RESEARCH

Open Access



Developing a high-performance AI model for spontaneous intracerebral hemorrhage mortality prediction using machine learning in ICU settings

Xiao-Han Vivian Yap^{1†}, Kuan-Chi Tu^{1†}, Nai-Ching Chen², Che-Chuan Wang¹, Chia-Jung Chen³, Chung-Feng Liu⁴, Tee-Tau Eric Nya^{1,5*†} and Ching-Lung Kuo^{1,4,6*†}

Abstract

Background Spontaneous intracerebral hemorrhage (SICH) is a devastating condition that significantly contributes to high mortality rates. This study aims to construct a mortality prediction model for patients with SICH using four various artificial intelligence (AI) machine learning algorithms.

Method A retrospective analysis was conducted on electronic medical records of SICH patients aged 20 and above, admitted to Chi Mei Medical Center's intensive care unit between January 2016 and December 2021. The study utilized 37 features related to mortality. Predictive models were developed using logistic regression, Random forest, LightGBM, XGBoost, and Multi-layer Perceptron (MLP), with assessments of feature importance, and Area under the curve (AUC).

Results A total of 1451 SICH patients were enrolled. Factors associated with mortality included lower initial GCS scores (p < 0.001), pupillary changes (P < 0.001), kidney disease (p < 0.001), and respiratory failure requiring intubation (p < 0.001). Negative correlations were observed between mortality and pupil light reflexes, as well as GCS components E(r=-0.4602), V (r=-0.4132), M(r=-0.4082). Positive correlations were identified with vasopressors (r=0.4464), FiO2 (r=0.3901), and sedative-hypnotic drugs (r=0.1178). XGBoost demonstrated the best predictive performance (AUC = 0.913), outperforming LR (0.899), RF (0.905), LightGBM (0.909), and MLP (0.892). The XGBoost model, utilizing both 18 and 36 features, continues to outperform both the Acute Physiology and Chronic Health Evaluation (APACHE II) (p < 0.001) and Sequential Organ Failure Assessment (SOFA) scoring systems (p < 0.001).

[†]Xiao-Han Vivian Yap, Kuan-Chi Tu, Tee-Tau Eric Nyam and Ching-Lung Kuo contributed equally to this work.

*Correspondence: Tee-Tau Eric Nya ronaldowen@gmail.com Ching-Lung Kuo kuojinnrung@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Conclusion This study successfully developed an AI mortality prediction model for SICH patients, with XGBoost exhibiting superior performance. The model, incorporating 18 key features, has been integrated into clinical practice assisting clinicians in treatment decisions and communication with patients' families.

Keywords Artificial intelligence, Spontaneous intracrebral hemorrhage, Mortality, Intensive care unit

Introduction

Spontaneous intracerebral hemorrhage (SICH) is a devastating condition, with early-term mortality ranging from 30 to 40%, and there has been minimal improvement over recent years [1]. Stroke is a leading cause of long-term disability in the United States, with approximately 10% of the 795,000 strokes per year being SICH [2]. 26% of individuals remain disabled in basic activities of daily living, and 50% experience reduced mobility due to hemiparesis [3].

Consequently, SICH remains a significant public health concern, affecting not only the well-being of patients but also imposing a substantial burden on social, economic, and healthcare resources [3].³ Therefore, developing an accurate method for predicting prognosis can be genuinely beneficial in clinical practice. This information can assist physicians in deciding whether to pursue conservative or aggressive treatment for the patient.

Several prognostic tools for predicting mortality in SICH have been proposed, encompassing factors such as age [4, 5], gender [6], blood pressure [7], initial Glasgow Coma Scale (GCS) [8, 9], pupillary changes [9, 10], mechanical ventilation requirement [11], and underlying comorbidities such as cardiovascular and cerebrovascular diseases [12]. Moreover, the Acute Physiology and Chronic Health Evaluation II (APACHE II) system [13, 14] and the Sequential Organ Failure Assessment (SOFA) score [15] are widely utilized disease classification systems for predicting mortality and severity of failed organs the Intensive Care Unit (ICU). The consideration of whether there are new predictive models that can assist or potentially replace these existing tools is of significant importance.

Machine learning, a form of artificial intelligence (AI) that learns patterns and rules from given information, offers advantages in detecting possible interactions among many attributes, making it useful in clinical prediction and identifying novel prognostic markers [16, 17]. Recent studies have applied machine learning to severity or outcome prediction models for neurological disorders, such as ischemic stroke [18], aneurysmal subarachnoid hemorrhage [19], and traumatic brain injury [20]. However, its application in predicting mortality after spontaneous ICH is still relatively rare [21–23]. Therefore, the development of new AI prognostic prediction models is worth pursuing.

Machine learning delivers precise predictions in complex scenarios [24]. Nevertheless, the "black-box" nature of AI, marked by a lack of explanation, remains a primary hindrance to its widespread clinical application. Explanatory AI (XAI), such as SHAP (SHapley Additive exPlanations), proves vital in comprehending essential clinical features for predicting diseases or patient outcomes [25]. As the most widely used XAI technique, SHAP is crucial for interpreting AI models.

Hospitals have recently started applying statistical and AI models with various algorithms, including logistic regression (LR) [26], random forest [27], Light Gradient Boosting Machine (LightGBM) [28], Extreme Gradient Boosting (XGBoost) [29], and Multi-layer Perceptron (MLP) [30]. Logistic Regression serves as a baseline model with high interpretability, while advanced machine learning methods like Random Forest, XGBoost, and LightGBM, based on ensemble decision tree algorithms, effectively capture non-linear relationships and interactions, often achieving superior performance in highdimensional and imbalanced clinical datasets. MLP introduces a neural network perspective, capable of modeling complex patterns within the data.

Integrating big-data-driven approaches and machine learning into our hospital information system (HIS), a real-time prediction system was developed for patients with traumatic brain injuries to prognosticate early mortality risk [31]. However, AI algorithms prediction in spontaneous ICH has not well established.

In this study, we hypothesize initial clinical parameters for predicting outcomes in SICU-admitted patients, using easily obtainable data. To achieve this, we employ machine-learning algorithms to analyze a vast amount of SICU data, predicting mortality risk after spontaneous ICH. Additionally, we compare four machine learning models with the existing APACHE II and SOFA scores. Furthermore, we utilize the SHAP technique to explain which clinical features are crucial for predicting mortality.

