UNIVERSA MEDICINA

pISSN: 1907-3062 / eISSN: 2407-2230

Univ Med 2026;45:13-26

DOI: https://doi.org/10.18051/UnivMed.2026.v45.13-26

ORIGINAL ARTICLE

Machine learning models for predicting 48-hour mortality in acute intracerebral hemorrhagic stroke

Intan Kemaladina[1], Syahrul[2], Taufik Fuadi Abidin[3], Nasrul Musadir[2], and Imran[2]

[1]Residency Program in Neurology, Department of Neurology, Faculty of Medicine, Universitas Syiah Kuala/Dr. Zainoel Abidin Hospital, Banda Aceh, Indonesia
[2]Department of Neurology, Faculty of Medicine, Universitas Syiah Kuala/Dr. Zainoel Abidin Hospital, Banda Aceh, Indonesia
[3]Department of Informatics, Faculty of Science and Mathematics, Universitas Syiah Kuala, Banda Aceh, Indonesia

*** Correspondence Author:**
intan.kemaladina@gmail.com

Date of first submission, November 15, 2025
Date of final revised submission, January 6, 2026
Date of acceptance, January 12, 2026

Cite this article as: Kemaladina I, Syahrul, Abidin TF, Musadir N, Imran. Machine learning models for predicting 48-hour mortality in acute intracerebral hemorrhagic stroke. Univ Med 2026;45:13-26

ABSTRACT

BACKGROUND

Identifying patients with intracerebral hemorrhagic (ICH) at high risk of mortality is crucial for timely intervention. Machine learning (ML) offers novel methodologies for precise predictive models for ICH. Therefore, the aim of this study was to develop an ML-based predictive model for 48-hour mortality in patients with acute hemorrhagic stroke.

METHODS

A cross-sectional study was conducted using secondary data from 657 patients diagnosed with acute ICH. Demographic, clinical, laboratory, and radiological variables were extracted from medical records. Data preprocessing included cleaning, normalization, and class balancing using the Synthetic Minority Oversampling Technique (SMOTE). Three supervised algorithms—Random Forest, Decision Tree, and Gaussian Naïve Bayes—were developed and evaluated using stratified 5-fold cross-validation. Model performance was assessed using accuracy, sensitivity, specificity, precision, recall, F1-score, and AUC.

RESULTS

Random Forest achieved the best overall performance for predicting 48-hour mortality, with an accuracy of 84.77%, F1-score of 84.63%, and AUC of 80.51, outperforming Decision Tree (AUC 61.12) and Gaussian Naïve Bayes (AUC 82.94). Random Forest most accurately identified >48-hour survival, with high sensitivity (93.5%) and PPV (92.9%), while Naïve Bayes provided the most reliable positive classification for this category (PPV 99.0; specificity 94.2%). For ≤24-hour mortality, Naïve Bayes showed the best detection performance (sensitivity 85.4%; NPV 98.7%).

CONCLUSION

Machine learning, particularly the Random Forest algorithm, enables reliable prediction of 48-hour mortality in patients with acute ICH using basic clinical and radiological data available at admission. The model offers practical potential for early risk stratification in emergency and critical care settings.

Keywords: Intracerebral hemorrhage, machine learning, early mortality, predictive model, 48-hour.

INTRODUCTION

Hemorrhagic stroke remains one of the most catastrophic neurological emergencies, contributing to a disproportionately high global burden of mortality and disability.[1,2] Ischemic stroke is the most common subtype, making up about 65–85% of all strokes, while intracerebral hemorrhage (ICH) represents 10–30% of cases globally.[3] Despite advances in neurocritical care and surgical management, ICH remains highly fatal.[4-6] Thirty-day mortality ranges from 30% to 44%, reaching up to 50% in severe cases, with many deaths occurring within the first week.[7] One-year mortality remains high at approximately 50–60%.[4] Early death, particularly within the first 10 days, is primarily associated with hematoma expansion, increased intracranial pressure, and secondary brain injury.[8-10] This acute and rapidly evolving phase represents a critical window in which accurate mortality prediction can guide clinical decision-making, resource prioritization, and treatment planning.

Several prognostic scoring systems, including the ICH score and the acute physiology and chronic health evaluation II (APACHE II), have been developed to estimate mortality risk in ICH.[11,12] Although both instruments are clinically valuable, each has significant limitations. The ICH score depends on clinical and radiological interpretation, which may vary between evaluators and requires specialized expertise.[13] The APACHE II system, widely used in intensive care settings, often undergoes simplification to facilitate manual calculation, which may reduce predictive accuracy.[6] These limitations highlight the need for an adaptive, objective, and efficient approach capable of integrating multiple variables to improve prognostic accuracy in hemorrhagic stroke.

Machine learning, a data-driven branch of artificial intelligence, offers the potential to address these challenges. By analyzing large, multidimensional datasets, machine learning algorithms can uncover complex and nonlinear interactions among clinical, laboratory, and radiological features that are often overlooked by conventional statistical models.[14] In the case of hemorrhagic stroke, this approach enables automated risk prediction based on patient-specific information, providing an opportunity to improve precision and reliability in outcome forecasting.[15-17]

Previous studies have explored machine learning applications in stroke prognosis; however, most research has focused on broad outcome measures, such as in-hospital or 30-day mortality.[18-22] Despite these advances, evidence remains limited for very short-term mortality prediction, especially to predict 48-hour mortality in acute hemorrhagic stroke. This outcome is clinically critical, as a large proportion of fatal events occur during the early acute phase, when intensive monitoring and therapeutic decisions are most influential.

Although machine learning has shown promise, the current literature remains inconclusive regarding which machine-learning approach offers the most reliable and clinically applicable performance. An analysis of 3,489 patients with acute ischemic stroke admitted to the intensive care unit, who survived and remained hospitalized beyond the first 48 hours, using data from the Medical Information Mart for Intensive Care IV (MIMIC-IV) database, demonstrated that machine-learning–based models have substantial capability in predicting the risk of in-hospital mortality in this clinical setting.[23] However, the study was conducted from intensive care settings and may not be directly applicable to earlier phases of care.

Moreover, early risk stratification in the emergency department, where initial clinical decisions and triage occur, has been minimally explored in prior machine learning studies. Therefore, the present study aimed to develop and evaluate a machine learning–based predictive model for 48-hour mortality in patients with acute hemorrhagic stroke using emergency department-based clinical data.

METHODS

Research design

A cross-sectional study was performed involving the development of a machine learning model based on secondary data from patients diagnosed with ICH. The study was conducted at Dr. Zainoel Abidin Hospital, Banda Aceh, Indonesia. The dataset included clinical, laboratory, and neuroimaging variables obtained from patients diagnosed with ICH between

January 2022 and December 2024. A supervised machine learning approach was implemented to develop a predictive model for 48-hour mortality in patients with acute intracerebral hemorrhage. The dataset was preprocessed through data cleaning, normalization, and exclusion of incomplete records. Class imbalance was addressed using the Synthetic Minority Oversampling Technique (SMOTE) to enhance model generalization across mortality categories. Three algorithms—Random Forest, Decision Tree, and Gaussian Naïve Bayes—were trained and validated using stratified 5-fold cross-validation. Model performance was assessed through multiple metrics, including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC).

