BMC Medical Informatics and Decision Making

RESEARCH

Open Access

Death risk prediction model for patients with non-traumatic intracerebral hemorrhage

Yidan Chen[1†], Xuhui Liu[2†], Mingmin Yan[3*] and Yue Wan[3*]

Abstract

Background This study aimed to assess the risk of death from non-traumatic intracerebral hemorrhage (ICH) using a machine learning model.

Methods 1274 ICH patients who met the specified inclusion and exclusion criteria were analyzed retrospectively in the MIMIC IV 3.0 database. Patients were randomly divided into training, validation, and testing datasets in a ratio of 6:2:2 based on the outcome distribution. Data from the Second Hospital of Lanzhou University were used as an external validation set. This study used LASSO regression and multivariable logistic regression analysis to screen for features. We then employed XGBoost to construct a machine-learning model. The model's performance was evaluated using ROC curve analysis, calibration curve analysis, clinical decision curve analysis, sensitivity, specificity, accuracy, and F1 score. Conclusively, the SHapley Additive exPlanations (SHAP) method was employed to interpret the model's predictions.

Results Deaths occurred in 572 out of the 1274 ICH cases included in the study, resulting in an incidence rate of 44.9%. The XGBoost model achieved a high AUC when predicting deaths in ICH patients (train: 0.814, 95%CI: 0.784 – 0.844; validation: 0.715, 95%CI: 0.653 – 0.777; test: 0.797, 95%CI: 0.743 – 0.851). The importance of SHAP variables in the model ranked from high to low was: 'GCS motor', 'Age', 'GCS eyes', 'Low density lipoprotein (LDL)', ' Albumin', ' Atrial fibrillation', and 'Gender'. The XGBoost model demonstrated good predictive performance in both the validation and external validation datasets.

Conclusions The XGBoost machine learning model we built has demonstrated strong performance in predicting the risk of death from ICH. Furthermore, the SHAP provides the possibility of interpreting machine learning results.

Keywords Non-traumatic intracerebral hemorrhage, Prediction model, Machine learning, SHAP

†Yidan Chen and Xuhui Liu contributed equally to this work.

*Correspondence:
Mingmin Yan
2009203020074@whu.edu.cn
Yue Wan
wylydia@aliyun.com
[1]Jianghan University School of Medicine, Wuhan, China
[2]Department of Neurology, The Second Hospital of Lanzhou University, Lanzhou, China
[3]Department of Neurology, School of Medicine, Jianghan University, Hubei No. 3 People's Hospital, Wuhan 430033, China

Introduction

Nontraumatic intracerebral hemorrhage (ICH) is the second most common type of stroke globally, affecting 20% [1] of the estimated 17.9 million stroke patients worldwide [2]. Despite advancements in understanding the natural history, treatment approaches, and prognosis of ICH, its incidence and mortality rates remain alarmingly high [3]. Within the first month after injury, 54% of ICH patients succumb to the condition, underscoring the urgent need for improved strategies to guide supportive treatment and prognostic management [4]. A critical step in improving outcomes is the early prediction of long-term results, such as mortality risk. At the time of admission, clinicians often have access to limited information, such as vital signs, Glasgow Coma Scale (GCS) scores, and laboratory test results. Accurately predicting patient mortality using these data is critical for enabling early intervention and informed clinical decision-making.

The large volume of data generated for ICH patients presents significant challenges for traditional statistical methods, which struggle to account for complex interactions and nonlinear relationships among variables. Machine learning (ML) algorithms, which excel at handling such complexities, have been widely adopted in the analysis of medical big data [5–7]. Nevertheless, clinical prediction models for ICH are rarely implemented in routine practice. This underutilization may partly stem from the limitations of existing models in terms of generalizability and predictive performance. Furthermore, few studies have specifically focused on mortality risk as the primary outcome in ICH patients. To address these gaps, this study aims to develop an interpretable ML model to accurately predict mortality risk in patients with nontraumatic ICH. Specifically, the objectives of this study are to: (1) identify key predictors of mortality using clinical data available at admission, such as vital signs, GCS scores, and laboratory test results; (2) evaluate the predictive performance of the proposed model and compare it with results from previous studies; and (3) ensure the interpretability of the model to enhance its potential application in routine clinical decision-making.

Literature review

The PubMed published literature on death prediction of ICH patients was searched by a combination of computer-based and manual searches was employed. The search was carried out by combining subject words and free words. English search terms included: ("intracerebral hemorrhage" OR "ICH" OR "non-traumatic intracerebral hemorrhage" OR "spontaneous intracerebral hemorrhage") AND ("mortality prediction" OR "risk prediction" OR "prognostic model" OR "predictive model") AND ("machine learning" OR "artificial intelligence" OR "XGBoost" OR "logistic regression" OR "SHAP" OR "Shapley additive explanations" OR "interpretability" OR "explainability") AND ("critical care" OR "intensive care" OR "clinical decision-making" OR "MIMIC database" OR "MIMIC-IV"). After excluding irrelevant studies, we conducted a comprehensive review of the selected studies to support the development of our model. We conducted a comprehensive review of the literature to support the development of our model. The key characteristics of the included literature are available in the supplementary Excel document (Supplementary File 1).

Design and methods

This study adhered to the reporting guidelines outlined in the STROBE guidelines. The overall workflow chart as illustrated in Fig. 1.

Data source

The datasets used for model training, validation, and testing are accessible through the MIMIC-IV 3.0 database (https://mimic.mit.edu/). Established in 2003 with finan cial support from the National Institutes of Health(NIH), the database represents a joint endeavor spearheaded by the Computational Physiology Laboratory at the Massachusetts Institute of Technology in partnership with Beth Israel Deaconess Medical Center (BIDMC), an affiliate of Harvard Medical School, and Philips Medical Systems [8]. The MIMIC IV database collected clinical data from more than 190,000 patients and 450,000 hospital admissions at BIDMC from 2008 to 2019. The author obtained access to the database (certificate number: 12228131).

To externally validate the prediction model, medical records of 129 patients with ICH treated at the Second Hospital of Lanzhou University (Lanzhou, China) from January 2022 to January 2024 were retrospectively analyzed. The study protocol was approved by the institutional review boards of the Second Hospital of Lanzhou University (Approval Number: [2024 A-1393]), which waived the need to obtain patient informed consent.

Inclusion criteria and exclusion criteria.

A total of 34 variables (Table S1) with a total of 11,178 data were collected in this study. These data correspond to patients who were admitted for the first time with an ICH. Consistent with previous studies [9], diagnoses in this study, including death, type 2 diabetes, pulmonary embolism, atrial fibrillation, lower extremity venous thrombosis, and hypertension, were based on the International Classification of Diseases (ICD-10) codes, as detailed in Additional Table S2.

