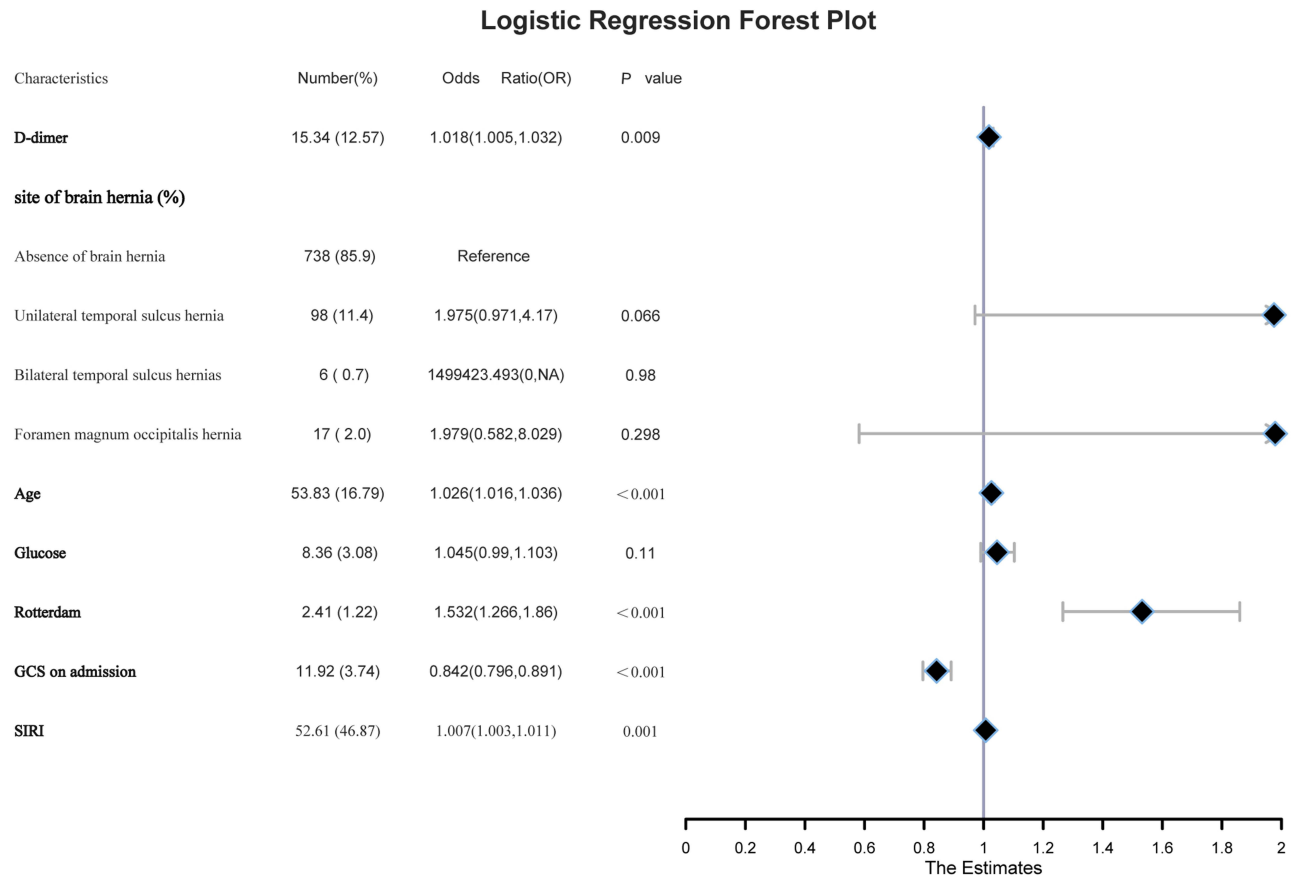

Figure 6 Multiple logistic regression forest plot in the training dataset.