Study subjects

A consecutive sampling method was employed to include all eligible cases of ICH recorded between January 2022 and December 2024. The inclusion criteria were: 1) adult patients aged 18 years or older with a confirmed diagnosis of ICH based on non-contrast head CT scans; 2) patients with complete clinical documentation encompassing medical history, laboratory investigations, and neuroimaging findings; and 3) patients who underwent stroke severity assessment using the National Institutes of Health Stroke Scale (NIHSS) at admission. The exclusion criteria encompassed: 1) patients with a primary diagnosis other than ICH; 2) patients presenting with complex comorbid conditions that could independently affect clinical outcomes; 3) patients with incomplete or inconsistent medical records; and 4) cases lost to follow-up during hospitalization.

Sample size determination

The minimum required sample size was estimated using the rule of thumb for machine learning classification analysis. The formula applied was $n \geq 10 \times k \times c$, where n denotes the minimum sample size, k represents the number of predictor variables, and c signifies the number of outcome classes.[24] In the present study, 12 independent variables were included, encompassing four clinical, four laboratory, three computed tomography (CT)-based variables, and one additional variable, with three outcome categories representing different mortality intervals. Applying the formula yielded a minimum of 360 patient records, which was determined to be adequate to ensure model stability and generalizability.

Data collection

Data were collected retrospectively from hospital medical records. Each entry in the dataset represented a single clinical episode of ICH. In cases where a patient experienced multiple admissions due to recurrent events, each hospitalization was considered a separate case to preserve data independence. Clinical data such as age, stroke onset, blood pressure, Glasgow coma scale (GCS), and National Institutes of Health Stroke Scale (NIHSS) scores were obtained from standardized hospital admission forms. Laboratory data, including leukocyte count and random blood glucose levels, were retrieved from hospital laboratory information systems. Neuroimaging data, including hemorrhage location, hematoma volume, and midline shift, were extracted from radiology reports verified by radiologists. Additional parameters such as surgical evacuation and pneumonia were derived from the patients' treatment and progress notes. All data were reviewed and verified to ensure accuracy and completeness before entry into the analytical database. Missing values were managed using median imputation for numerical variables to reduce bias and preserve data variability. Categorical data were encoded numerically to enable computational analysis. Prior to modeling, exploratory data analysis was conducted to assess data distribution, identify outliers, and visualize potential relationships among variables. The complete dataset was then randomly divided into training and testing subsets, maintaining proportional representation of outcome classes. The training subset (80%) was used for model development and hyperparameter optimization, while the training (or testing) subset (20%) served for performance evaluation and external validation.

Model development

Model development was performed using Python software version 3.10 (Python Software Foundation, Beaverton, Oregon, USA) on the Google Collaboratory platform (Google LLC, Mountain View, California, USA). Data preprocessing steps included standardization of continuous variables and encoding of categorical variables to ensure algorithm compatibility. Three supervised machine learning algorithms were applied: Random Forest, Decision Tree, and Naïve

Bayes classifier. The Random Forest algorithm was utilized as an ensemble learning approach that integrates multiple decision trees to minimize overfitting and enhance predictive robustness. The Decision Tree model was implemented for its interpretability, as it allows visualization of hierarchical decision paths and facilitates understanding of variable interactions in clinical settings. The Naïve Bayes classifier, based on probabilistic reasoning, was chosen for its computational efficiency and suitability for small to medium-sized datasets. Hyperparameter tuning was conducted for each algorithm through grid search optimization to identify the parameter configurations that achieved the best predictive performance. The target output for all models was the 48-hour mortality classification, comprising three categories as previously defined. Model training involved fitting the algorithms to the training data, learning from patterns and interactions between input features and known outcomes. Feature importance ranking was derived primarily from the Random Forest model to identify the most influential predictors of short-term mortality. Class imbalance was addressed using SMOTE to enhance model generalization across mortality categories. The three algorithms—Random Forest, Decision Tree, and Gaussian Naïve Bayes—were trained and validated using stratified 5-fold cross-validation.

Outcome measurements

The independent variables included demographic, clinical, laboratory, and neuroimaging parameters identified in prior literature as relevant predictors of mortality in hemorrhagic stroke. These parameters included patient age, time from symptom onset to hospital arrival, systolic blood pressure at admission, Glasgow Coma Scale (GCS) score, NIHSS score, leukocyte count, random blood glucose at admission, hemorrhage location, hematoma volume, presence of midline shift, surgical evacuation, and occurrence of pneumonia complications. The dependent variable was 48-hour mortality, classified into three distinct outcome categories: death within 24 hours after onset, death within 24–48 hours after onset, and survival beyond 48 hours.

Model evaluation

Our model evaluation used a comprehensive set of performance metrics—sensitivity (recall), specificity, positive predictive value (PPV or precision), negative predictive value (NPV), F1-score, and ROC-AUC, the latter being an overall measure of the model's discriminatory ability.[25] These metrics were selected for their clinical relevance: sensitivity is crucial for identifying at-risk patients (minimizing false negatives), while specificity helps reduce unnecessary interventions. Accuracy measures the overall proportion of correct classifications, while precision assesses the proportion of correctly identified positive cases among all positive predictions. Recall, also referred to as sensitivity, quantifies the proportion of actual positive cases correctly identified by the model. The F1-score, representing the harmonic mean of precision and recall, was employed to balance predictive capability, particularly in the presence of class imbalance. The AUC provided a comprehensive assessment of discriminative ability across multiple probability thresholds and was considered the most important indicator for evaluating clinical applicability.[25] All performance metrics were calculated using the *scikit-learn* and *NumPy* Python libraries. The algorithm demonstrating the highest AUC and F1-score was selected as the optimal model due to its superior combination of accuracy, robustness, and generalizability. Cross-validation was performed to confirm model stability, and receiver operating characteristic (ROC) curves were constructed to visually compare classification performance among algorithms.

Ethical approval

The study protocol was reviewed and approved by the Ethics Committee for Health Research, Dr. Zainoel Abidin Hospital, Banda Aceh, Indonesia (Approval number: 060/ETIK-RSUDZA/2025), in accordance with the principles of the Declaration of Helsinki. Written informed consent was obtained from the patients or their legal guardians prior to enrolment.