Data cleaning

After excluding variables with missing data rates greater than 30%, 22 variables were retained for analysis, encompassing a total of 1274 cases (Figure S1). The proportions

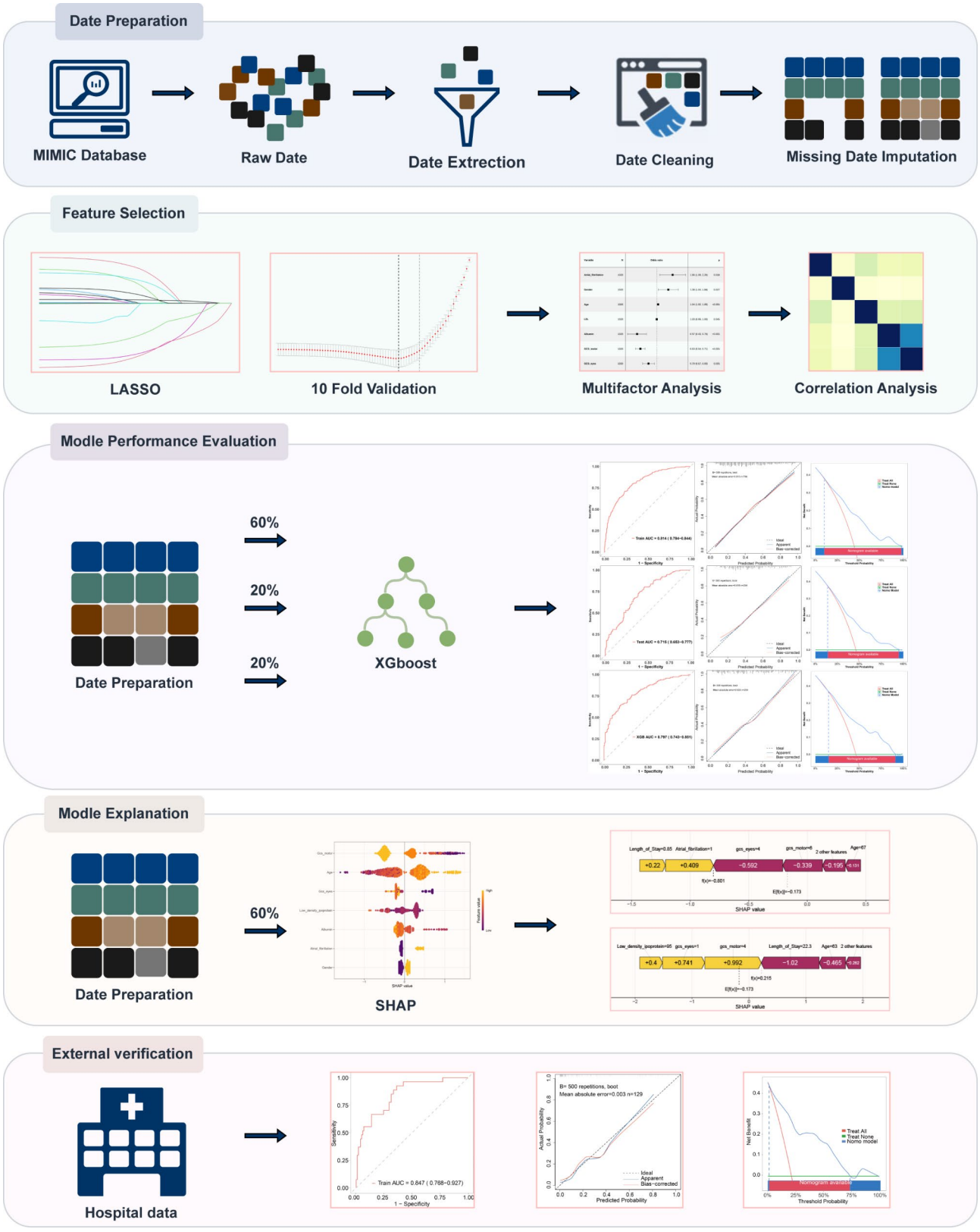

Fig. 1 Workflow diagram for the study

of missing data for these retained variables are summarized in Table S3.

Missing data imputation

All absent data points were filled using predictive mean matching methods, with five imputations carried out using the Multivariate Imputation via Chained Equations (MICE) function in R [10]. The data were then split into three subsets (training, validation, and testing) in a 6:2:2 ratio based on the distribution of outcomes. The baseline characteristics of the three groups, including clinical characteristics and mortality rates, are summarized in Table 1. No statistically significant differences were observed among the three groups.

Feature collection and screening

We employed a meticulous strategy for selecting variables to pinpoint the most pertinent predictors for constructing the prediction model. As a first step, we applied the least absolute shrinkage and selection operator (LASSO) regression, recognized as a potent technique for high-dimensional predictions [11]. Concurrently, we ascertained the optimal λ value through ten-fold cross-validation. Due to the large number of characteristic variables in this study, using λ min (0.010) as the optimal λ value would include 22 variables in the final model, making it overly complex and potentially leading to overfitting. In contrast, selecting λ 1se (0.027) as the optimal λ value results in a more streamlined model