Method

Ethics

The Chi Mei Medical Center's Institutional Review Board granted ethics approval (11107-012) for this study. All procedures were conducted by the authors in compliance with applicable laws and regulations. Due to the retrospective nature of the study, the ethics committee decided to waive the requirement for informed consent.

Flow chart and the prediction device content of the current study

Our investigation adhered to the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) guidelines (Supplemental Table 1). Figure 1 depicts the flowchart illustrating the integration of the AI prediction model for SICH patients in the ICU, utilizing 36 feature variables for training. Various models, including logistic regression (LR), random forest (RF), LightGBM, XGBoost, and Multi-layer Perceptron (MLP) were trained on 70% of the data and validated on a 30% test set through random splitting. To mitigate concerns of overfitting that might arise from a small dataset, we employed the 5-fold cross-validation technique to build the models.

To address the imbalance in the dataset, characterized by more negative cases (survival) than positive cases (mortality), we applied the Synthetic Minority Oversampling Technique (SMOTE) [32] to achieve equal representation during the final model training with each algorithm. Figure 2 illustrates our AI prediction device for SICH in the ICU, providing insight into the system's architecture and modules.

Patient selection

This study retrospectively enrolled 1451 patients aged 20 and above with spontaneous SICH. These patients were admitted to the ICU at Chi Mei Medical Center in Tainan, Taiwan, between January 2016 and December

2021. The electronic medical records were screened, and those containing the following diagnostic codes were included: [ICD-9] ICD-10, ICD-9: 431*, 432.9*; ICD-10: I61.0 - I62.9, indicating SICH, unspecified. Due to low missing rates (<30%) for features, data with missing or ambiguous values were excluded.

Features selection and model building

The initial 36 features were selected by experts' opinions based on their knowledge of the subject matter, previous research findings, and clinical relevance [4-16]. These 36 features were collected from medical records at the time of patient admission to the intensive care unit, with the endpoint being the discharge status. These features include age, gender, height, weight, systolic blood pressure (SBP), diastolic blood pressure (DBP), body temperature (BT), pulse, and respiratory rate (RR). Neurological indicators consisted of Glasgow Coma Scale (GCS) components-eyes open, verbal response, and motor response-along with pupil reflex and size (right and left) and muscle power of all four extremities (Muscle LUE, Muscle LEE, Muscle RUE, Muscle RLE). Additional features encompassed the Inspired Fraction of Oxygen (FiO2), presence of an endotracheal tube (Endo), external ventricular drain (EVD), and intracranial pressure (ICP). Medical histories, including hypertension, diabetes mellitus, heart disease, cerebrovascular disease, gastrointestinal disease, liver disease, kidney disease, and cancer, were also included. Furthermore, variables related to



Fig. 1 Schematic diagram of the study's workflow. ICU, intensive care unit; SICH, intracerebral hemorrhage; SMOTE, Synthetic Minority Oversampling Technique; HIS, hospital information system



Fig. 2 Artificial intelligence prediction device of SICH in the intensive care unit. GCS, Glascow Coma Scale

interventions or treatments, such as vasopressors, sedative/hypnotic drugs, and nicardipine, were considered. We excluded patients with missing or obviously erroneous values for these feature variables.

To identify the correlation between 36 features and mortality, we use the Spearman's correlation coefficient methods [33] alongside SHAP (SHapley Additive Explanations) analysis.

Model performance measurement

In this study, we assessed the performance of the machine learning models using accuracy [34], sensitivity, and specificity [35], F1-score [36], as well as the Area under the Curve (AUC) of the Receiver Operating Characteristic curve (ROC) [37] and the DeLong test [38]. A higher AUC value indicates a better-performing model, reflecting its ability to distinguish between the two classes across various threshold levels. The DeLong test specifically compares the areas under two or more correlated ROC curves to indicate a significant difference in performance between the models.

To enhance our understanding of how each feature contributes to the associated outcome, we employ SHAP (SHapley Additive explanations) analysis [39], the most widely used technique for explaining the importance of clinical features in predicting various diseases or patient prognosis.

Statistical analysis

We conducted significant testing using the t-test for numerical variables and the Chi-square test for categorical variables. Additionally, Spearman's correlation method was employed to assess the strength of the correlation between each feature and mortality. The ROC and the AUC were utilized to estimate the cutoff value for variables and their reliability in prognosis. For this analysis, we used commercial statistical software (SPSS for Windows, Version 15, SPSS Inc., Chicago, IL, USA). P-values less than 0.05 were considered statistically significant.

Machine learning analyses, including data preprocessing, model training, hyperparameter tuning, and visualization, were conducted using Python 3.11.5. The following libraries and their versions were used: numpy (1.25.2), pandas (2.0.3), imbalanced-learn (0.11.0), lightgbm (4.1.0), xgboost (1.7.3), matplotlib (3.7.3), scikitlearn (1.3.2), and shap (0.43.0). These tools ensured robust and reproducible analyses."

Result

Demographics and clinical profiles in patients with SICH

The present study comprised 1,451 patients, including 966 males and 485 females, with an average age of 64.54 years (mean \pm SD: 14.50). Among them, 285 patients succumbed to the condition, resulting in a total mortality rate of 19.6% (285 out of 1,451). Comparative analysis between the group experiencing mortality and the nonmortality group revealed lower blood pressure control, lower initial GCS scores, pupillary changes, the need for intubation due to respiratory failure, and comorbidities with kidney and DM diseases. The model training process incorporated 36 variable features, with 26 demonstrating significant differences related to mortality (p-value < 0.05). Comprehensive characteristics and the significance of these features in traumatic brain injury patients are presented in Table 1. Additionally, both the