RESULTS

Characteristics of the included patients

A total of 746 patients diagnosed with acute ICH were included in the analysis (**Table 1**). The mean age was 57.28 ± 12.19 years. Male patients constituted 57.1% of the cohort, whereas females accounted for 42.9%. Upon hospital admission, the mean GCS score was 11.23 ± 3.49, while the mean NIHSS score was 15.49 ± 8.01, indicating a wide spectrum of neurological deficits from mild

to severe. The mean systolic blood pressure on arrival was 189.20 ± 30.33 mmHg. The mean random blood glucose level was 141.78 ± 50.06 mg/dL. The mean leukocyte count was 11,911.17 ± 3,921.74/µL. Radiological evaluation revealed the presence of midline shift in 572 patients (76.68%). Regarding clinical outcomes, 605 patients (81.1%) survived beyond 48 hours, 31 patients (4.2%) died within 24–48 hours, and 49 patients (6.6%) died within the first 24 hours. Clinical outcome data were incomplete for 61 patients (8.2%).

Table 1. General characteristics of the research subjects (n=746)

Variable	n (%)
Age (years)	57.28 ± 12.19
Sex	
Male	57.1
Female	42.9
GCS score	11.23 ± 3.49
NIHSS score	15.49 ± 8.01
Systolic blood pressure (mmHg)	189.20 ± 30.33
Random blood glucose (mg/dL)	141.78 ± 50.06
Leukocyte count (/µL)	11,911.17 ± 3,921.74
Radiological findings	
Midline shift	572 (76.7)
Clinical outcomes	
Survived >48 h	605 (81.1)
Died 24–48 h	31 (4.2)
Died <24 h	49 (6.6)
Unknown	61 (8.2)

Note: Data presented as mean ± SD, except sex, radiological findings and clinical outcomes: n (%). GCS: Glasgow coma scale; NIHSS: National Institutes of Health stroke scale

Data preprocessing

The dataset comprised demographic, clinical, laboratory, radiological, and outcome variables from 746 patients with acute ICH. The primary outcome variable categorized mortality into three groups: survival beyond 48 hours, death within 24–48 hours, and death within 24 hours. Preprocessing began with exploratory data analysis, which included evaluation of variable distribution, data types, outlier detection, and assessment of missing values. Missing data were identified across several variables, with the highest proportion being observed in GCS (7.91%) and systolic blood pressure (7.77%). The proportion of missing values was below 10% for all variables, indicating that deletion of incomplete cases would not substantially bias the analysis. Consequently, records containing missing data were excluded to preserve dataset integrity and analytical consistency. Subsequent data cleaning included verification of variable consistency, correction of data entry errors, and alignment of variable formats according to the operational definitions. Several ordinal variables initially stored as numeric values were recoded to reflect categorical classifications. After preprocessing, 657 complete cases were retained for model development. The cleaned dataset accurately represented the clinical spectrum of patients and ensured valid input for the modeling phase.

The distribution of clinical outcomes after the cleaning process showed 588 patients who survived >48 hours (89.5%), 28 patients who died within 24–48 hours (4.3%), and 41 patients who died within <24 hours (6.2%). The number of deaths was substantially lower than the number of survivors, creating a class imbalance. This condition poses a risk of prediction bias, in which the model becomes more likely to classify patients as survivors, reflecting the majority class. To address this issue, the Synthetic Minority Oversampling Technique (SMOTE) was applied to generate new synthetic samples in the minority classes. This approach enables the model to be trained on a more balanced distribution, thereby improving its ability to detect patients at high risk of mortality. The data distribution following the application of the SMOTE technique shows a balanced composition, with each class represented by 588 patients.

The feature importance analysis of the Random Forest model after applying SMOTE to the full dataset provides an overview of the relative contribution of each variable in predicting clinical outcomes (**Figure 1**). The GCS score emerged as the most dominant predictor, contributing 19.92%, indicating that the patient's level of consciousness is the most influential factor in determining the outcome. Leukocyte count ranked second with a contribution of 11.47%, followed by the NIHSS score at 10.54% and blood glucose level at 9.28%, demonstrating that neurological and inflammatory parameters play key roles in model prediction. Blood pressure showed a moderate contribution (8.82%), whereas age, evacuation, midline shift, hemorrhage location, and hematoma volume had relatively lower influences, with values ranging from 5–8%. Onset and pneumonia exhibited the smallest contributions, at 3.78% and 2.81% respectively, indicating that these variables were less significant in the model compared with the others.

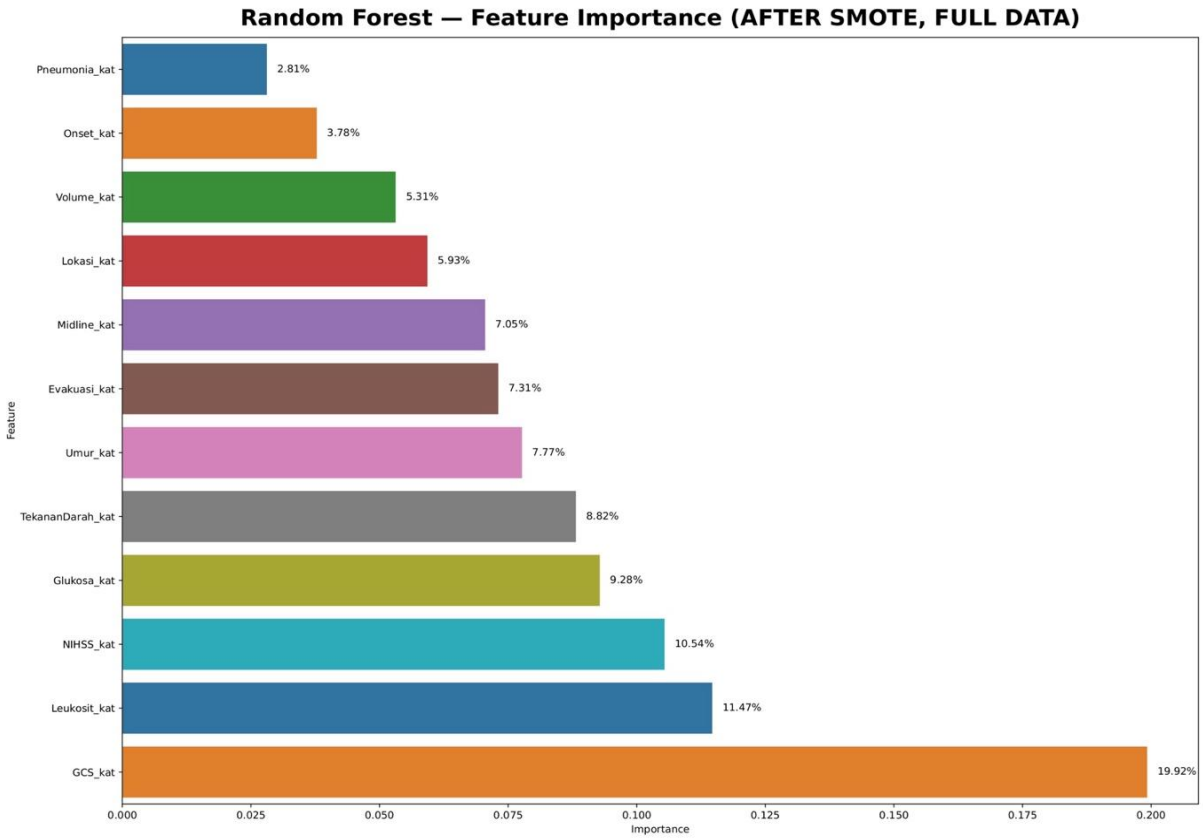

Figure 1. Feature importance plot of the Random Forest model after applying SMOTE to the entire dataset. The GCS score is the strongest predictor of clinical outcomes (19.92%), followed by leukocyte count (11.47%), NIHSS score (10.54%), and blood glucose level (9.28%). Other variables, including blood pressure, age, hemorrhage location, midline shift, hematoma volume, onset, and pneumonia, show smaller contributions, indicating a relatively lower influence on the model's predictions.