Table 1 Baseline Data for the training, validation, and testing datasets

	Total(*n* = 1274)	Train(*n* = 764)	Validation (*n* = 255)	Test(*n* = 255)	*p*
Age(years)	69.00 (59.00, 79.00)	69.00 (59.00,80.00)	69.00 (58.50,77.00)	70.00 (59.00,80.00)	0.582
PO2(mmHg)	136.00 (81.00, 136.00)	136.00(91.00,136.00)	136.00 (71.00,136.00)	136.00 (72.50,136.00)	0.002
Platelet(K/μL)	215.00 (171.25, 263.75)	215.00 (169.00,259.25)	213.00 (171.50,275.00)	217.00 (180.50,259.00)	0.331
ALT(IU/L)	22.00 (15.00, 33.00)	22.00 (15.00,33.00)	23.00 (15.00,33.00)	20.00 (15.00,33.00)	0.852
AST(IU/L)	27.00 (19.00, 46.00)	27.00 (19.00,45.00)	28.00 (20.00,48.00)	26.00 (19.00,40.50)	0.373
PCO2(mmHg)	40.00 (38.00, 41.00)	40.00 (38.00,41.00)	40.00 (37.50,42.00)	40.00 (37.00,42.00)	0.627
White blood cell(K/μL)	8.60 (6.70, 11.90)	8.50 (6.70,11.62)	8.50 (6.80,12.90)	9.10 (6.80,11.80)	0.232
Creatinine(mg/dL)	0.90 (0.70, 1.10)	0.90 (0.70,1.10)	0.80 (0.70,1.05)	0.90 (0.70,1.20)	0.422
Cholesterol(mg/dL)	172.00 (156.00, 180.00)	172.00 (156.00,179.00)	172.00 (161.50,178.00)	172.00 (153.00,186.50)	0.918
LDL(mg/dL)	95.00 (83.00, 100.00)	95.00 (83.00,100.00)	95.00 (84.50,99.00)	95.00 (80.00,100.00)	0.906
HDL(mg/dL)	52.00 (45.00, 53.75)	52.00 (46.00,54.00)	52.00 (45.50,52.00)	52.00 (45.00,53.50)	0.595
PH(units)	7.40 (7.39, 7.41)	7.40 (7.39,7.42)	7.40 (7.38,7.41)	7.40 (7.39,7.42)	0.419
Albumin(g/dL)	3.80 (3.60, 4.20)	3.80 (3.60,4.20)	3.80 (3.60,4.20)	3.80 (3.60,4.20)	0.822
GCS min	12.00 (9.00, 14.00)	12.00 (10.00,14.00)	11.00 (9.00,14.00)	12.00 (9.00,14.00)	0.052
GCS motor	5.00 (5.00, 6.00)	5.00 (5.00,6.00)	5.00 (5.00,6.00)	5.00 (5.00,6.00)	0.439
GCS eyes	3.00 (2.00, 3.00)	3.00 (2.00,3.00)	3.00 (2.00,3.00)	3.00 (2.00,3.00)	0.238
Death:					0.996
No	702 (55.10%)	421 (55.10%)	140 (54.90%)	141 (55.29%)	
Yes	572 (44.90%)	343 (44.90%)	115 (45.10%)	114 (44.71%)	
Type 2 diabetes:					0.205
No	935 (73.39%)	573 (75.00%)	185 (72.55%)	177 (69.41%)	
Yes	339 (26.61%)	191 (25.00%)	70 (27.45%)	78 (30.59%)	
Atrial fibrillation:					0.943
No	1040 (81.63%)	626 (81.94%)	207 (81.18%)	207 (81.18%)	
Yes	234 (18.37%)	138 (18.06%)	48 (18.82%)	48 (18.82%)	
Pulmonary embolism:					0.936
No	1194 (93.72%)	716 (93.72%)	238 (93.33%)	240 (94.12%)	
Yes	80 (6.28%)	48 (6.28%)	17 (6.67%)	15 (5.88%)	
Deep vein thrombosis:					0.261
No	1201 (94.27%)	725 (94.90%)	235 (92.16%)	241 (94.51%)	
Yes	73 (5.73%)	39 (5.10%)	20 (7.84%)	14 (5.49%)	
Gender:					0.851
Male	651 (51.10%)	386 (50.52%)	131 (51.37%)	134 (52.55%)	
Female	623 (48.90%)	378 (49.48%)	124 (48.63%)	121 (47.45%)	
Hypertension:					0.224
No	417 (32.73%)	257 (33.64%)	72 (28.24%)	88 (34.51%)	
Yes	857 (67.27%)	507 (66.36%)	183 (71.76%)	167 (65.49%)	

with 7 variables, while still maintaining good predictive performance. Therefore, λ 1se was ultimately chosen as the optimal λ value for this study. Subsequently, we conducted a multivariable logistic regression analysis, incorporating the features chosen by the LASSO regression model, to identify statistically significant predictors. Lastly, we undertook a correlation analysis to alleviate any adverse effects of multicollinearity on the model.

XGBoost machine learning

XGBoost is a non-parametric approach leveraging the training of numerous sequential decision trees, enabling it to optimize and treat a wide array of variable types and imbalanced datasets [12]. In order to build a more accurate and generalizable model, the training and test data were normalized to prevent data leakage, and grid search was used for hyperparameter tuning to improve model performance. The optimized parameters of the XGBoost algorithm are as follows: objective: binary: logistic, learning_rate: 0.3, max_depth: 4, min_child_weight: 6, reg_lambda: 0.5. To assess the model's performance, we utilized the ROC curve, calibration curve, and clinical decision curve analysis (DCA). The predicted probabilities were transformed into binary outcomes using a threshold of 0.5. Subsequently, we computed and reported accuracy, sensitivity, specificity, and other relevant metrics for both the training and validation datasets [13].

SHAP

The R package "shapviz" is dedicated to interpreting the predictions of machine learning models by providing visual explanations that are grounded in SHAP (Shapley Additive exPlanations) values [14]. SHAP values illustrate the extent to which each feature contributes to the model's predictions, either in a positive or negative direction. A feature importance plot highlights the features that most significantly affect the model's predictions, with rankings determined by the average absolute SHAP values. Additionally, the generation of a force diagram is based on two samples selected, providing a visual representation of SHAP values for single-sample prediction and interpretation.

Statistics analysis

Data visualization and statistical analysis were carried out using R version 4.3.1. In general, for continuous variables, data that are normally distributed are described by the mean ± standard deviation, while skewed data are described by the median and interquartile range (IQR). Categorical variables are expressed as frequencies (percentages). The independent samples t-test or nonparametric tests were applied to evaluate differences between groups for continuous variables, whereas the Chi-square test was used for analyzing categorical variables. In this study, the "tidyverse," "pROC," "CBCgrps," "rms," and "rmda" packages were employed for data collation and visualization.

Results

Basic characteristics

The study included 1,274 patients with cerebral hemorrhage, of which 572 died, resulting in a mortality rate of 44.9%. Some basic characteristics of patients are shown in Table 2. The following variables showed statistically significant differences between the two groups: Age ($p < 0.001$), PO2 ($p < 0.001$), platelet ($p = 0.041$), AST ($p = 0.888$), cholesterol ($p = 0.008$), low density lipoprotein(LDL) ($p = 0.002$), albumin ($p < 0.001$), GCS min ($p < 0.001$), GCS motor ($p < 0.001$), GCS eyes($p < 0.001$), type 2 diabetes ($p = 0.038$), atrial fibrillation($p = 0.003$) and gender($p = 0.007$). The comparison of the following indicators between the two groups revealed no statistical significance: ALT, PCO2, white blood cell(WBC), creatinine, high density lipoprotein(HDL), PH, pulmonary embolism, deep vein thrombosis and hypertension, among others (p-value > 0.05).

LASSO regression for feature selection

We preliminarily selected predictive factors for death in patients with intracerebral hemorrhage using LASSO regression. (Fig. 2A). Using a lambda value set to one standard deviation from the minimum lambda, where the error remains within one standard error of the minimum, we were able to identify seven variables that exhibited the highest predictive power: 'Age', 'Albumin', 'LDL', 'Atrial fibrillation', 'GCS motor', 'GCS eyes', and 'Gender' (Fig. 2B). These seven variables, which demonstrated the strongest association with the outcome variable, were selected to ensure the model's simplicity and to address concerns about overfitting. By conducting a multivariate logistic regression analysis, we confirmed that these variables were all independent risk factors for death (Fig. 2C). The Spearman correlation test was used to examine 5 continuous variables. The heatmap results showed that there is no significant correlation between the variables, indicating that there is no multicollinearity (Fig. 2D).