5 mm, intraventricular hemorrhage, subarachnoid hemorrhage, epidural space-occupying lesions, and compression or disappearance of the basaltic pools,[23] and provides an objective, comprehensive and convenient method.[27] Previous studies have focused on the use of this score to predict intracranial hypertension[24] and mortality,[25] and the present study is the first to investigate its application in predicting the risk of HAP in patients with TBI. By combining data on basic clinical characteristics with cranial imaging findings, we found that the Rotterdam CT score is not only an independent predictor of poor outcomes in patients with TBI but also improves the efficacy of pneumonia diagnosis. Our analyses showed that the

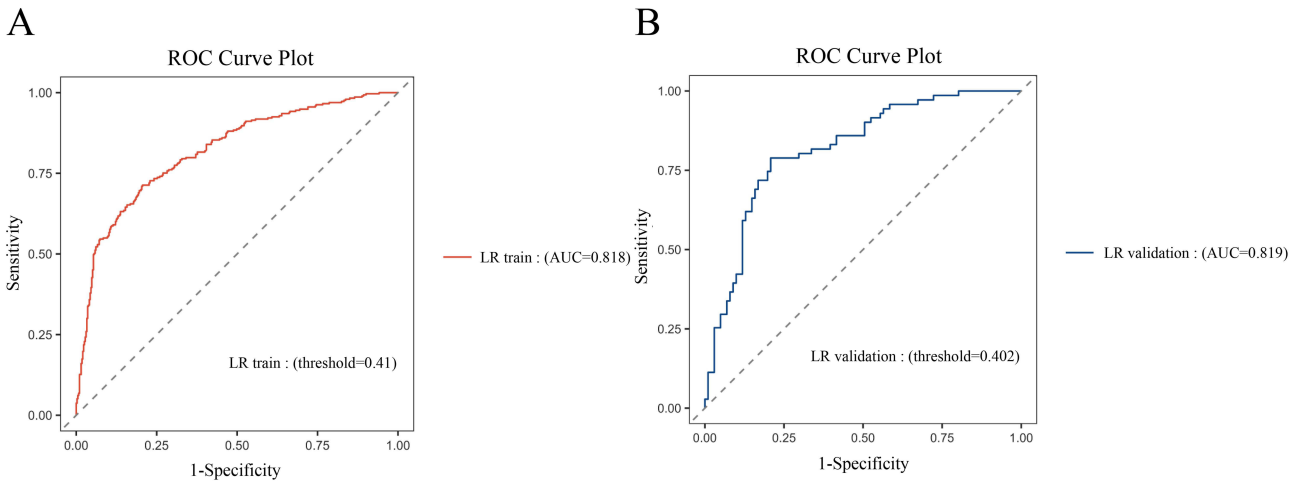

Figure 7 (**A** and **B**) represents the ROC curves in the training and validation datasets, respectively.