Model evaluation

Random Forest achieved the highest accuracy at 84.77%, precision of 84.57%, recall of 84.77%, F1-score of 84.63%, and AUC of 80.51% (**Table 2**). The balanced precision and recall values indicate consistent performance across survival and mortality classes. Decision Tree reached an accuracy of 80.98% and precision of 85.24%, with a lower AUC of 61.12%, suggesting limited discrimination despite adequate classification capacity. Gaussian Naïve Bayes recorded the lowest accuracy at 68.35%, but attained the highest precision (89.83%) and an AUC of 82.94%, reflecting acceptable discrimination but weaker sensitivity for mortality detection. Overall, Random Forest emerged as the most optimal algorithm, showing the best trade-off between accuracy, sensitivity, and F1-score, consistent with its ensemble architecture that mitigates the limitations of single-tree models. Decision Tree remained valuable for interpretability, while Naïve Bayes provided a computationally efficient yet less balanced alternative.

Table 2. Comparative performance of three machine learning models—Random Forest, Decision Tree, and Naïve Bayes (Gaussian)—based on key evaluation metrics, including accuracy, precision, recall, F1-score, and area under the curve

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
Random Forest	84.77	84.57	84.77	84.63	80.51
Decision Tree	80.98	85.24	80.98	82.90	61.12
Naïve Bayes (Gaussian)	68.35	89.83	68.35	75.18	82.94

Note : AUC : area under curve

Table 3. Comparative performance of Random Forest, Decision Tree, and Naïve Bayes models in predicting 48-hour mortality in acute intracerebral hemorrhagic stroke, evaluated using class-specific sensitivity, specificity, positive predictive value, and negative predictive value across ≤24-hour, 24–48-hour, and >48-hour outcome categories

Model	Class	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Random Forest	≤24 hours	14.6	94.2	14.3	94.3
	24–48 hours	3.6	96.5	4.3	95.7
	>48 hours	93.5	39.1	92.9	41.5
Decision Tree	≤24 hours	29.3	91.1	17.9	95.1
	24–48 hours	3.6	94.6	2.9	95.7
	>48 hours	88.4	49.3	93.7	33.3
Naïve Bayes	≤24 hours	85.4	71.3	16.5	98.7
	24–48 hours	3.6	96.2	4.0	95.7
	>48 hours	70.7	94.2	99.0	27.4

Note : PPV : positive predictive value; NPV : negative predictive value

Random Forest demonstrated the strongest performance for predicting survival beyond 48 hours, with high sensitivity (93.5%) and PPV (92.9%), indicating robust identification of patients unlikely to experience early mortality, although specificity remained limited (39.1%) (**Table 3**). Decision Tree showed a similar pattern for the >48-hour category, achieving high sensitivity (88.4%) and PPV (93.7%), but with moderate specificity (49.3%). Naïve Bayes yielded the highest PPV for >48 hours (99.0%) with high specificity (94.2%), suggesting highly reliable positive classification, albeit with lower sensitivity (70.7%). For the ≤24-hour category, Naïve Bayes outperformed the other models in sensitivity (85.4%) and NPV (98.7%), reflecting superior detection of very early mortality, whereas Random Forest and Decision Tree exhibited limited sensitivity despite high specificity. Across all models, prediction of the intermediate 24–48-hour category was consistently poor, with sensitivity remaining very low (3.6%) despite high specificity (>94%) and NPV (>95%), indicating persistent difficulty in discriminating this time window.

The Random Forest model achieved superior class balance, accurately identifying patients with acute mortality within 24 hours (93.2%) and maintaining stable detection performance across other outcome categories (**Figure 2**). Although misclassification occurred in some patients surviving beyond 48 hours, the model had more consistent prediction stability compared with Decision Tree and Naïve Bayes. Decision Tree achieved good accuracy for early mortality (<24 hours) but struggled to distinguish patients who died within 24–48 hours from survivors. Gaussian Naïve Bayes demonstrated a strong bias toward the survival class, correctly predicting most survivors but misclassifying the majority of mortality cases, reflecting the limitation of the independence assumption among clinical variables. In summary, Random Forest provided the most reliable and clinically applicable performance for early mortality prediction in patients with acute intracerebral hemorrhage. The model demonstrated stable accuracy, balanced sensitivity and specificity, and robust discriminatory power across multiple evaluation metrics.

Table 4 presents the evaluation results of the primary model using 12 features and the comparison model using 11 features after the hemorrhage evacuation variable was removed. In the primary model, the Random Forest algorithm demonstrated the strongest performance, with an accuracy of 84.77%, precision of 84.57%, recall of 84.77%, F1-score of 84.63%, and AUC of 80.51%. The balanced combination of high accuracy and F1-score indicates that Random Forest maintained consistent predictive ability for both survival and mortality classes. After the hemorrhage evacuation feature was excluded, the model performance decreased only slightly, with an accuracy of 84.30%, F1-score of 84.20%, and AUC of 80.30%. This minimal change suggests that the hemorrhage evacuation variable did not exert a significant influence on model performance, allowing the model to remain stable even when the feature was removed.

For the Decision Tree algorithm, the primary model with 12 features produced an accuracy of 80.98%, precision of 85.24%, recall of 80.98%, F1-score of 82.90%, and AUC of 61.12%. After the hemorrhage evacuation feature was excluded, the model performance remained relatively stable, with a slight improvement in AUC to 63.50%, while accuracy and F1-score remained within the

range of 80.50–83.60%. These findings indicate that the hemorrhage evacuation feature did not contribute meaningfully to the discriminative capability of the Decision Tree model and may have introduced minor noise in the classification process.