Table 1 Demographics and significances in SICH patients

Feature	Overall		P-Value		
		No	Yes		
	N=1451	N=1166	N=285		
Gender				0.463	
Female, n (%)	485 (33.43)	384 (32.93)	101 (35.44)		
Male, n (%)	966 (66.57)	782 (67.07)	184 (64.56)		
Age, mean (SD)	64.54 (14.50)	63.97 (14.49)	66.88 (14.34)	0.002	
Height, mean (SD)	163.04 (9.25)	163.01 (9.23)	163.19 (9.33)	0.772	
Weight, mean (SD)	64.99 (15.77)	65.83 (16.12)	61.55 (13.74)	< 0.001	
Systolic blood pressure (SBP), mean (SD)	157.97 (35.10)	160.28 (33.87)	148.53 (38.40)	< 0.001	
Diastolic blood pressure (DBP), mean (SD)	83.33 (18.67)	84.74 (18.23)	77.55 (19.35)	< 0.001	
Body temperature (BT), mean (SD)	36.57 (0.71)	36.56 (0.60)	36.63 (1.06)	0.238	
Pulse, mean (SD)	85.14 (18.28)	84.12 (16.46)	89.31 (23.94)	0.001	
Respiratory rate (RR), mean (SD)	17.49 (4.34)	17.52 (4.18)	17.37 (4.95)	0.648	
Glasgow Coma Scale - eye opening (GCS_E) mean (SD)	2.73 (1.39)	3.05 (1.28)	1.38 (0.94)	< 0.001	
Glasgow Coma Scale - verbal response (GCS_V) mean (SD)	2 81 (1 82)	3 18 (1 79)	1 27 (0 91)	< 0.001	
Glasgow Coma Scale - motor response (GCS_M) mean (SD)	4 42 (1 99)	4 76 (1.91)	3.06 (1.70)	< 0.001	
Left Pupil			5.66 (1.7.6)	(0.00)	
Pupil reflex (-), n (%)	298 (20.54)	120 (10.29)	178 (62.46)	< 0.001	
Pupil reflex (+), n (%)	1153 (79.46)	1046 (89.71)	107 (37.54)		
Pupil size(L), mean (SD)	3.13 (1.00)	2.97 (0.72)	3.76 (1.57)	< 0.001	
Right Pupil					
Pupil reflex (-), n (%)	298 (20.54)	119 (10.21)	179 (62.81)	< 0.001	
Pupil reflex (+), n (%)	1153 (79.46)	1047 (89.79)	106 (37.19)		
Pupil size(R), mean (SD)	3.11 (0.97)	2.97 (0.71)	3.66 (1.54)	< 0.001	
Muscle power - left upper extremity (Muscle LUE), mean (SD)	2.33 (1.80)	2.65 (1.78)	1.04 (1.20)	< 0.001	
Muscle power - left lower extremity (Muscle LEF), mean (SD)	2.34 (1.77)	2.65 (1.75)	1.06 (1.17)	< 0.001	
Muscle power - right upper extremity (Muscle RUE), mean (SD)	2.36 (1.85)	2.68 (1.84)	1.06 (1.26)	< 0.001	
Muscle power - right lower extremity (Muscle RI F), mean (SD)	2.36 (1.81)	2.69 (1.80)	1.03 (1.14)	< 0.001	
Inspired fraction of oxygen (EiO2), mean (SD)	29.59 (12.59)	27.43 (9.94)	38.41 (17.53)	< 0.001	
Endotracheal tube (Endo)	25.05 (12.057)	2,113 (515 1)	00.11 (17.00)		
No.n (%)	621 (42 80)	565 (48 46)	56 (19 65)	< 0.001	
Yes n (%)	830 (57 20)	601 (51 54)	229 (80 35)		
External ventricular drain (EVD)	030 (37.20)	001 (01.01)	229 (00.00)		
No.n (%)	1159 (79.88)	927 (79 50)	232 (81 40)	0.525	
Yes n (%)	292 (20 12)	239 (20 50)	53 (18 60)	0.525	
Intracranial pressure (ICP), p. (%)	202 (20112)	200 (20.00)	55 (10.00)		
No.n (%)	1216 (83.80)	978 (83 88)	238 (83 51)	0.951	
Yes n (%)	235 (16 20)	188 (16 12)	47 (16 49)	0.551	
Surgery n (%)	221 (15 23)	184 (15 78)	37 (12 98)	0 277	
Medication used	221 (15.25)	101 (15.70)	57 (12.50)	0.277	
Vacontestors n (%)	200 (13 78)	72 (6 17)	128 (44 91)	< 0.001	
Sedative/ Hyppotic n (%)	811 (55.89)	618 (53 00)	193 (67 72)	< 0.001	
Nicardinine n (%)	800 (55.13)	645 (55.32)	155 (54 39)	0.828	
Systemic disease	000 (35.15)	015 (55.52)	135 (31.55)	0.020	
Hypertension n (%)	878 (60 51)	694 (59 52)	184 (64 56)	0 1 3 5	
Diabetes mellitus n (%)	336 (23.16)	251 (21 53)	85 (29 82)	0.004	
Heart disease n (%)	223 (15 37)	162 (13 89)	61 (21 40)	0.00-	
Cerebrovascular disease n (%)	258 (17 78)	194 (16 64)	64 (22.46)	0.002	
Gastrointestinal disease n (%)	104 (7 17)	73 (6 26)	31 (10.88)	0.027	
Liver Disease in (%)	116 (7 00)	83 (7 1 2)	33 (11 58)	0.01	
Kidney disease, n (%)	101 (6.96)	61 (5.23)	40 (14.04)	< 0.001	

Table 1 (continued)

Feature	Overall	Mortality		P-Value
		No	Yes	
	N=1451	N=1166	N=285	
Cancer, n (%)	60 (4.14)	44 (3.77)	16 (5.61)	0.218
APACHE II, mean (SD)	15.04 (7.28)	13.25 (6.19)	22.35 (6.83)	< 0.001
Sequential Organ Failure Assessment (SOFA score), mean (SD)	3.60 (2.61)	3.00 (2.18)	6.06 (2.80)	< 0.001

Table 2	Spearman correlation	coefficient (r) b	etween features	and mortality,	, sorted by	absolute val	ues. Bold text	: absolute value
greater	than 0.2; <i>Italic text</i> : abso	olute value great	er than 0.1					

Feature	r	Feature	r	Feature	r
pupil_reflex+(R)	-0.517	DBP	-0.153	BT	-0.042
pupil_reflex+(L)	-0.513	SBP	-0.119	RR	-0.041
GCS_E	-0.460	Weight	-0.102	Surgery	-0.031
GCS_V	-0.413			Gender	-0.021
GCS_M	-0.408	pupil_size(L)	0.18	EVD	-0.019
Muscle_RLE	-0.354	pupil_size(R)	0.150		
Muscle_LLEE	-0.352	Kidney disease	0.137	Heart disease	0.083
Muscle_LUE	-0.350	Sedative/Hypnotic	0.118	Age	0.081
Muscle_RUE	-0.337			Diabetes mellitus	0.078
				Pulse	0.074
APACHE II	0.464			Gastrointestinal disease	0.071
Vasopressors	0.446			Liver Disease	0.065
SOFA	0.431			Cerebrovascular disease	0.060
FiO2	0.390			Hypertension	0.041
Endo	0.231			Cancer	0.037
				Height	0.011
				ICP	0.004

APACHE II score and SOFA score exhibited high significance in predicting the in-hospital mortality rate of SICH patients in the intensive care unit.