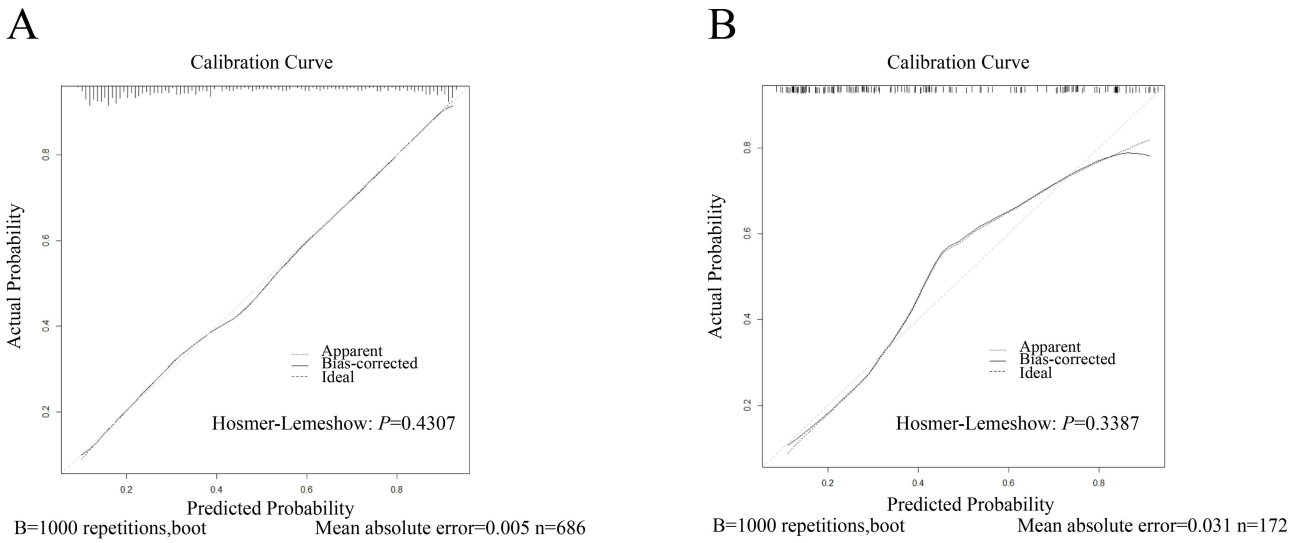

Figure 8 (**A** and **B**) represents the calibration curves for the training and validation datasets, respectively.

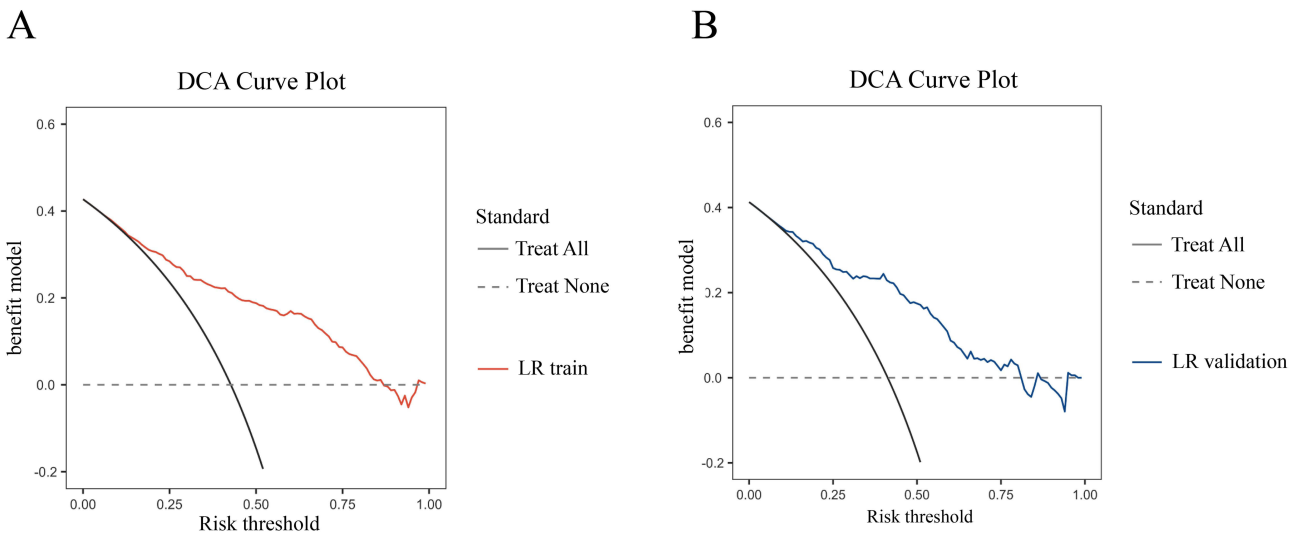

Figure 9 (**A** and **B**) represent the DCA curves in the training dataset and validation dataset, respectively.