XGBoost ML model performance evaluation

The XGBoost model exhibits superior AUC performance (train: 0.814, 95%CI: 0.784 – 0.844; validation: 0.715, 95%CI: 0.653 – 0.777; test: 0.797, 95%CI: 0.743 – 0.851) in predicting mortality among ICH patients (Fig. 3A, D, G). The model shows good generalization ability, as its performance on the training and testing datasets is consistent, suggesting strong adaptability to new data. Additionally, the F1 scores for the training and testing datasets

Table 2 Baseline characteristics of the study population

Level	Total($n=1274$)	Non-Death($n=702$)	Death($n=572$)	*p*
Age(years)	69.00(59.00;79.00)	66.00 (55.00;76.00)	74.00 (64.00;82.25)	<0.001
PO2(mmHg)	136.00 (81.00;136.00)	136.00 (113.25;136.00)	136.00 (66.75;142.25)	<0.001
Platelet(K/μL)	215.00 (171.25;263.75)	218.00 (181.00;260.50)	210.00 (162.75;267.00)	0.041
ALT(IU/L)	22.00 (15.00;33.00)	22.00 (15.00;33.00)	22.00 (15.00;33.00)	0.888
AST(IU/L)	27.00 (9.00;46.00)	25.00 (18.00;41.00)	29.50 (21.00;48.00)	<0.001
PCO2(mmHg)	40.00 (8.00;41.00)	40.00 (40.00;40.00)	40.00 (36.00;44.00)	0.404
WBC(K/μL)	8.60 (6.70;11.90)	8.50 (6.70;11.50)	8.70 (6.70;12.30)	0.389
Creatinine(mg/dL)	0.90 (0.70;1.10)	0.90 (0.70;1.10)	0.90 (0.70;1.20)	0.076
Cholesterol(mg/dL)	172.00 (156.00;180.00)	172.00 (155.25;191.00)	172.00 (159.00;172.00)	0.008
LDL(mg/dL)	95.00 (83.00;100.00)	95.00 (83.00;110.00)	95.00 (83.00;95.00)	0.002
HDL(mg/dL)	52.00 (45.00;53.75)	52.00 (43.00;56.00)	52.00 (49.00;52.00)	0.460
PH(units)	7.40 (7.39;7.41)	7.40 (7.40;7.40)	7.40 (7.36;7.42)	0.051
Albumin(g/dL)	3.80 (3.60;4.20)	3.90 (3.70;4.30)	3.80 (3.40;4.10)	<0.001
GCS min	12.00 (9.00;14.00)	13.00 (11.00;14.00)	11.00 (8.00;14.00)	<0.001
GCS motor	5.00 (5.00;6.00)	6.00 (5.00;6.00)	5.00 (4.00;6.00)	<0.001
GCS eyes	2.663 (2.605,2.722)	2.964 (2.897,3.032)	2.294 (2.201,2.386)	<0.001
Type 2 diabetes (%)				0.038
No	935 (73.39%)	532 (75.78%)	403 (70.45%)	
Yes	339 (26.61%)	170 (24.22%)	169 (29.55%)	
Atrial fibrillation (%)				0.003
No	1040 (81.63%)	594 (84.62%)	446 (77.97%)	
Yes	234 (18.37%)	108 (15.38%)	126 (22.03%)	
Pulmonary embolism (%)				0.406
No	1194 (93.72%)	662 (94.30%)	532 (93.01%)	
Yes	80 (6.28%)	40 (5.70%)	40 (6.99%)	
Deep vein thrombosis (%)				0.676
No	1201 (94.27%)	664 (94.59%)	537 (93.88%)	
Yes	73 (5.73%)	38 (5.41%)	35 (6.12%)	
Gender (%)				0.007
Male	651 (51.10%)	383 (54.56%)	268 (46.85%)	
Female	623 (48.90%)	319 (45.44%)	304 (53.15%)	
Hypertension (%)				0.878
No	417 (32.73%)	228 (32.48%)	189 (33.04%)	
Yes	857 (67.27%)	474 (67.52%)	383 (66.96%)	

Describe the statistical results table. This results table counts the mean of each continuous variable and its 95% confidence interval, as well as the frequency and percentage of categorical variables. The p-values are derived from the results of comparing the means of two groups (t-test), or comparing the means of more than two groups (ANOVA)

are similar, indicating the model's balanced performance in terms of sensitivity and precision. Moreover, specificity remains stable across all three datasets, particularly in the training and validation datasets, where the model effectively identifies negative cases. (Table 3). The definitions of all the performance metrics used in this study, along with their respective calculation methods, are provided in the supplementary statement.

We analyzed calibration curves and clinical decision curves from the training, validation, and testing datasets to assess the accuracy of the XGBoost model in predicting the risk of death in patients with ICH. By utilizing the Bootstrap resampling method to assess the XGBoost model 500 times, we observed that the model's calibration curve deviates only slightly from the ideal linear relationship, indicating a high consistency between model predictions and observed outcomes (Fig. 3B, E and H). Additionally, we used decision curve analysis. The X-axis represents the threshold probability for predicting death risk, while the Y-axis indicates the net benefit. The blue line reflects the performance of the XGBoost model across the three datasets, showing its predictive improvement. The red line represents the scenario where all patients are treated based on the XGBoost model, and the green line represents the assumption that no patients are treated based on the XGBoost model [15]. Our study demonstrated a wide threshold range for net benefit across all three datasets, indicating the clinical utility of the XGBoost model in decision-making. (Fig. 3C, F and I).