Correlation between features and mortality (Spearman correlation coefficient)

Both right and left pupil light reflexes exhibited a noteworthy negative correlation with mortality, featuring correlation coefficients of -0.517 and -0.513, respectively, suggesting that higher values of these features are associated with reduced mortality rates. Similarly, attributes related to GCS E, V, and M components displayed negative correlations with mortality, indicating that an increase in these features is associated with lower mortality. Conversely, vasopressors, FiO2, kidney disease, and sedative-hypnotic drugs demonstrated positive correlations with mortality, featuring higher correlation coefficients, suggesting that an increase in these features is associated with elevated mortality rates. It's crucial to note that vasopressors exhibited a relatively strong correlation coefficient of 0.446, signifying a significant association with higher mortality. Additionally, some other features like muscle status, cancer, and ICP also displayed a positive correlation with mortality, although these correlations were weaker (Table 2).

SHAP analysis of feature importance in XGBoost with 36 variables

Figure 3 demonstrates the feature importance in the bestperforming predictive model, XGBoost, using 36 variables for post-ICH outcomes. The analysis utilizes SHAP values to evaluate each feature's contribution to the model's predictions, aiding in identifying and ranking relevant attributes. The SHAP summary plot (left) illustrates the direction and magnitude of the feature impacts on the predictions, while the mean absolute SHAP values (right) rank features based on their overall influence. Key variables such as pupil light reflex, GCS scores, vasopressors, and muscle status exhibit the highest impact, indicating their critical roles in predicting mortality. These findings underscore the importance of neurological and physiological parameters in constructing accurate and reliable predictive models for post-ICH outcomes.

Mortality prediction models in five different AI algorithms

Through ROC analysis and AUC calculations, we identified models for mortality risk prediction using 36 feature variables. The XGBoost-based model demonstrated the best predictive performance with an AUC of 0.913, followed by LightGBM (AUC = 0.909), Random Forest (AUC = 0.905), Logistic Regression (AUC = 0.899), and MLP (AUC = 0.892) (Fig. 4). Additionally, the XGBoostbased model exhibited the highest accuracy (0.833) for



Fig. 3 SHAP summary plot (a) and mean absolute SHAP values (b) for the best model (XGBoost) with 36 Features

mortality risk prediction, with a sensitivity of 0.826 and specificity of 0.834, as detailed in Table 3. The details of the 5-fold cross-validation results are shown in Supplemental Table 2, while the hyperparameter ranges used during the grid search are provided in Supplemental Table 3.

Performance and feature importance of the XGBoost model using the top 18 feature variables

Since the XGBoost-based model demonstrated the best predictive performance using 36 features, we proceeded to select the top 18 features, ranging from vasopressors to the left pupil size, in the XGBoost 36-feature model (Fig. 3) to create a new predictive model. This streamlined model achieved impressive performance metrics, including an AUC of 0.913, an accuracy of 0.828, sensitivity of 0.826, specificity of 0.829, and an F1-score of 0.654, demonstrating its ability to maintain strong predictive power while reducing the number of features.

To identify these top features, we performed SHAP (SHapley Additive explanations) analysis [37], which provides a deeper understanding of how each feature contributes to the predicted outcomes. In Fig. 5(a), the color of the SHAP plot represents the original feature values, with red indicating higher values and blue indicating

lower values. A broader spread of SHAP values suggests a stronger influence on the outcome. For example, patients using vasopressors (represented by red dots) are associated with an increased risk of death (indicated by positive SHAP values), whereas higher values of GCS_E and left pupil light reflex are associated with a reduced risk of mortality.

Figure 5(b) ranks the influence of features on the outcome based on their mean absolute SHAP values. The top seven influential features include vasopressors, GCS_E, left pupil light reflex, right pupil light reflex, FiO2, Muscle_LUE, and EVD. These findings highlight the significance of these variables in accurately predicting post-ICH outcomes and demonstrate the effectiveness of the 18-feature XGBoost model in providing reliable and interpretable predictions.

The DeLong test compares XGBoost-based models with different feature combinations and conventional tools (APACHE II and SOFA scores) in predicting mortality

In terms of sensitivity, the 18-feature model performs slightly better with a sensitivity of 0.826, while the 36-feature model achieves a sensitivity of 0.802. This suggests that the 18-feature model is marginally better at correctly identifying positive results, but the difference is



Fig. 4 Analysis of receiver operating characteristic curves (ROC), area under the curve (AUC), plotting sensitivity versus 1-specificity for, logistic regression (LR) (orange), random forest (black), LightGBM (green), XGBoost (pink dashed line) and MLP (blue dashed line) using the 36 feature variables

 Table 3
 Model performance with 36 feature variables

Algorithm	Accuracy	Sensitivity	Specificity	F1-	AUC
				score	
Logistic Regression	0.787	0.779	0.789	0.590	0.899
Random Forest	0.839	0.779	0.854	0.657	0.905
LightGBM	0.849	0.756	0.871	0.663	0.909
XGBoost	0.844	0.756	0.866	0.657	0.913
MLP	0.846	0.756	0.869	0.660	0.892

not statistically significant. In the DeLong test comparison between the 18-feature and 36-feature models, the p-value of 1 indicates that there is no statistically significant difference in AUC between the two models. Therefore, the two models exhibit similar predictive capabilities, and neither significantly outperforms the other. The AI model, utilizing both 18 and 36 features, continues to outperform both the APACHE II and SOFA scoring systems in all performance metrics, including accuracy, sensitivity, specificity, and AUC. This underscores the superior predictive performance of the AI model for patient mortality and its enhanced ability to accurately identify high-risk patients compared to traditional scoring systems (Table 4).