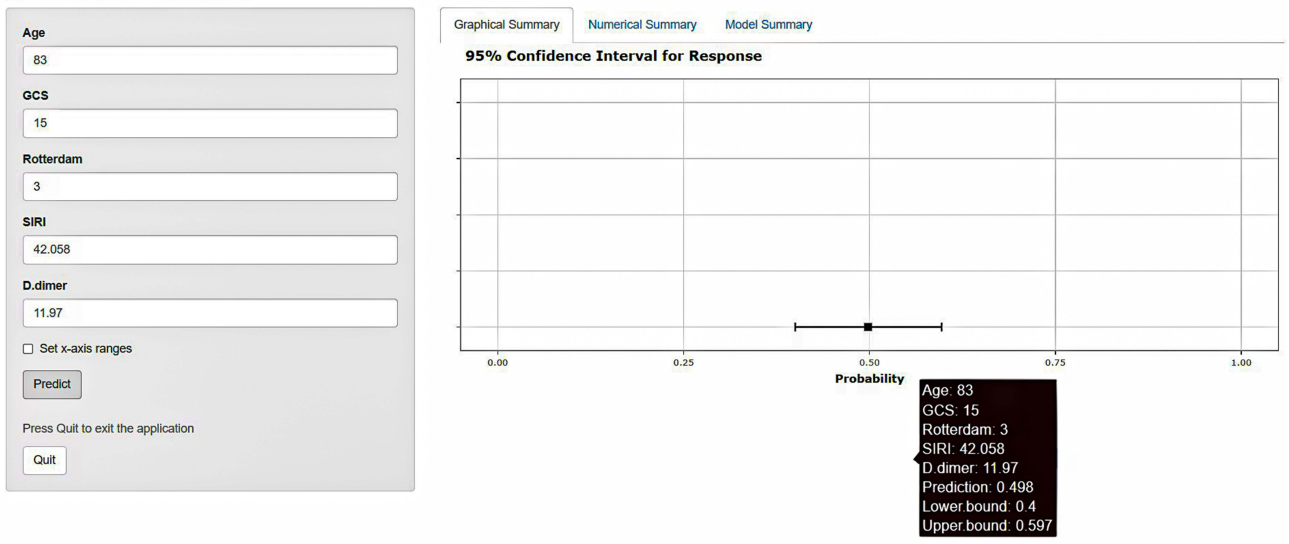

Figure 10 Web-based dynamic nomogram used to predict HAP after TBI. Predicted probabilities and 95% confidence intervals can be found on the right side of the page by entering the precise values of the respective variables on the left side.

Rotterdam CT score presented a significant correlation with the occurrence of HAP. We hypothesize that the mechanism is large as follows: elevated Rotterdam CT scores usually reflect severe brain damage in patients, such as compression or loss of brain pools and cerebral hemorrhage. These brain injuries may affect the respiratory control center, leading to impaired regulation of respiratory function, which in turn causes inadequate lung ventilation and retention of respiratory secretions,[28] increasing the risk of pulmonary infection in patients. In addition, more severe neurological impairment may be accompanied by decreased cough and swallowing reflexes, making it difficult for patients to effectively clear respiratory secretions,[29] thus increasing the risk of HAP. Within the current field of research, the association between SIRI and the development of HAP after TBI and its underlying mechanisms are not fully understood. However, the association of SIRI with stroke-associated pneumonia (SAP) after hemorrhagic stroke has been empirically supported.[30] It has been suggested that a sustained inflammatory response may lead to decreased immune system function and increased patient susceptibility to SAP. In addition, SIRI was found to predict the risk of SAP and adverse outcomes at discharge in patients with hemorrhagic stroke (ICH). Further, Yan et al revealed by constructing a response curve (RCS) model that in ischaemic stroke patients, the risk of pneumonia was significantly increased when the SIRI value reached or exceeded 2.74 (OR: 5.82, 95% CI: 4.54–7.49), implying that higher SIRI was a significant risk factor for the development of SAP in acute ischaemic stroke patients.[31] This coincides with our findings. Consequently, early identification of patients at high risk of pneumonia and the use of suitable therapies may be made possible by monitoring SIRI. Chen et al found that D-dimer levels were elevated early in the development of TBI and may predict short- and long-term mortality in patients with TBI.[32] Elevated D-dimer levels have been linked to the degree of tissue damage following traumatic brain injury, according to recent research.[32,33] Tissue factor (TF) is released from the compromised blood-brain barrier throughout the body during the acute phase of traumatic brain injury (TBI), which causes a sharp rise in plasma D-dimer levels.[34,35] Elevated D-dimer enhances local and systemic inflammatory responses and immunosuppression by modulating local immune responses and releasing pro-inflammatory factors (eg, IL-6).[36] A 2021 correlation study indicated that D-dimer, a traditional marker of fibrinolysis, fluctuates faster than serum CRP and ESR during inflammation and may be effective for early diagnosis of infections,[37] which supports our conjecture.