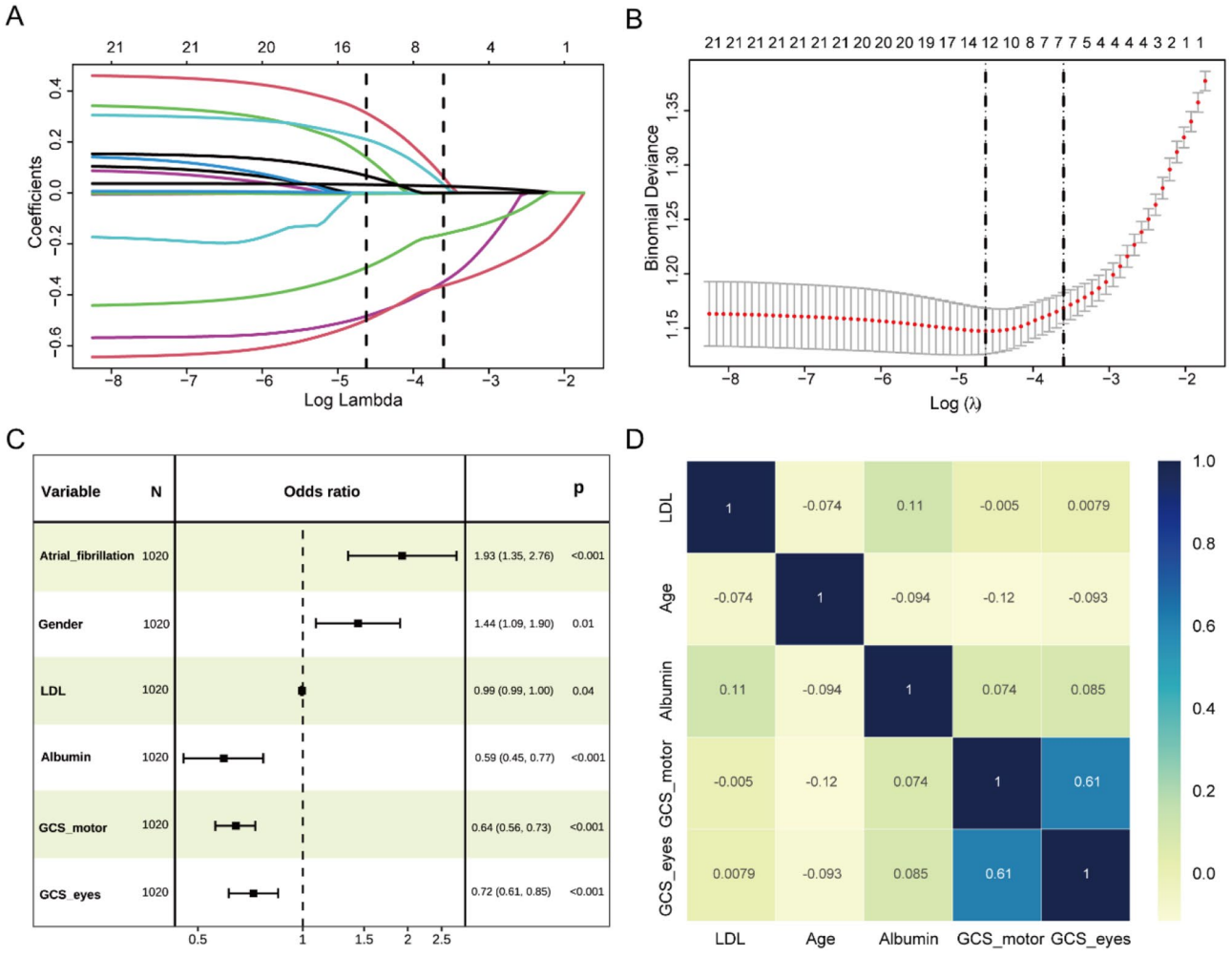

Fig. 2 **A**. Lasso coefficient path plots for 22 variables. **B**. Cross-validation curves (10-fold cross-validation). **C**. Forest plot displaying adjusted odds ratios (ORs) with 95% confidence intervals (CIs) for factors associated with the outcome, as determined by multivariate logistic regression analysis. **D**. Correlation heat map of continuous variables

Interpretation of the XGBoost ML model

The SHAP beeswarm plot reveals how individual features influence the model's predictions. (Fig. 4A). The features included are ranked from most to least important: 'GCS motor', 'Age', 'GCS eyes', ' LDL', ' Albumin', ' Atrial fibrillation', and 'Gender', respectively. In the predictive model, a greater SHAP value for a feature indicates a higher probability of mortality. To illustrate the XGBoost model's assessment of a single observation features' contributions, we present an interpretation of the prediction based on the SHAP model for both cases. The color coding reflects the impact of each feature on the prediction: purple signifies a detrimental influence on the forecast (with an arrow to the left indicating a decrease in the SHAP value), while yellow denotes a beneficial effect (with an arrow to the right indicating an increase in the SHAP value). The color bars length signifies the magnitude of the contribution, while E[f(x)] represents the SHAP reference value, which corresponds to the model's mean prediction. For the 'true positive' patient group, the XGBoost model forecasted mortality with a SHAP value of -0.261, surpassing the baseline and suggesting a heightened likelihood of death, as depicted in Fig. 4B. In contrast, the 'true negative' patient group had a SHAP value of -0.794, which was below the reference threshold, suggesting no occurrence of death, as depicted in Fig. 4C.

External validation of the model

The comparison of model-relevant variable data for the external independent patient cohort is shown in Table S4. The AUC for external validation is 0.847 (95% CI: 0.768–0.927), which is comparable to the training dataset (0.814) and superior to the internal validation dataset (0.715), indicating high discriminative ability. The calibration curve demonstrated excellent agreement between the predicted probabilities and observed outcomes, with a mean absolute error of 0.003, further supporting the reliability of the predictions. The DCA