Interface presentation of AI in real-world clinical application within the Chi Mei hospital healthcare system

After a series of analyses, we have concluded that the XGBoost-based model, using a combination of 18 features, is more lightweight. As a result, we have integrated it into the hospital system to aid clinical doctors and nurses in treatment and to facilitate communication with patients' families. The "Original" column represents data for the current status, displaying information from the time of admission to the ICU. The "Adjust" column allows the observer to modify the values of each feature to understand the effect of each feature on the risk of mortality, serving as a reference for treatment (see Fig. 6).

The model, developed in Python using scikit-learn, outputs files in Pickle format (PKL). The user interface, built with Visual Studio[®] using Visual Basic (version 17.7), retrieves patient feature values through web APIs



Fig. 5 (a) SHAP global explanation on the 18-feature model (XGBoost model) (b) SHAP absolute value of each feature on the 18-feature model (XGBoost model)

Table 4 P-values from DeLong's test for AUC comparisons between XGBoost models (36 vs. 18 features) and scoring systems APACHE II and SOFA

Model	Accuracy	Sensitivity	Specificity	F1-score	AUC	P-value (vs. XGBoost 36) *	P-value (vs. XGBoost 18) **
XGBoost (36 features)	0.844	0.756	0.866	0.657	0.913	-	1
XGBoost (18 features)	0.828	0.826	0.829	0.654	0.913	1	-
APACHE II	0.725	0.733	0.723	0.490	0.826	< 0.001	< 0.001
SOFA	0.727	0.698	0.734	0.520	0.788	< 0.001	< 0.001

Note: *P-value (vs. XGBoost 36): DeLong's test comparing AUC with XGBoost (36 features). **P-value (vs. XGBoost 18): DeLong's test comparing AUC with XGBoost (18 features)

(application interfaces) connected to the HIS, calls the PKL model file, and returns risk probabilities. Developers can customize the interface using tools such as Visual Studio, PyCharm, or Jupyter Notebook, based on their preferences. (The model PKL file and interface source codes can be requested from the corresponding author).

A comparison with related studies

Table 5 demonstrates several strengths that enhance its value in the field of mortality prediction, including a larger sample size, diverse feature variables, high predictive performance, real-world applicability, and a comprehensive approach.

Discussion

Summary and novelty of current study

This is the first study to combine feature variables to predict the risk of ICU mortality in patients with SICH using an AI model. The XGBoost-based model was found to be superior to traditional scoring systems, such as APACHE II and SOFA. Moreover, this approach has been implemented in a clinical system and aids in clinical decision-making, planning by the medical team, and shared decision-making with patients. These results underscore the potential of using machine learning models, particularly lightweight implementations like the 18-feature XGBoost model, in assisting clinical decision-making. Such models can enhance the identification of high-risk ICH patients, streamline resource allocation, and ultimately improve patient outcomes.

Strategies for addressing data imbalance

To address data imbalance, we applied SMOTE during training, ensuring robust model training and validation. The 5-fold cross-validation results (Supplemental Table 2) further confirmed the stability of our approach. Additionally, we mitigated overfitting by analyzing learning curves (Supplemental Fig. 1), which showed convergence between training and validation performances, demonstrating the reliability of the models.

The XGBoost model with 36 features achieved an AUC of 0.981 ± 0.027 , while the 18-feature XGBoost model achieved a similarly high AUC of 0.977 ± 0.029 . These results highlight the stability and robustness of the XGBoost model across different feature sets, even with the application of SMOTE.



Machine learning algorithm predicts mortality in intensive care unit for patients with Spontaneous Intracerebral Hemorrhage

Feature Original Adjust GCS E 1 1 GCS V 2 2 3 GCS_M 3 Left Pupil reflex Yes Yes **Right Pupil reflex** Yes Yes 2.5 2.5 Left Pupil size **Right Pupil size** 3 3 Muscle_LUE 2 2 1 Muscle_LLE 1 Muscle RUE 2 2 Muscle RLE 1 1 Fio2 60 60 EVD Yes Yes Surgery No No vasopressors No No Hypertension Yes Yes **Diabetes mellitus** No No Kidney disease No No

Fig. 6 Interface presentation of AI in practical application within the Chi Mei Hospital healthcare system

Study	This Study	Nie et al., 2021 [<mark>21</mark>]	Lim et al., 2021 [<mark>22</mark>]	Guo et al., 2022 [<mark>23</mark>]
Setting	Intensive care unit	Intensive care unit	In hospital patient	Emergency room
Patient number	1451	760	297	751
Study method	5 ML method + 2 scoring system	6 ML method + 1 scoring system	2 ML method + 1 scoring system	6 ML method + 1 scoring system
Feature variables	36 feature variables	72 feature variables	15 feature variables	3–19 feature variables
Outcome	Mortality	Mortality	Mortality	Mortality
Testing results (AUC)	0.913	0.819	0.900	0.844
Best predicting model	XGBoost	Random forest	Support Vec- tor Machine	Logistic regression
Real world application	Yes	No	No	Yes

	Table 5	Com	parison	with	recent	studies
--	---------	-----	---------	------	--------	---------

ML: Machine learning

To further address concerns about overfitting, we plotted the learning curves for both the 36-feature and 18-feature XGBoost models (Supplemental Fig. 1). The learning curves showed a clear convergence between training and validation performances, indicating that the

models-maintained robustness and did not overfitting, even with the use of SMOTE for data balancing.

Demographics and clinical picture

In the current study, hypertension is diagnosed based on past medical history and records. It represents a long-term risk factor, potentially increasing stroke risk through mechanisms such as atherosclerosis and vascular narrowing [40]. Post-stroke elevated SBP results from acute physiological changes, possibly due to autonomic nervous system regulation or acute stress responses [41].

Our data suggests an unexpected finding: post-stroke higher systolic blood pressure (SBP) is associated with a lower mortality rate in patients with spontaneous intracerebral events; contrary to the common belief that hypertension is a stroke risk factor [7]. The relationship between blood pressure and mortality is complex, involving various interacting factors rather than a straightforward linear association. This phenomenon is consistent to previous studies that high blood pressure has lower mortality in different situation [42, 43].