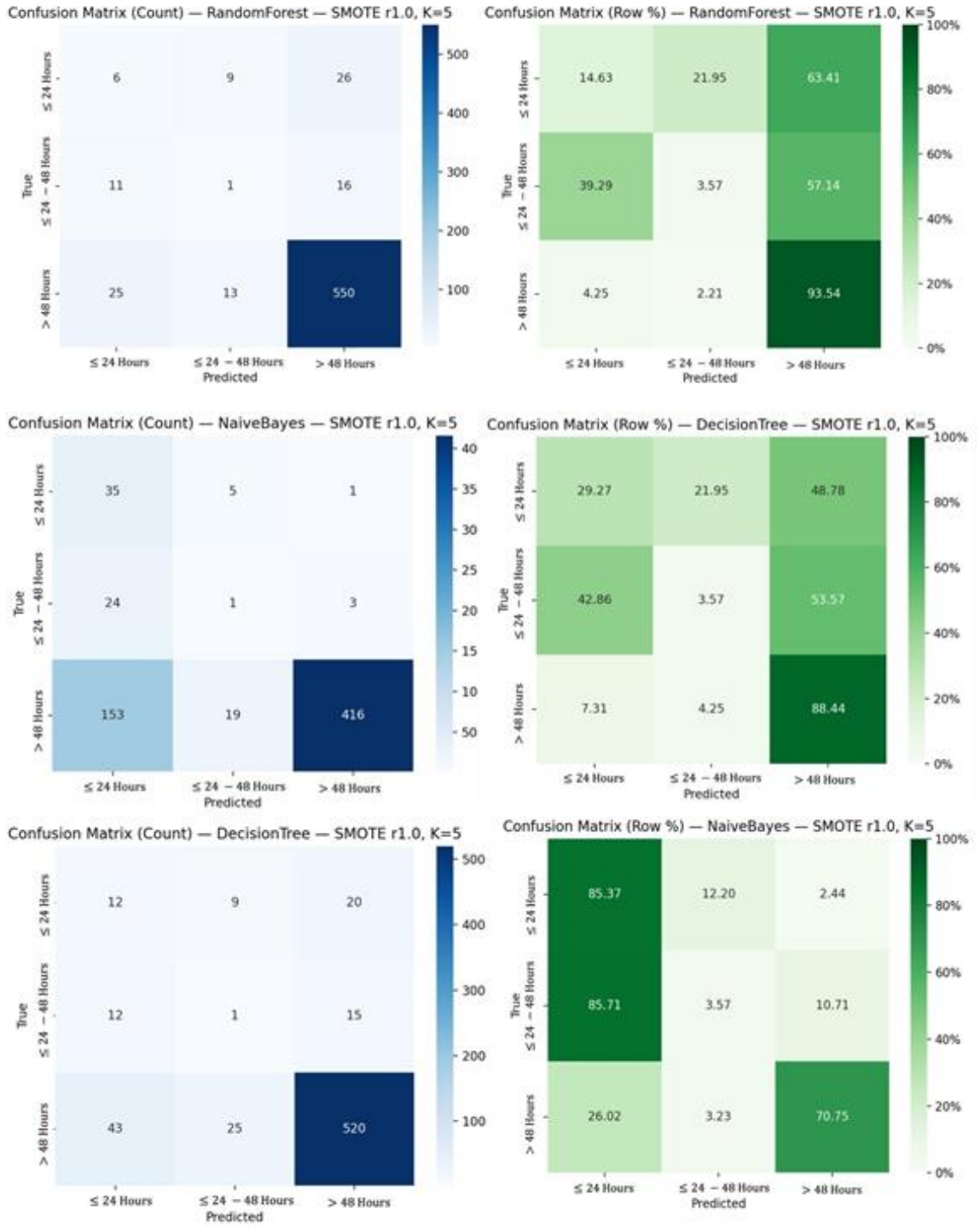

Figure 2. A heatmap-based confusion matrix derived from four machine-learning algorithms (Random Forest, Decision Tree, and Gaussian Naive Bayes) illustrates the distribution of predicted outcomes for patients with acute intracerebral hemorrhage compared with the actual clinical status. Darker shading indicates a higher proportion of correct classifications, with the Random Forest model demonstrating the most consistent performance, particularly in the group of patients who survived beyond 48 hours

Table 4. Performance metrics of machine-learning algorithms using 12 features vs. 11 features

Algorithm	Metric	Model with 12 Features (%)	Model with 11 Features (%)
Random Forest	Accuracy	84.77	84.30
	Precision	84.57	84.10
	Recall	84.77	84.30
	F1-score	84.63	84.20
	AUC	80.51	80.30
Decision Tree	Accuracy	80.98	82.20
	Precision	85.24	85.20
	Recall	80.98	82.20
	F1-score	82.90	83.60
	AUC	61.12	63.00
Naïve Bayes	Accuracy	68.35	69.60
	Precision	89.83	89.70
	Recall	68.35	69.60
	F1-score	75.18	76.20
	AUC	82.94	83.10