This study combines some of the imaging indices to explore, for the first time, the risk factors associated with the development of HAP in patients with TBI, and produces a simple dynamic column chart to guide the clinic. This study does, however, still have certain shortcomings. First, this is a single-center retrospective study, and although we screened the optimal model from several different machine-learning algorithms to try to ensure the accuracy of the prediction, there is inevitably a selection bias. We hope that clinical data from multiple centers will be used as external validation of

the model in the future. Second, this study collected as many clinically available variables as possible, but some other clinical information may have had confounding effects and influenced the risk of pneumonia. For example, influencing patients' hormone and immunosuppressant use may increase the risk of pneumonia.

Conclusions

In this study, we have implemented five distinct machine learning algorithms to develop predictive models for HAP in patients with TBI. Among these, the LR model has shown superior performance, effectively highlighting its robustness in handling clinical data. This approach has enabled us, for the first time, to pinpoint key independent risk factors for HAP in TBI patients, namely age, Rotterdam, GCS, SIRI, and D-dimer levels. Our predictive model significantly enhances clinical utility by employing a dynamic nomogram format, providing healthcare professionals with an effective tool to swiftly identify patients at risk of pneumonia. This allows for the implementation of individualized treatment strategies for high-risk patients. Specific interventions include targeted nebulization, physical expectoration techniques, and intensive cough training. These are supplemented by the proactive use of prophylactic antibiotics and ongoing monitoring of relevant clinical indicators. Such personalized interventions are designed to optimize respiratory function and reduce the complications associated with hospital-acquired pneumonia (HAP), potentially decreasing its incidence and significantly improving patient outcomes. However, future enhancements to the prediction model can be realized by conducting a larger, well-structured prospective study, which should include external validation to confirm its effectiveness.

Data Sharing Statement

Research data are not available at this time.

Ethical Approval and Consent to Participate

The study protocol was approved by the Ethics Committee of the Second Affiliated Hospital of Fujian Medical University (2023-287). This study was conducted in accordance with the Declaration of Helsinki. The need for written informed consent was waived by the Ethics Committee of the Second Affiliated Hospital of Fujian Medical University due to the non-interventional design of the study. Additionally, this research exclusively utilized previously collected medical record information from which all personally identifiable information had been removed, ensuring no risk to the subjects and no adverse effects on their rights and health.

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

The work was supported by Quanzhou Science and Technology Program (2022NS067).

Disclosure

The authors declare that they have no competing interests.

References

1. Lefevre-Dognin C, Cogné M, Perdrieau V, Granger A, Heslot C, Azouvi P. Definition and epidemiology of mild traumatic brain injury. *Neurochirurgie*. 2021;67:218–221. doi:10.1016/j.neuchi.2020.02.002
2. Mukherjee S, Arisi GM, Mims K, Hollingsworth G, O'Neil K, Shapiro LA. Neuroinflammatory mechanisms of post-traumatic epilepsy. *J Neuroinflammation*. 2020;17:193. doi:10.1186/s12974-020-01854-w
3. Pearn ML, Niesman IR, Egawa J, et al. Pathophysiology associated with traumatic brain injury: current treatments and potential novel therapeutics. *Cell Mol Neurobiol*. 2017;37:571–585. doi:10.1007/s10571-016-0400-1