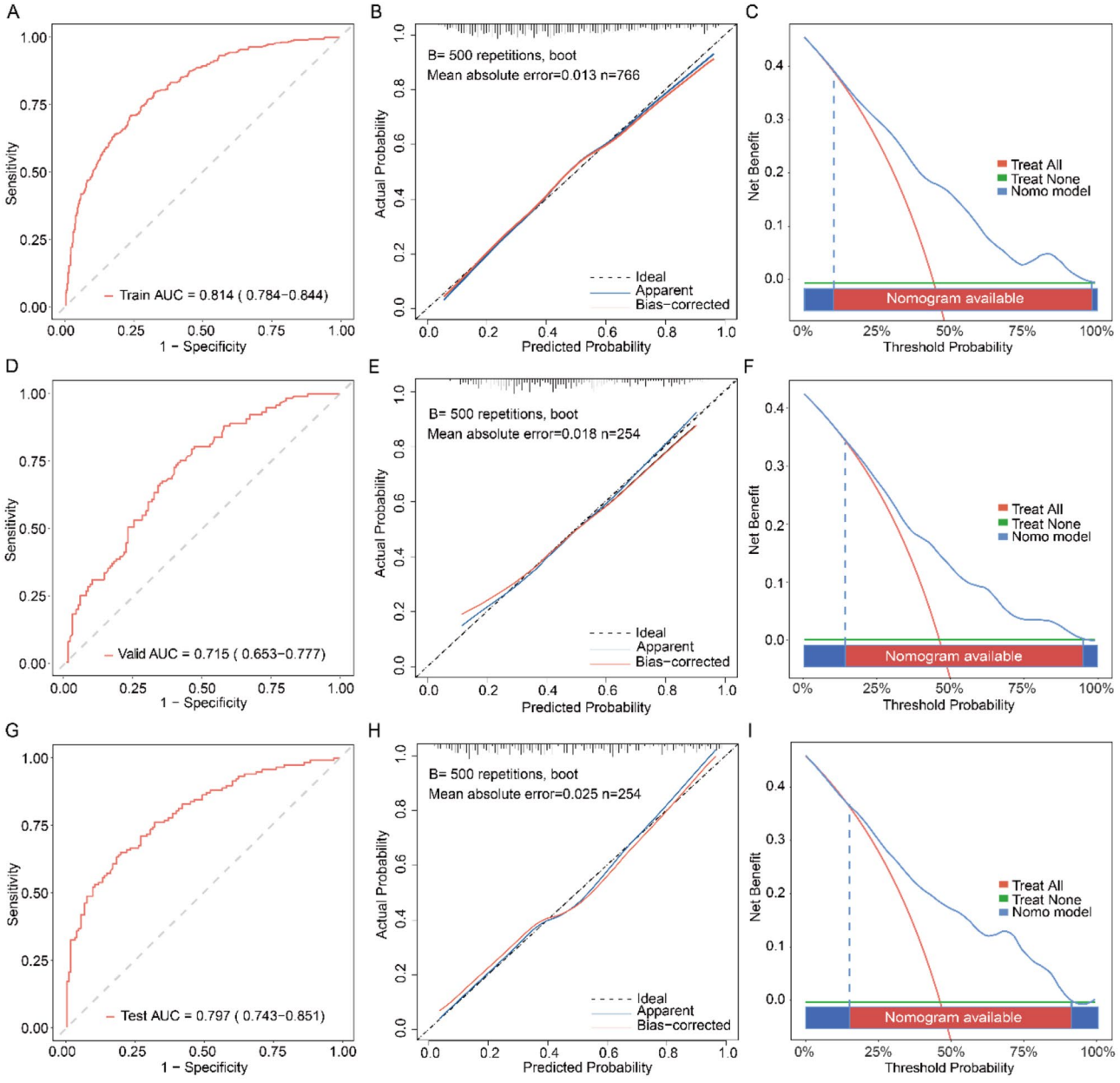

Fig. 3 (**A**-**C**) ROC curve, calibration curve and clinical decision curves of the model on the training set. (**D**-**F**) ROC curve, calibration curve, and clinical decision curves of the model on the validation set. (**G**-**I**) ROC curve, calibration curve, and clinical decision curves of the model on the testing set

Table 3 Performance Metrics of the XGBoost Model on the training, validation, and Testing datasets

	Training set	Validation set	Testing set
Accuracy	0.7402(0.7076, 0.7709)	0.6378(0.5754, 0.697)	0.689 (0.6281, 0.7454)
Sensitivity	0.7071	0.5470	0.7949
Specificity	0.7664	0.7153	0.5985
Precision	0.7050	0.6214	0.6284
F1 score	0.7061	0.5818	0.7019

for external validation shows that the model effectively predicts the net benefit for ICH mortality risk across a threshold probability range of 1–73%. (Fig. 5)

Discussion

In our study, we employed the SHAP-interpretable XGBoost model as a novel approach to predict mortality risk in ICH patients. The model's performance was assessed using both the training and validation datasets. Evaluation metrics, including the ROC curve, calibration curve, and others, demonstrated the model's high predictive accuracy. Furthermore, we delved into the

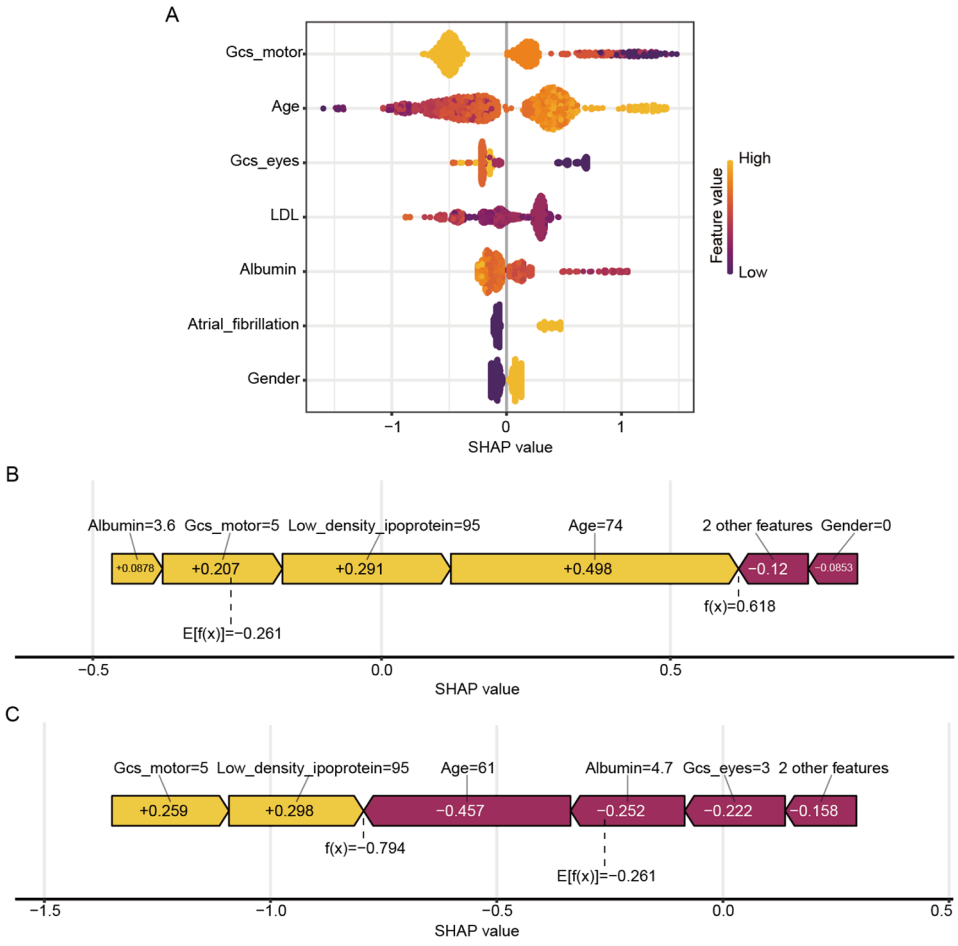

Fig. 4 **A**. SHAP variable importance chart, displaying the included features arranged in descending order based on their average absolute SHAP values. **B** and **C**. SHAP force plots for two cases: Each feature's contribution is denoted by color, with purple signifying a negative impact on the prediction (indicated by an arrow pointing left, resulting in a decrease in SHAP value), and yellow signifying a positive impact (indicated by an arrow pointing right, increasing in SHAP value). The length of the color bar represents the magnitude of the contribution, while E[f(x)] corresponds to the SHAP reference value, which is the model's mean prediction. f(x) denotes the individual SHAP value

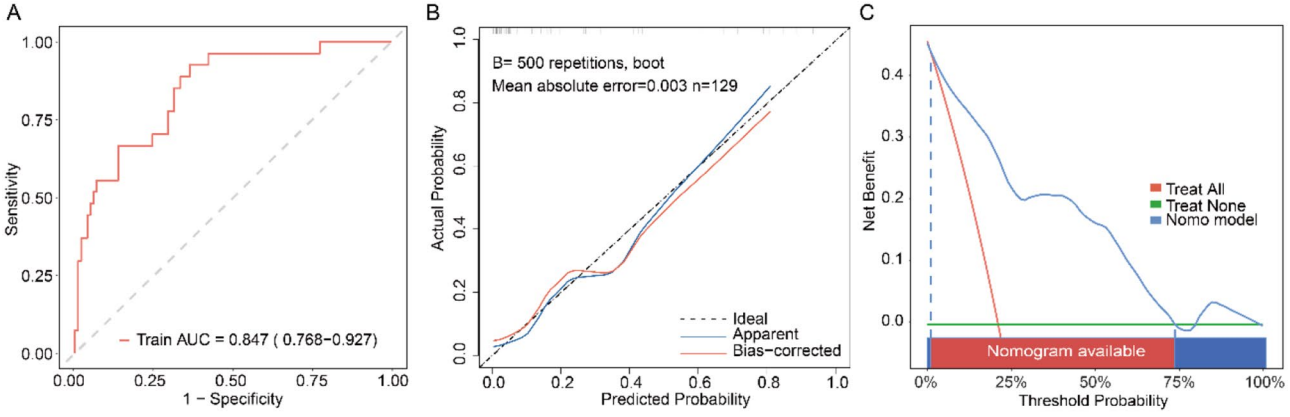

Fig. 5 (**A**-**C**) ROC, calibration, and clinical decision curve analysis for external validation

interpretation of the XGBoost model using SHAP. The SHAP analysis revealed that the key features, in order of importance, are 'GCS motor', 'Age', 'GCS eyes', 'LDL', ' Albumin', ' Atrial fibrillation', and 'Gender'.