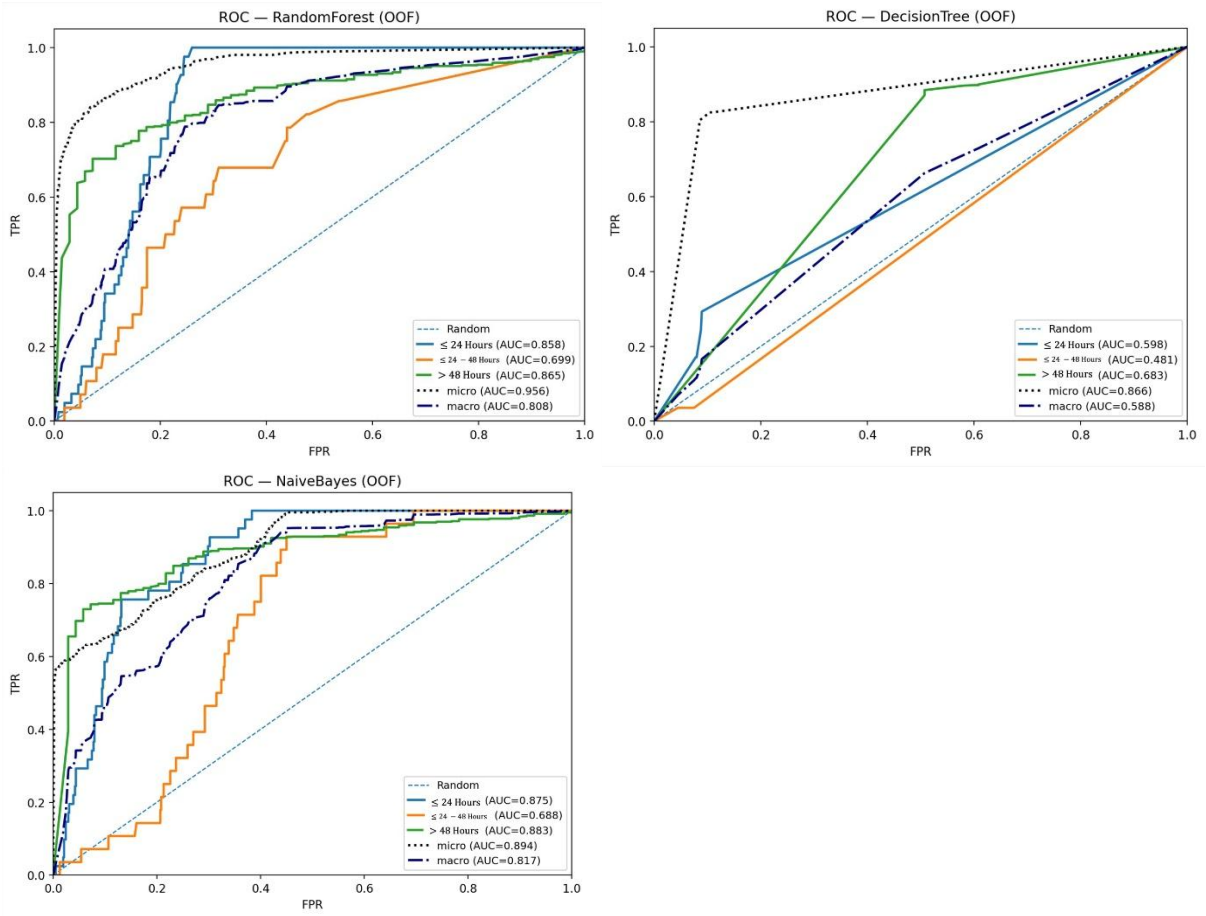

Figure 3. Receiver operating characteristic (ROC) curves from the evaluation of three machine-learning algorithms—Random Forest, Naive Bayes, and Decision Tree—in predicting mortality among patients with acute intracerebral hemorrhagic stroke

In contrast, the Gaussian Naive Bayes algorithm exhibited more variable performance. In the primary model, the accuracy was 68.35%, precision 89.83%, recall 68.35%, F1-score 75.18%, and AUC 82.94%. After the hemorrhage evacuation feature was removed, the model demonstrated consistent improvement, with an accuracy of 69.60%, F1-score of 76.20%, and AUC of 83.10%. This improvement suggests that in probabilistic models such as Naive Bayes,

removing features that are not fully independent of other variables can enhance prediction stability by reducing irrelevant redundancy or potential bias.

Figure 3 illustrates that the ROC curve for the Random Forest algorithm demonstrates excellent classification performance across all time-to-death categories. The model achieved an AUC of 0.858 for ≤24 hours, 0.699 for 24–48 hours, and 0.865 for >48 hours, with a micro-AUC of 0.956 and a macro-AUC of 0.808. These values indicate strong discriminative ability in distinguishing high-risk patients from those who survive, particularly in the ≤24-hour and >48-hour periods. The curve approaching the upper-left corner of the plot reflects optimal sensitivity and specificity, positioning Random Forest as the most stable and accurate algorithm among the three.

The ROC curve for the Decision Tree algorithm shows relatively lower classification performance compared with the other two algorithms. The AUC values were 0.598 for ≤24 hours, 0.481 for 24–48 hours, and 0.683 for >48 hours, with a micro-AUC of 0.866 and a macro-AUC of 0.588. The distance of the curve from the ideal diagonal indicates limited discriminative ability, especially in predicting mortality in the 24–48-hour interval. These findings highlight the tendency of Decision Tree models to overfit the training data, resulting in reduced performance when applied to the test set.

The ROC curve for the Naive Bayes algorithm demonstrates good performance, although slightly below that of Random Forest. The model achieved AUC values of 0.875 for ≤24 hours, 0.688 for 24–48 hours, and 0.883 for >48 hours, with a micro-AUC of 0.894 and a macro-AUC of 0.817. These findings indicate that Naive Bayes provided consistent predictions for the ≤24-hour and >48-hour periods, with a noticeable drop in performance for the 24–48-hour category. Nevertheless, the curve's proximity to the ideal area suggests that the model maintains favorable sensitivity and accuracy in identifying patients at high risk of mortality.

DISCUSSION

In the present study, three machine learning algorithms—Random Forest, Decision Tree, and Gaussian Naïve Bayes—were compared to develop a predictive model for 48-hour mortality in patients with acute intracerebral hemorrhage. Other algorithms such as support vector machine (SVM), logistic regression, extreme gradient boosting (XGBoost), artificial neural network (ANN), and K-nearest neighbor (KNN) were not included at this stage of model development. The stratified 5-fold cross-validation results demonstrated that Random Forest achieved the best overall performance, with an accuracy of 84.77%, F1-score of 84.63%, and AUC of 80.51%. The superiority of the Random Forest model can be attributed to its ensemble architecture, which aggregates multiple decision trees to capture complex non-linear interactions among clinical variables, thereby enhancing both stability and generalizability.[26] This finding is consistent with previous studies reporting that Random Forest provides higher predictive reliability than single-tree models in clinical outcome prediction.[26,27]

The Decision Tree model, while conceptually simple and highly interpretable, showed limited predictive power with an AUC of only 61.12%. A high rate of misclassification was observed, particularly in cases where survivors were incorrectly predicted as non-survivors. This outcome highlights the inherent limitation of single-tree models, which often fail to capture the heterogeneity and complex interdependence among clinical predictors in medical datasets. Despite these limitations, the Decision Tree algorithm remains valuable for exploratory analysis and identifying key mortality-related predictors due to its transparent rule-based structure. Furthermore, the Gaussian Naïve Bayes model displayed a strong bias toward the survival category, demonstrating high precision but low recall for mortality detection. This pattern aligns with the model's underlying assumption of predictor independence, which is rarely met in multifactorial clinical data. Nonetheless, the relatively high AUC of 82.94% suggests that this algorithm retains a satisfactory discriminatory capacity and may serve as a complementary model when computational simplicity is prioritized.

Confusion matrix visualization confirmed these findings. The Random Forest model achieved the highest accuracy in identifying patients who died within 24 hours (93.2%), followed by Decision Tree, which produced similar but less consistent results due to frequent misclassification between the 24–48-hour and >48-hour survival categories. In contrast, the Naïve Bayes model tended to predict most cases as long-term survivors (>48 hours), resulting in diminished sensitivity for acute mortality detection.

The present findings align closely with previous studies demonstrating the robustness of Random Forest in stroke mortality prediction. Abujaber et al.[19] reported a higher accuracy (0.954) using a multiethnic registry dataset, which may be explained by a larger and more diverse sample compared with the single-center dataset used in the present study. Nonetheless, the consistent advantage of Random Forest across studies underscores its reliability as a predictive framework. Similarly, Peng et al.[27] found that Random Forest outperformed artificial neural networks (ANN), support vector machines (SVM), and logistic regression in predicting 30-day mortality among patients with spontaneous ICH, with an AUC of 0.87, corroborating the present findings.