In the current study, we found that patients with a history of hypertension had higher SBP in both survival and mortality cases compared to those without hypertension. In clinical practice, patients with higher blood pressure often take antihypertensive drugs to manage their condition, which can result in this variable being influenced by external intervention, making it less objective. Consequently, blood pressure was not included among the 18 features in the final model.

Comorbidities have been mortality risk factors in ICH [12]. In current study, the observed associations between mortality and factors such as DM (p = 0.004), kidney diseases (p < 0.001), and heart disease (p = 0.002) suggest that these variables serve as crucial early indicators of patient outcomes. This emphasizes the importance of timely interventions to effectively manage these factors in ICH patients.

Elevated APACHE II and SOFA scores are associated with a higher likelihood of mortality. Our results reveal the significant predictive power of APACHE II and SOFA scores in in-hospital mortality [14, 15], emphasizing the potential of these scoring systems as valuable tools in risk assessment. This finding supports their continued use in intensive care units, potentially leading to improved patient care and resource allocation.

Correlation between features and mortality

The strong negative correlation between pupil light reflexes and GCS components, signifying an increased mortality risk with a decrease in these features, is consistent with prior studies [13, 44]. This highlights the reliability of assessing neurological function, particularly pupils and GCS, as predictors of patient outcomes. These findings underscore the potential use of these physiological indicators as predictive markers for adverse outcomes in critically ill patients, aiding clinicians in prioritizing interventions, especially in ICH cases.

Consistent with previous research [44], positive correlations, particularly the strong association of vasopressor use, increased FiO2, and endotracheal intubation with higher mortality, draw attention to modifiable risk factors that clinicians should monitor closely. For example, the use of vasopressors, elevated FiO2, and endotracheal intubation indicate complex treatment strategies, compromised physiological states, potential delays in initiating appropriate treatment, or responses to deteriorating conditions. The need for these interventions suggests a higher severity of illness, contributing to an increased risk of mortality.

Feature importance in the best model (XGBoost) with 36 features

Figure 3 illustrates the SHAP analysis of feature importance in the best-performing predictive model, XGBoost, using 36 features for post-ICH outcomes. The SHAP summary plot (Fig. 3a) visualizes the direction and magnitude of each feature's impact on the model's predictions, while the mean absolute SHAP values (Fig. 3b) rank features based on their overall contribution to the model.

Key variables, such as pupil light reflex, GCS components, vasopressors, and muscle status, emerge as the most impactful factors in predicting mortality. These findings highlight the critical importance of neurological and physiological parameters in developing accurate and reliable predictive models. Additionally, the consistent prominence of these features underscores their potential as robust indicators for clinical decision support tools and further model refinement.

AUC for mortality prediction in five different AI algorithms

Table 3 showcases robust AUC values ranging from 0.892 to 0.913 for all five AI-based models, with the XGBoostbased model notably leading with an AUC of 0.913. This highlights the potential of machine learning algorithms in clinical outcome predictions for SSICH patients.

The reasons why XGBoost is considered the best mortality model can be attributed to several factors. (1) The larger sample size (1451 patients) favors XGBoost's stability and generalization. With training on 36 features, the model captures intricate relationships crucial for mortality prediction. (2) XGBoost's proficiency in handling non-linear relationships aligns well with the challenges in mortality prediction, especially in addressing non-linearities in the top influential feature variables [45].

Feature importance in the best model (XGBoost) with 18 features

Based on the simpler yet effective model is advantageous in real-world clinical settings, we selected the top 18 features from the XGBoost 36 feature model (Fig. 3d). To boost the practical use of AI in clinical settings, we employed the widely used SHAP (SHapley Additive exPlanations) technique for explaining clinical feature significance in predicting diseases or patient prognosis [25]. Notably, the leading seven features most influential features—vasopressors, GCS_M, GCS_V, GCS_E, Muscle_LUE, Muscle_RLE are all risk factors for mortality. This choice emphasizes the importance of model interpretability and practicality. Further research could explore the ideal balance between model complexity and performance, crucial for practical clinical applications.

Considering features selection in the final model

At our hospital, we primarily use the APACHE II [13] and SOFA [15] assessment tools for clinical decision-making in ICU. APACHE II estimates ICU mortality based on laboratory values and patient age and signs, considering both acute and chronic diseases. The SOFA score quantifies six distinct scores, one for each of the severity of failed organs, encompassing respiratory, coagulation, liver, cardiovascular, renal, and central nervous system functions. These tools assist in predicting patient mortality, identifying high-risk patients, and effectively communicating with patients and their families to explain their medical condition in the ICU.

To compare the AI models with APACHE II and SOFA scores, we employed the DeLong test. The results revealed that the ML models generally outperformed the traditional tools. Therefore, the AI models remain valuable tools for clinical practice, offering improved predictive performance and more accurate risk assessment.

In clinical application, the choice between the 18-feature and 36-feature models may depend on practical considerations, such as model complexity and resource requirements. Whether combining APACHE II, SOFA, and our AI model to establish a new predictive model is worth evaluating in the future remains to be seen.

Real-world application

The integration of the 18-feature based XGBoost model into the hospital system marks a significant advance toward practical clinical implementation. Its simplicity and efficiency make it suitable for daily application. This software improves the capability to engage with prediction functions, permitting manual adjustment of parameter values to reassess outcomes. For example, it can show how a decrease in FiO2 increases risk. This enhances the contribution of our research.

Strengths and limitations of the current study

Our study possesses several strengths. Firstly, it had a larger sample size of 1451 patients, enhancing the statistical power and generalizability of findings. A larger sample size often renders study results more reliable. Secondly, we employed both a comprehensive set of 36 features and a more concise set of 18 features, providing flexibility in model development and practical application. This demonstrates the adaptability of our approach. Thirdly, the high AUC values obtained (0.913 for the 36 features and 0.913 for the 18 features) reflect the efficacy of the predictive models in differentiating mortality risk. Such robust predictive performance can significantly aid in clinical decision-making. Fourthly, our study explicitly states that integration of the XGBoost-based model into the healthcare system in a real-world context, suggesting its potential clinical utility and relevance in healthcare settings.