Several studies have reported a potential association between age and mortality subsequent to ICH. Daniel et al. found that in their analysis of ICH patients from diverse racial and ethnic backgrounds, advanced age was strongly associated with higher post-ICH mortality rates [16]. A global meta-analysis on the incidence and mortality of cerebral hemorrhage similarly showed that the incidence increases with age, particularly after 85 years old [16]. Consistent with these findings, our analysis also reveals that in ICH patients, higher SHAP values for age are linked to increased mortality rates (Fig. 4A). Furthermore, there is literature to explore the impact of gender differences on prognosis after ICH, and the mortality rate of men with ICH is higher than that of women, especially in the early stages after ICH [17].

The Glasgow Coma Scale (GCS) was created to standardize the evaluation of neurologically compromised patients, aiding in the triage of injury severity and guiding management decisions for personalized care [18]. GCS motor scores have been found to be significant predictors of survival and neurological recovery, while lower GCS eye scores are associated with higher mortality rates and poorer prognosis [19]. It can be intuitively found in our SHAP value graph that the lower the score of GCS motor and GCS eyes, pushing the predicted value towards positive, may imply a higher risk of death (Fig. 4A). Healthcare Professionals should standardize the GCS evaluation criteria. Patients with low GCS motor and GCS eye scores require close monitoring, and their level of care should be enhanced accordingly.

LDL, a lipoprotein found in the blood, has been implicated in vascular health. Research suggests that excessively low levels of LDL may compromise the structural integrity of blood vessel walls, increasing the risk of ICH [20]. Albumin, a crucial plasma protein primarily synthesized by the liver, exerts antioxidant and anti-inflammatory effects. These properties contribute to the mitigation of the inflammatory response following cerebral hemorrhage and safeguard brain tissue from oxidative stress-induced damage. Insufficient albumin levels can worsen vascular leakage and cerebral edema, thereby increasing the risk of mortality [21, 22]. Thus, it is crucial to carefully select appropriate treatments for high-risk patients and implement dietary conditioning programs. Additionally, patients with atrial fibrillation receiving anticoagulant therapy—such as warfarin or apixaban—may have an elevated risk of cerebral hemorrhage and mortality [23–25].

The use of AI in healthcare raises critical ethical concerns, including data privacy, algorithmic transparency, fairness, and bias. To address these concerns, robust data protection measures must be implemented to minimize the risks of data leakage and misuse. In parallel, enhancing the transparency and interpretability of AI algorithms is essential for building trust among healthcare professionals and patients, ensuring that decisions are both fair and understandable [26, 27]. To further ensure fairness, diverse datasets and fairness assessments should be employed to prevent bias and guarantee equitable outcomes across all demographics [28]. Ultimately, public trust in AI applications hinges on transparent communication and strong data privacy protections, both of which significantly influence acceptance [29]. Moreover, the level of public trust in AI systems is closely related to the level of transparency and data privacy protection of AI systems, improving public understanding of AI technology is an effective way to enhance trust, and educating the public and addressing cultural differences in AI perception is critical to promoting widespread acceptance of AI technology in healthcare [30].

We utilized LASSO regression and multivariable logistic regression analysis to screen variables efficiently while simultaneously reducing their number, this approach was adopted. It improved clinical applicability and effectively mitigated the risk of model overfitting. The seven parameters included in our model are all common clinical indicators, with advantages of accessibility and convenience, which contribute to clinical application and popularization of the model. Our machine-learning-developed model serves as a screening tool for the identification of high-risk patients and provides information that aids clinical decision-making. Lastly, the model developed in this study undergoes external validation, wherein the results of said validation demonstrate its commendable discriminative capabilities and practical utility in forecasting death.

This study is not without its limitations. Firstly, this was a retrospective study, selection bias was inevitable. Further prospective studies are required to improve the level of evidence to support our findings. Secondly, it is still a challenge to interpret machine learning models, although SHAP enhances model interpretability [31]. The decision-making process of the model may still be unclear. Consequently, our next step involves integrating predictive models into digital health record systems to forecast individual patient cases by extracting patient information and indicators. Furthermore, we aim to present the forecasted outcomes directly to all users, thereby enhancing the practical applicability of the model. Although the SHAP method has been used to visually illustrate the relative importance of features, a web- or portable electronic equipment-based user-friendly program integrating our developed predictive model should be designed and produced to improve the chances of early detection

of the risk of death in ICH patients. Finally, we only consider the most commonly used XGBoost machine learning methods, some of the more advanced methods may show higher predictive power [31, 32]. We encourage collaboration from teams at different centers to provide additional multi-center data, thereby expanding the scope and enhancing the robustness of our research.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12911-025-02865-4.

Supplementary Material 1

Supplementary Material 2

Acknowledgements
Thanks to Jianghan University and Hubei NO. 3 People's Hospital for supporting this study.

Author contributions
Y. C. wrote the main manuscript text and prepared all the graphs and tables. X. L. processed the data. M. Y. was responsible for the experimental design. Y. W. supervised and guided the whole process. All authors reviewed the manuscript.

Funding
The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by grants from the National Natural Science Foundation of China (No. 82301439), the Natural Science Foundation of Hubei Provincial (Nos. 2022CFB870 and 2022CFB299), and a grant from the Health. Commission of Hubei Province Scientific Research Project (No. WJ2023M114).

Data availability
The data that support the findings of this study are available from MIMIC-IV. Access to the database can be obtained through PhysioNet at https://mimic.mit.edu/. Researchers wishing to access the MIMIC-IV database must complete the required training and agree to the data use agreement available on the PhysioNet website.