Differences between studies are likely influenced by sample size, feature selection, and model parameterization. In contrast, the present study demonstrated that Random Forest was superior to Decision Tree in detecting 48-hour mortality, a clinically significant endpoint often overlooked in previous studies. Fernandez-Lozano et al.[26] further support the robustness of Random Forest, showing strong predictive power for both short- and medium-term outcomes (AUC range: 0.79–0.95).

When the performance of the present study's model is compared to previously published research, it becomes evident that the locally developed Random Forest model achieved favorable results (AUC 0.805; F1-score 0.846) despite focusing on a narrower, early mortality endpoint (within 48 hours). Although the AUC value was slightly lower than that reported in studies with larger multicenter datasets, the consistent stability of Random Forest in early mortality prediction highlights its applicability to local clinical data.[26] The results also indicate that model performance depends strongly on the type of clinical outcome, the algorithm used, and dataset characteristics. Logistic regression demonstrated high accuracy in long-term mortality prediction, whereas Decision Tree performed best in in-hospital mortality prediction.[27,28] Conversely, the present study confirmed that Random Forest remained robust for short-term mortality prediction, reflecting its capacity to handle nonlinear relationships and class imbalance.

The findings of the present study differ from a similar previous study that identified other machine learning approaches, particularly gradient boosting or regression-based models, as the optimal predictors of stroke-related mortality. While the study of Cummins et al.,[23] conducted in intensive care unit settings, reported superior performance of models such as XGBoost or logistic regression; the present study demonstrated that Random Forest achieved the most reliable performance for predicting 48-hour mortality in acute intracerebral hemorrhage using emergency department data. This discrepancy may reflect differences in clinical context, stroke subtype, data availability, and outcome timeframe. Specifically, early mortality prediction in hemorrhagic stroke relies on rapidly obtainable and often heterogeneous variables, a setting in which ensemble bagging methods such as Random Forest may be more robust than boosting or parametric models. These contrasting results underscore that the optimal machine learning approach for stroke mortality prediction is highly context-dependent rather than universal.

In the present study, Decision Tree yielded the lowest performance among all models, being characterized by reduced AUC and F1-score values. This limitation can be explained by the model's tendency toward overfitting, where decision boundaries are excessively tailored to the training data, reducing its generalizability to new datasets.[28,29] This effect is exacerbated in complex clinical data, such as hemorrhagic stroke, where predictor interactions are often nonlinear and interdependent. Furthermore, Decision Tree models are sensitive to noise and class imbalance, both of which are common in medical datasets.[29] In contrast, ensemble methods such as Random Forest mitigate these issues by combining multiple trees, thus reducing variance and improving model stability.[30]

From a clinical perspective, the present study contributes novel evidence supporting the role of machine learning, especially Random Forest, as an effective tool for early outcome prediction in hemorrhagic stroke. The focus on the early phase (within 48 hours) represents a unique aspect rarely explored in previous research, offering valuable insight for acute decision-making. Implementation of such predictive models in emergency and critical care settings could assist physicians in promptly identifying high-risk patients using basic clinical and radiological parameters—such as GCS, NIHSS, systolic blood pressure, random blood glucose, and CT findings—available upon admission. This would allow prioritization of high-risk individuals for

intensive monitoring or surgical intervention, potentially improving survival outcomes. Furthermore, although Random Forest provided the most balanced performance, simpler models such as logistic regression may remain useful in ischemic stroke or settings with limited computational resources. These models are more interpretable and transparent, aligning with the practical needs of clinicians in low-resource environments.

Several limitations should be acknowledged when interpreting these findings. First, biochemical variables and comorbid conditions were not comprehensively analyzed. Factors such as lipid profiles, renal function, coagulation parameters, and chronic disease history (including diabetes mellitus, cardiovascular disease, and chronic kidney disease) are known to influence both the severity of intracerebral hemorrhage and recovery trajectories. Exclusion of these variables may have reduced the model's ability to capture certain mortality determinants. Second, the generalizability of the models is limited by the single-center design. As a tertiary referral hospital, Dr. Zainoel Abidin Hospital, Banda Aceh, primarily manages patients with more severe presentations, which may not represent the full clinical spectrum encountered in smaller healthcare facilities or primary care centers. Consequently, the predictive performance of the models may differ in other settings with distinct population characteristics, healthcare resources, or clinical management protocols. Future research should aim to validate the models using multicenter or national datasets encompassing diverse populations and broader clinical variables. Expanding data sources would not only enhance external validity but also facilitate development of hybrid models capable of predicting both mortality and functional outcomes, ultimately improving the clinical utility of machine learning applications in stroke care.

CONCLUSIONS

Machine learning, particularly the Random Forest algorithm, demonstrated strong capability in predicting 48-hour mortality among patients with acute ICH. Using readily available clinical and radiological variables, the model achieved high accuracy and robust discrimination, confirming the suitability of ensemble methods for complex clinical data. The focus on 48-hour mortality prediction provides valuable potential for rapid risk stratification and clinical decision support in acute stroke care. Broader validation across multicenter datasets and inclusion of additional variables such as biochemical markers and comorbidities are recommended to enhance model generalizability and clinical applicability.

Conflict of Interest

No relevant disclosures.

Acknowledgement

The authors extend their sincere gratitude to the Department of Informatics, Faculty of Science and Mathematics, Universitas Syiah Kuala, Banda Aceh, Indonesia, for their essential contributions to this study. Their support in the development, implementation, and optimization of machine learning algorithms was pivotal to the success of the predictive modeling. The collaboration significantly enhanced the technical rigor and computational reliability of the research.

Author Contributions

IK: Conceptualization, methodology, data curation, formal analysis, validation, visualization, writing – original draft, writing – review and editing, project administration; SS: Data acquisition, methodology, supervision, writing – review and editing; TFA: Supervision, methodology, resources, critical revision, writing – review and editing; NM: Supervision, methodology, critical revision, writing – review and editing; II: Supervision, methodology, critical revision, writing – review and editing. All authors have read and approved the final manuscript.

Funding

This study received no external funding.

Data Availability Statement

Derived data supporting the findings of this study are available from the corresponding author on request.

Declaration of Use of AI in Scientific Writing

Nothing to declare.