However, several limitations in the current study should be acknowledged. First, as a retrospective observational study, there is the possibility of miscoded feature variables. Researchers have limited control over the data collection process, which may introduce biases or confounding factors. Second, imaging parameters such as size of intra-cerebral hemorrhage, midline shift and the presence/absence of brain ventricles have not been quantitatively incorporated into our ML model. Therefore, the potential confounding effects of the numerous features utilized require further exploration. Third, our study relied on data from a single intensive care unit, limiting the generalizability of findings to other healthcare contexts with different patient populations and treatment protocols. Therefore, establishing a multi-center data sharing platform will enhance the usability of data by allowing artificial intelligence machines to engage in federated learning [46]. Fourth, while SMOTE effectively addressed data imbalance in the training set, it carries the potential risk of overfitting due to the generation of synthetic samples. However, the learning curves and cross-validation results confirmed that our XGBoost models maintained robust performance without overfitting. Future work may explore alternative data balancing methods or hybrid approaches to further optimize model training. Finally, most importantly, someone should have knowledge not only on mortality risk but on functional outcome as well and at a specified follow-up check. This is what can alter decision making. Consequently, there is a need for larger prospective studies with more comprehensive data collection and the inclusion of additional variables to be considered in the future.

Conclusion

Our study developed a mortality prediction model for spontaneous ICH patients using machine learning, highlighting XGBoost's superior performance. The integrated 15-feature model is now part of clinical practice at Chi Mei Hospital, aiding treatment decisions. It's important to note that our AI predictive tool is a clinical aid, not a replacement for a doctor's judgment. Before implementing AI-based policies, thorough evaluations on ethics and societal impact, including privacy and fairness considerations, are crucial.

Abbreviations

, is bit c matrio	
SICH	Spontaneous intracerebral hemorrhage
AI	Artificial intelligence
AUC	Area under curve
APACHE II	Acute Physiology and Chronic Health Evaluation II
SOFA	Sequential Organ Failure Assessment
GCS	Glascow Coma Scale
ICU	Intensive care unit
XAI	Explanatory artificial intelligence
Shap	SHapley Additive exPlanations
HIS	Hospital information system
TRIPOD	Transparent Reporting of a Multivariable Prediction Model for
	Individual Prognosis or Diagnosis
RF	Random forest
LR	Logistic Regression
LightGBM	Light Gradient Boosting Machine
XGBoost	Extreme Gradient Boosting
MLP	Multi-layer Perceptron
SMOTE	Synthetic Minority Oversampling Technique
HR	Heart rate; RR: respiratory rate
SBP	Systolic blood pressure
DBP	Diastolic blood pressure
FiO2	Inspired fraction of oxygen

- EVD External ventricular drainage
- ICP Intracranial pressure
- ROC Receiver operating characteristic
- PKL Pickle format
- APIs Application interfaces

Supplementary Information

The online version contains supplementary material available at https://doi.or g/10.1186/s12911-025-02984-y.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgments

The authors would like to thank all of the researchers, especially Yu-Ting Shen who extended their unwavering support in this study.

Author contributions

Kuan-Chi Tu, Ching-Lung Kuo and Nai-Ching Chen conceived and designed the experiments. Chia-Jung Chen, and Chung-Feng Liu performed the experiments, Ching-Lung Kuo and Tee-Tau Eric Nyam analyzed the data, Che-Chuan Wang and Nai-Ching Chen contributed reagents/materials/analysis tools, and Ching-Lung Kuo, Xiao-Han Vivian Yap and Tee-Tau Eric Nyam wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding

This research received CMFHR11091 grant from Chi-Mei Medical Center.

Data availability

Due to patient privacy concerns within the Chi Mei Medical Center's Health Information Network, the primary data supporting this article cannot be shared publicly. Nonetheless, de-identified data will be made available upon reasonable request to the corresponding author.

Declarations

Ethics approval and consent to participate

Ethics Approval and Consent to Participate This study was approved by the Institutional Review Board of Chi Mei Medical Center (approval number: 11107-012). The requirement for informed consent was waived by the ethics committee due to the study's retrospective nature. All procedures were conducted in accordance with relevant laws, regulations, and the principles outlined in the Declaration of Helsinki (https://www.wma.net/policies-post/w ma-declaration-of-helsinki/).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Neurosurgery, Chi Mei Medical Center, Tainan 710402, Taiwan

²Department of Nursing, Chi Mei Medical Center, Tainan 710402, Taiwan ³Department of Information Systems, Chi Mei Medical Center, Tainan 710402, Taiwan

⁴Department of Medical Research, Chi Mei Medical Center, Tainan 710402. Taiwan

⁵Center of General Education, Chia Nan University of Pharmacy and Science, Tainan 717301, Taiwan

⁶School of Medicine, College of Medicine, National Sun Yat-Sen University, Kaohsiung, Taiwan Received: 17 October 2024 / Accepted: 21 March 2025 Published online: 28 March 2025