Declarations

Ethics approval and consent to participate
The MIMIC-IV database was approved by the Massachusetts Institute of Technology and Beth Israel Deaconess Medical Center, and consent was obtained for the original data collection. Additionally, the institutional review boards of the Second Hospital of Lanzhou University also approved our study and waived the need for informed consent due to the retrospective nature of this study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 28 October 2024 / Accepted: 13 January 2025
Published online: 22 January 2025

References

1. Qureshi AI, Mendelow AD, Hanley DF. Intracerebral haemorrhage. Lancet. 2009;373:1632–44.
2. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, et al. Heart Disease and Stroke Statistics-2020 update: a Report from the American Heart Association. Circulation. 2020;141:e139–596.
3. An SJ, Kim TJ, Yoon B-W, Epidemiology. Risk factors, and clinical features of Intracerebral Hemorrhage: an update. J Stroke. 2017;19:3–10.
4. Van Asch CJ, Luitse MJ, Rinkel GJ, Van Der Tweel I, Algra A, Klijn CJ. Incidence, case fatality, and functional outcome of intracerebral haemorrhage over time, according to age, sex, and ethnic origin: a systematic review and meta-analysis. Lancet Neurol. 2010;9:167–76.
5. Hale AT, Stonko DP, Brown A, Lim J, Voce DJ, Gannon SR et al. Machine-learning analysis outperforms conventional statistical models and CT classification systems in predicting 6-month outcomes in pediatric patients sustaining traumatic brain injury. 2018. https://doi.org/10.3171/2018.8.FOCUS17773
6. Wu E, Marthi S, Asaad WF. Predictors of mortality in traumatic intracranial hemorrhage: a National Trauma Data Bank Study. Front Neurol. 2020;11:587587.
7. Mortality Prediction in Cerebral Hemorrhage Patients Using Machine Learning Algorithms in Intensive Care Units. - PubMed. https://pubmed.ncbi.nlm.nih.gov/33551969/. Accessed 28 Aug 2024.
8. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data. 2023;10:1.
9. Wei X, Chen X, Zhang Z, Wei J, Hu B, Long N, et al. Risk analysis of the association between different hemoglobin glycation index and poor prognosis in critical patients with coronary heart disease-A study based on the MIMIC-IV database. Cardiovasc Diabetol. 2024;23:113.
10. Qi L, Wang Y-F, Chen R, Siddique J, Robbins J, He Y. Strategies for imputing missing covariates in accelerated failure time models. Stat Med. 2018;37:3417–36.
11. Li L, Tu B, Xiong Y, Hu Z, Zhang Z, Liu S, et al. Machine learning-based Model for Predicting prolonged mechanical ventilation in patients with congestive heart failure. Cardiovasc Drugs Ther. 2024;38:359–69.
12. Duckworth C, Guy MJ, Kumaran A, O'Kane AA, Ayobi A, Chapman A, et al. Explainable machine learning for real-time hypoglycemia and hyperglycemia prediction and Personalized Control recommendations. J Diabetes Sci Technol. 2024;18:113–23.
13. Ershadi MM, Rise ZR. Fusing clinical and image data for detecting the severity level of hospitalized symptomatic COVID-19 patients using hierarchical model. Res Biomed Eng. 2023;39:209–32.
14. Bifarin OO. Interpretable machine learning with tree-based shapley additive explanations: application to metabolomics datasets for binary classification. PLoS ONE. 2023;18:e0284315.
15. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and interpreting decision curve analysis: a guide for investigators. Eur Urol. 2018;74:796–804.
16. Frontiers.| Epidemiology of intracerebral hemorrhage: A systematic review and meta-analysis. https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2022.915813/full. Accessed 4 Sep 2024.
17. James ML, Cox M, Xian Y, Smith EE, Bhatt DL, Schulte PJ et al. Sex and Age Interactions and Differences in Outcomes After Intracerebral Hemorrhage. J Womens Health. 2002. 2017;26:380–8.
18. Maserati M, Fetzick A, Puccio A. The Glasgow Coma Scale (GCS): deciphering the Motor Component of the GCS. J Neurosci Nurs. 2016;8:311.
19. Dey A, Ghosh S, Bhuniya T, Koley M, Bera A, Guha S, et al. Clinical theragnostic signature of Extracellular vesicles in Traumatic Brain Injury (TBI). ACS Chem Neurosci. 2023;14:2981–94.
20. Abrantes CS, Pintalhão M, Tavares S, Fonseca L, Chaves PC. Anticoagulation after intracerebral hemorrhage in patients with atrial fibrillation: between Scylla and Charybdis. Neurol Sci. 2022;43:2441–8.
21. Peng Q, Hou J, Wang S, Zhou F, Wang EY. Hypersensitive C-reactive protein-albumin ratio predicts symptomatic intracranial hemorrhage after endovascular therapy in acute ischemic stroke patients. BMC Neurol. 2021;21:47.
22. He J, Zhang Y, Li T, Deng H, Wang P, Chong W, et al. Glucose-albumin ratio as new biomarker for predicting mortality after intracerebral hemorrhage. Neurosurg Rev. 2023;46:94.
23. Lopes RD, Guimarães PO, Kolls BJ, Wojdyla DM, Bushnell CD, Hanna M, et al. Intracranial hemorrhage in patients with atrial fibrillation receiving anticoagulation therapy. Blood. 2017;129:2980–7.
24. Zhao B, Yuan Y, Li Z, Chen Y, Gao Y, Yang B et al. Risk of intracranial hemorrhage in patients using anticoagulant therapy for atrial fibrillation after cerebral microbleeds combined with acute ischemic stroke: a meta-analysis. Front Neurol. 2024;15.

25. Pham HN, Sainbayar E, Ibrahim R, Lee JZ. Intracerebral hemorrhage mortality in individuals with atrial fibrillation: a nationwide analysis of mortality trends in the United States. J Interv Card Electrophysiol. 2024;67:1117–25.
26. Razai MS, Al-Bedaery R, Bowen L, Yahia R, Chandrasekaran L, Oakeshott P. Implementation challenges of artificial intelligence (AI) in primary care: perspectives of general practitioners in London UK. PLoS ONE. 2024;19:e0314196.
27. Teng Z, Li L, Xin Z, Xiang D, Huang J, Zhou H, et al. A literature review of artificial intelligence (AI) for medical image segmentation: from AI and explainable AI to trustworthy AI. Quant Imaging Med Surg. 2024;14:9620–52.
28. Chen F, Wang L, Hong J, Jiang J, Zhou L. Unmasking bias in artificial intelligence: a systematic review of bias detection and mitigation strategies in electronic health record-based models. J Am Med Inf Assoc JAMIA. 2024;31:1172–83.
29. Grzybowski A, Jin K, Wu H. Challenges of artificial intelligence in medicine and dermatology. Clin Dermatol. 2024;42:210–5.
30. Williams M, Karim W, Gelman J, Raza M. Ethical data acquisition for LLMs and AI algorithms in healthcare. NPJ Digit Med. 2024;7:377.
31. Rainio O, Teuho J, Klén R. Evaluation metrics and statistical tests for machine learning. Sci Rep. 2024;14:6086.
32. Fazai R, Abodayeh K, Mansouri M, Trabelsi M, Nounou H, Nounou M, et al. Machine learning-based statistical testing hypothesis for fault detection in photovoltaic systems. Sol Energy. 2019;190:405–13.

Publisher's note