4. Bajinka O, Simbilyabo L, Tan Y, Jabang J, Saleem SA. Lung-brain axis. *Crit Rev Microbiol*. 2022;48:257–269. doi:10.1080/1040841X.2021.1960483
5. Heuer JF, Pelosi P, Hermann P, et al. Acute effects of intracranial hypertension and ARDS on pulmonary and neuronal damage: a randomized experimental study in pigs. *Intensive Care Med*. 2011;37:1182–1191. doi:10.1007/s00134-011-2232-2
6. Parker A, Fonseca S, Carding SR. Gut microbes and metabolites as modulators of blood-brain barrier integrity and brain health. *Gut Microbes*. 2020;11:135–157. doi:10.1080/19490976.2019.1638722
7. Li Y, Liu C, Xiao W, Song T, Wang S. Incidence, risk factors, and outcomes of ventilator-associated pneumonia in traumatic brain injury: a meta-analysis. *Neurocrit Care*. 2020;32:272–285. doi:10.1007/s12028-019-00773-w
8. Martin-Loeches I, Rodriguez AH, Torres A. New guidelines for hospital-acquired pneumonia/ventilator-associated pneumonia: USA vs. *Eur Curr Opin Crit Care*. 2018;24:347–352. doi:10.1097/MCC.0000000000000535
9. Kalil AC, Metersky ML, Klompas M, et al. Management of adults with hospital-acquired and ventilator-associated pneumonia: 2016 clinical practice guidelines by the infectious diseases society of America and the American thoracic society. *Clin Infect Dis*. 2016;63:e61–111. doi:10.1093/cid/ciw353
10. Klompas M, Branson R, Cawcutt K, et al. Strategies to prevent ventilator-associated pneumonia, ventilator-associated events, and nonventilator hospital-acquired pneumonia in acute-care hospitals: 2022 Update. *Infect Control Hosp Epidemiol*. 2022;43:687–713. doi:10.1017/ice.2022.88
11. Schwarz S. Prophylactic antibiotic therapy for preventing poststroke infection. *Neurotherapeutics*. 2016;13:783–790. doi:10.1007/s13311-016-0466-y
12. Meisel A. Preventive antibiotic therapy in stroke: pASSed away? *Lancet*. 2015;385:1486–1487. doi:10.1016/S0140-6736(15)60076-9
13. Harms H, Prass K, Meisel C, et al. Preventive antibacterial therapy in acute ischemic stroke: a randomized controlled trial. *PLoS One*. 2008;3:e2158. doi:10.1371/journal.pone.0002158
14. Dragan V, Wei Y, Elligsen M, Kiss A, Walker SAN, Leis JA. Prophylactic antimicrobial therapy for acute aspiration pneumonitis. *Clin Infect Dis*. 2018;67:513–518. doi:10.1093/cid/ciy120
15. Reizine F, Asehnoune K, Roquilly A, et al. Effects of antibiotic prophylaxis on ventilator-associated pneumonia in severe traumatic brain injury. A post hoc analysis of two trials. *J Crit Care*. 2019;50:221–226. doi:10.1016/j.jcrc.2018.12.010
16. Pei Y, Huang Y, Pan X, et al. Nomogram for predicting 90-day mortality in patients with Acinetobacter baumannii-caused hospital-acquired and ventilator-associated pneumonia in the respiratory intensive care unit. *J Int Med Res*. 2023;51:3000605231161481. doi:10.1177/03000605231161481
17. Weber DJ, Allette YM, Wilkes DS, White FA. The HMGB1-RAGE inflammatory pathway: implications for brain injury-induced pulmonary dysfunction. *Antioxid Redox Signal*. 2015;23:1316–1328. doi:10.1089/ars.2015.6299
18. Kerr N, de Rivero Vaccari JP, Dietrich WD, Keane RW. Neural-respiratory inflammasome axis in traumatic brain injury. *Exp Neurol*. 2020;323:113080. doi:10.1016/j.expneurol.2019.113080
19. Wang Y, Chen Y, Chen R, et al. Development and validation of a nomogram model for prediction of stroke-associated pneumonia associated with intracerebral hemorrhage. *BMC Geriatr*. 2023;23:633. doi:10.1186/s12877-023-04310-5
20. Wang R-H, Wen W-X, Jiang Z-P, et al. The clinical value of neutrophil-to-lymphocyte ratio (NLR), systemic immune-inflammation index (SII), platelet-to-lymphocyte ratio (PLR) and systemic inflammation response index (SIRI) for predicting the occurrence and severity of pneumonia in patients with intracerebral hemorrhage. *Front Immunol*. 2023;14:1115031. doi:10.3389/fimmu.2023.1115031