References

- 1. Hankey GJ, Stroke. Lancet. 2017;389:641-54.
- Greenberg SM, Ziai WC, Cordonnier C, et al. 2022 guideline for the management of patients with spontaneous intracerebral hemorrhage: A guideline from the American heart association/american stroke association. Stroke. 2022;53(7):e282–361.
- 3. Katan M, Luft A. Global burden of stroke. Semin Neurol. 2018;38(2):208–11.
- Camacho E, LoPresti MA, Bruce S, et al. The role of age in intracerebral hemorrhages. J Clin Neurosci. 2015;22(12):1867–70.
- Pasi M, Casolla B, Kyheng M, et al. Long-term mortality in survivors of spontaneous intracerebral hemorrhage. Int J Stroke. 2021;16(4):448–55.
- Ganti L, Shameem M, Houck J, et al. Gender disparity in Stoke: women have higher ICH scores than men at initial ED presentation for intracerebral hemorrhage. J Natl Med Assoc. 2023;115(2):186–90.
- Francoeur CL, Mayer SA, VISTA-ICH Collaborators. Acute blood pressure and outcome after intracerebral hemorrhage: the VISTA-ICH cohort. J Stroke Cerebrovasc Dis. 2021;30(1):105456.
- Hemphill JC 3rd, Bonovich DC, Besmertis L, Manley GT, Johnston SC. The ICH score: a simple, reliable grading scale for intracerebral hemorrhage. Stroke. 2001;32(4):891–7.
- Kalita J, Misra UK, Vajpeyee A, Phadke RV, Handique A, Salwani V. Brain herniations in patients with intracerebral hemorrhage. Acta Neurol Scand. 2009;119(4):254–60.
- Chen JW, Gombart ZJ, Rogers S, Gardiner SK, Cecil S, Bullock RM. Pupillary reactivity as an early indicator of increased intracranial pressure: the introduction of the neurological pupil index. Surg Neurol Int. 2011;2:82.
- Wankhade BB, Kumar A, Mudassir S, Ranjan A. Clinicoradiological and biochemical predictors of mortality in hospitalized patients of spontaneous intracerebral hemorrhage. J Neuroanaesthesiol Crit Care. 2023;10:46–50.
- 12. Faghih-Jouybari M, Raof MT, Abdollahzade S, et al. Mortality and morbidity in patients with spontaneous intracerebral hemorrhage: A single-center experience. Curr J Neurol. 2021;20(1):32–6.
- Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985;13(10):818–29.
- Huang Y, Chen J, Zhong S, Yuan J. Role of APACHE II scoring system in the prediction of severity and outcome of acute intracerebral hemorrhage. Int J Neurosci. 2016;126(11):1020–4.
- Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related organ failure Assessment) score to describe organ dysfunction/failure. On behalf of the working group on Sepsis-Related problems of the European society of intensive care medicine. Intensive Care Med. 1996;22(7):707–10.
- Trentino KM, Schwarzbauer K, Mitterecker A, et al. Machine Learning-Based mortality prediction of patients at risk during hospital admission. J Patient Saf. 2022;18(5):494–8.
- Seki T, Kawazoe Y, Ohe K. Machine learning-based prediction of in-hospital mortality using admission laboratory data: A retrospective, single-site study using electronic health record data. PLoS ONE. 2021;16(2):e0246640.
- 18. Jabal MS, Joly O, Kallmes D, et al. Interpretable machine learning modeling for ischemic stroke outcome prediction. Front Neurol. 2022;13:884693.
- de Jong G, Aquarius R, Sanaan B, et al. Prediction models in aneurysmal subarachnoid hemorrhage: forecasting clinical outcome with artificial intelligence. Neurosurgery. 2021;88(5):E427–34.
- Matsuo K, Aihara H, Nakai T, Morishita A, Tohma Y, Kohmura E. Machine learning to predict In-Hospital morbidity and mortality after traumatic brain injury. J Neurotrauma. 2020;37(1):202–10.
- Nie X, Cai Y, Liu J, et al. Mortality prediction in cerebral hemorrhage patients using machine learning algorithms in intensive care units. Front Neurol. 2021;11:610531.
- Lim MJR, Quek RHC, Ng KJ, et al. Machine learning models prognosticate functional outcomes better than clinical scores in spontaneous intracerebral haemorrhage. J Stroke Cerebrovasc Dis. 2022;31(2):106234.
- 23. Guo R, Zhang R, Liu R, et al. Machine Learning-Based approaches for prediction of patients' functional outcome and mortality after spontaneous intracerebral hemorrhage. J Pers Med. 2022;12(1):112.
- 24. Ley C, Martin RK, Pareek A, Groll A, Seil R, Tischer T. Machine learning and conventional statistics: making sense of the differences. Knee Surg Sports Traumatol Arthrosc. 2022;30(3):753–7.

- Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). Comput Methods Programs Biomed. 2022;226:107161.
- Meurer WJ, Tolles J. Logistic regression diagnostics: Understanding how well a model predicts outcomes. JAMA. 2017;317(10):1068–9.
- 27. Breiman L. Random forests. Mach Learn. 2001;45:5-32.
- Ke GL, Meng Q, Finley T, Wang TF, Chen W, Ma WD et al. LightGBM: A highly efficient gradient boosting decision tree. Neural Inform Process Syst. 2017.
- Wang X, Zhu T, Xia M, et al. Predicting the prognosis of patients in the coronary care unit: A novel Multi-Category machine learning model using XGBoost. Front Cardiovasc Med. 2022;9:764629.
- Rahman A, Debnath T, Kundu D, Khan SI, Aishi AA, Sazzad S, Sayduzzaman M, Band SS. Machine learning and deep learning-based approach in smart healthcare: recent advances, applications, challenges and opportunities. AIMS Public Health. 2024;11:58–109.
- Tu KC, Eric Nyam TT, Wang CC, et al. A Computer-Assisted system for early mortality risk prediction in patients with traumatic brain injury using artificial intelligence algorithms in emergency room triage. Brain Sci. 2022;12(5):612.
- Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321–57.
- Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med. 2018;18(3):91–3.
- 34. Swets JA. Measuring the accuracy of diagnostic systems. Science. 1988;240(4857):1285–93.
- Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol. 2008;56(1):45–50.
- Sokolova M, Japkowicz N, Szpakowicz S. (2006, December). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. In Australasian joint conference on artificial intelligence (pp. 1015–1021). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Jin H, Lin CX. Using AUC and accuracy in evaluating learning algorithms. IEEE Trans Knowl Data Eng. 2005;17:299–310.

- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837–45.
- Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc. 2017;4768–4777.
- Johansson BB. Hypertension mechanisms causing stroke. Clin Exp Pharmacol Physiol. 1999;26(7):563–5.
- Al-Qudah ZA, Yacoub HA, Souayah N. Disorders of the autonomic nervous system after hemispheric cerebrovascular disorders: an update. J Vasc Interv Neurol. 2015;8(4):43–52.
- 42. Liao JC, Ho CH, Liang FW, et al. One-year mortality associations in Hemodialysis patients after traumatic brain injury -- an eight-year population-based study. PLoS ONE. 2014;9(4):e93956.
- Cheng CY, Ho CH, Wang CC, Liang FW, Wang JJ, Chio CC, Chang CH, Kuo JR. One-Year mortality after traumatic brain injury in liver cirrhosis Patients–A Ten-Year Population-Based study. Med (Baltim). 2015;94:e1468.
- Vo HK, Nguyen CH, Vo HL. High In-Hospital mortality incidence rate and its predictors in patients with intracranial hemorrhage undergoing endotracheal intubation. Neurol Int. 2021;13(4):671–81.
- Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA. 2016;13–17,785–794.
- Tajabadi M, Grabenhenrich L, Ribeiro A, Leyer M, Heider D. Sharing data with shared benefits: artificial intelligence perspective. J Med Internet Res. 2023;25:e47540.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.