References

1. Lee TH. Intracerebral hemorrhage. Cerebrovasc Dis Extra 2025;15:1-8. doi: 10.1159/000542566.
2. Puy L, Parry-Jones AR, Sandset EC, Dowlatshahi D, Ziai W, Cordonnier C. Intracerebral

haemorrhage. Nat Rev Dis Primers 2023;9:14. doi: 10.1038/s41572-023-00424-7.
3. Parry-Jones AR, Krishnamurthi R, Ziai WC, et al. World Stroke Organization (WSO): Global intracerebral hemorrhage factsheet 2025. Int J Stroke 2025;20:145-50. doi: 10.1177/17474930241307876.
4. Magid-Bernstein J, Girard R, Polster S, et al. Cerebral hemorrhage: pathophysiology, treatment, and future directions. Circ Res 2022;130:1204-29. doi: 10.1161/CIRCRESAHA.121.319949.
5. Hillal A, Ullberg T, Ramgren B, Wasselius J. Computed tomography in acute intracerebral hemorrhage: neuroimaging predictors of hematoma expansion and outcome. Insights Imaging 2022;13:180. doi: 10.1186/s13244-022-01309-1.
6. Mutchmore A, Lamontagne F, Chasse M, Moore L, Mayette M. Automated APACHE II and SOFA score calculation using real-world electronic medical record data in a single center. J Clin Monit Comput 2023;37:1023-233. doi: 10.1007/s10877-023-01010-8.
7. Ray SK, Sarkar MSR, Ahmed KMA, Ahmed KMA, et al. Predicting 30-day outcomes in primary intracerebral hemorrhage using the intracerebral hemorrhage score: a study in Bangladesh. Cureus 2024;16:e73227. doi: 10.7759/cureus.73227.
8. Bautista W, Adelson PD, Bicher N, Themistocleous M, Tsivgoulis G, Chang JJ. Secondary mechanisms of injury and viable pathophysiological targets in intracerebral hemorrhage. Ther Adv Neurol Disord 2021;14:17562864211049208. doi: 10.1177/17562864211049208.
9. Li Z, You M, Long C, et al. Hematoma expansion in intracerebral hemorrhage: an update on prediction and treatment. Front Neurol 2020;11:702. doi: 10.3389/fneur.2020.00702.
10. Wan Y, Holste KG, Hua Y, Keep RF, Xi G. Brain edema formation and therapy after intracerebral hemorrhage. Neurobiol Dis 2023;176:105948. doi: 10.1016/j.nbd.2022.105948.
11. Fakiri MO, Uyttenboogaart M, Houben R, van Oostenbrugge RJ, Staals J, Luijckx GJ. Reliability of the intracerebral hemorrhage score for predicting outcome in patients with intracerebral hemorrhage using oral anticoagulants. Eur J Neurol 2020;27:2006-13. doi: 10.1111/ene.14336.
12. Gürbüz H, Topçu H. Estimating the outcomes of intracerebral haemorrhage with intracerebral haemorrhage score and acute physiology and chronic health evaluation-II score: a multicentre study. Turk J Anaesthesiol Reanim 2022;50:410-5. doi: 10.5152/TJAR.2022.21422.
13. Witsch J, Siegerink B, Nolte CH, et al. Prognostication after intracerebral hemorrhage: a review. Neurol Res Pract 2021;3:22. doi: 10.1186/s42466-021-00120-5.
14. Dhillon SK, Ganggayah MD, Sinnadurai S, Lio P, Taib NA. Theory and practice of integrating machine learning and conventional statistics in medical data analysis. Diagnostics (Basel) 2022;12:2526. doi: 10.3390/diagnostics12102526.
15. Abujaber AA, Albalkhi I, Imam Y, et al. Machine learning-based prediction of 90-day prognosis and in-hospital mortality in hemorrhagic stroke patients. Sci Rep 2025;15:16242. doi: 10.1038/s41598-025-90944-x.
16. Wei H, Huang X, Zhang Y, et al. Explainable machine learning for predicting neurological outcome in hemorrhagic and ischemic stroke patients in critical care. Front Neurol 2024;15:1385013. doi: 10.3389/fneur.2024.1385013.
17. Cao Y, Deng H, Liu S, et al. Development and validation of a machine learning-based risk prediction model for stroke-associated pneumonia in older adult hemorrhagic stroke. Front Neurol 2025;16:1591570. doi: 10.3389/fneur.2025.1591570.
18. Kurtz P, Peres IT, Soares M, Salluh JIF, Bozza FA. Hospital length of stay and 30-day mortality prediction in stroke: a machine learning analysis of 17,000 ICU admissions in Brazil. Neurocrit Care 2022;37(Suppl 2):313-21. doi: 10.1007/s12028-022-01486-3.
19. Abujaber AA, Albalkhi I, Imam Y, Nashwan A, Akhtar N, Alkhawaldeh IM. Machine learning-based prognostication of mortality in stroke patients. Heliyon 2024;10:e28869. doi: 10.1016/j.heliyon.2024.e28869.
20. Wang W, Otieno JA, Eriksson M, Wolfe CD, Curcin V, Bray BD. Developing and externally validating a machine learning risk prediction model for 30-day mortality after stroke using national stroke registers in the UK and Sweden. BMJ Open 2023;13:e069811. doi: 10.1136/bmjopen-2022-069811.
21. Schwartz L, Anteby R, Klang E, Soffer S. Stroke mortality prediction using machine learning: systematic review. J Neurol Sci 2023;444:120529. doi: 10.1016/j.jns.2022.120529.
22. Wang W, Kiik M, Peek N, et al. A systematic review of machine learning models for predicting outcomes of stroke with structured data. PLoS One 2020;15:e0234722. doi: 10.1371/journal.pone.0234722.
23. Cummins JA, Gerber BS, Fukunaga MI, Henninger N, Kiefe CI, Liu F. In-hospital mortality prediction among intensive care unit patients with acute ischemic stroke: a machine

learning approach. Health Data Sci 2025;5:0179. doi: 10.34133/hds.0179.
24. Khosravi B, Weston AD, Nugen F, et al. Demystifying statistics and machine learning in analysis of structured tabular data. J Arthroplasty 2023;38:1943-7. doi: 10.1016/j.arth.2023.08.045.
25. Li J. Area under the ROC curve has the most consistent evaluation for binary classification. PLoS One 2024;19:e0316019. doi: 10.1371/journal.pone.0316019.
26. Fernandez-Lozano C, Hervella P, Mato-Abad V, et al. Random forest-based prediction of stroke outcome. Sci Rep 2021;11:10071. doi: 10.1038/s41598-021-89434-7.
27. Peng SY, Chuang YC, Kang TW, Tseng KH. Random forest can predict 30-day mortality of spontaneous intracerebral hemorrhage with remarkable discrimination. Eur J Neurol 2010;17:945-50. doi: 10.1111/j.1468-1331.2010.02955.x.
28. Abujaber A, Yaseen S, Imam Y, Nashwan A, Akhtar N. Machine learning-based prediction of one-year mortality in ischemic stroke patients. Oxf Open Neurosci 2024;3:kvae011. doi: 10.1093/oons/kvae011.
29. Matsumoto K, Nohara Y, Soejima H, Yonehara T, Nakashima N, Kamouchi M. Stroke prognostic scores and data-driven prediction of clinical outcomes after acute ischemic stroke. Stroke 2020;51:1477-83. doi: 10.1161/STROKEAHA.119.027300.
30. 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. doi: 10.3389/fneur.2020.610531.