21. Li J, Luo H, Chen Y, et al. Comparison of the predictive value of inflammatory biomarkers for the risk of stroke-associated pneumonia in patients with acute ischemic stroke. *Clin Interv Aging*. 2023;18:1477–1490. doi:10.2147/CIA.S425393
22. Li D, Liu Y, Jia Y, et al. Association between malnutrition and stroke-associated pneumonia in patients with ischemic stroke. *BMC Neurol*. 2023;23:290. doi:10.1186/s12883-023-03340-1
23. Amakhian AO, Obi-Egbedi-Ejakpovi EB, Morgan E, Adeyekun AA, Abubakar MM. Correlation between optic nerve sheath diameter at initial head CT and the Rotterdam CT score. *Cureus*. 2023;15:e41995. doi:10.7759/cureus.41995
24. Cnossen MC, Huijben JA, van der Jagt M, et al. Variation in monitoring and treatment policies for intracranial hypertension in traumatic brain injury: a survey in 66 neurotrauma centers participating in the CENTER-TBI study. *Crit Care*. 2017;21:233. doi:10.1186/s13054-017-1816-9
25. Voormolen DC, Zeldovich M, Haagsma JA, et al. Outcomes after complicated and uncomplicated mild traumatic brain injury at three-and six-months post-injury: results from the CENTER-TBI study. *J Clin Med*. 2020;9:1525. doi:10.3390/jcm9051525
26. Zawiah M, Khan AH, Abu Farha R, et al. Predictors of stroke-associated pneumonia and the predictive value of neutrophil percentage-to-albumin ratio. *Postgrad Med*. 2023;135:681–689. doi:10.1080/00325481.2023.2261354
27. Garza N, Toussi A, Wilson M, Shahlaie K, Martin R. The increasing age of TBI patients at a single level 1 trauma center and the discordance between GCS and CT Rotterdam scores in the elderly. *Front Neurol*. 2020;11:112. doi:10.3389/fneur.2020.00112
28. Meyfroidt G, Bouzat P, Casaer MP, et al. Management of moderate to severe traumatic brain injury: an update for the intensivist. *Intensive Care Med*. 2022;48:649–666. doi:10.1007/s00134-022-06702-4
29. Terré R, Mearin F. Prospective evaluation of oro-pharyngeal dysphagia after severe traumatic brain injury. *Brain Inj*. 2007;21:1411–1417. doi:10.1080/02699050701785096
30. Xu M, Wang J, Zhan C, et al. Association of follow-up neutrophil-to-lymphocyte ratio and systemic inflammation response index with stroke-associated pneumonia and functional outcomes in cerebral hemorrhage patients: a case controlled study. *Int J Surg*. 2024. doi:10.1097/JS9.0000000000001329
31. Yan D, Dai C, Xu R, Huang Q, Ren W. Predictive ability of systemic inflammation response index for the risk of pneumonia in patients with acute ischemic stroke. *Gerontology*. 2023;69:181–188. doi:10.1159/000524759
32. Chen X, Wang X, Liu Y, et al. Plasma D-dimer levels are a biomarker for in-hospital complications and long-term mortality in patients with traumatic brain injury. *Front Mol Neurosci*. 2023;16:1276726. doi:10.3389/fnmol.2023.1276726
33. Kowalski EA, Chen J, Hazy A, et al. Peripheral loss of EphA4 ameliorates TBI-induced neuroinflammation and tissue damage. *J Neuroinflammation*. 2019;16:210. doi:10.1186/s12974-019-1605-2
34. Hubbard WB, Sim MMS, Saatman KE, Sullivan PG, Wood JP. Tissue factor release following traumatic brain injury drives thrombin generation. *Res Pract Thromb Haemost*. 2022;6:e12734. doi:10.1002/rth2.12734
35. Kwaan HC. The central role of fibrinolytic response in trauma-induced coagulopathy: a hematologist's perspective. *Semin Thromb Hemost*. 2020;46:116–124. doi:10.1055/s-0039-3402428

36. Medcalf RL, Keragala CB, Draxler DF. Fibrinolysis and the immune response in trauma. *Semin Thromb Hemost*. 2020;46:176–182. doi:10.1055/s-0040-1702170
37. Chen X, Li H, Zhu S, Wang Y, Qian W. Is D-dimer a reliable biomarker compared to ESR and CRP in the diagnosis of periprosthetic joint infection? *Bone Joint Res*. 2020;9:701–708. doi:10.1302/2046-3758.910.BJR-2020-0172.R2