STM JOURNALS

Journal of
Artificial Intelligence
Research & Advances

ISSN: 2395-6720
Volume 9, Issue 2, 2022
DOI Journal: 10.37591/JoAIRA

http://computers.stmjournals.com/index.php?journal=JoAIRA&page=index

Hybrid Feature Selection Approach for Stability Prediction for Intracerebral Hemorrhage Patients

Balaji Ganesh Rajagopal*

Abstract

Intra Cerebral Hemorrhage (ICH) stability estimation is useful for enhancing diagnosis accuracy, selecting the best course of treatment, and clinically evaluating variations with healthy individuals. Due to their low predictive value, the clinical application of several ICH progression scoring systems is constrained. The dataset includes clinical parameters including age, Glasgow Coma Scale (GCS), and CT Angiography (CTA) spot that have been retrieved using geometric features from the segmented bleeding volume to predict ICH stability. To train and compare stability estimation, many cutting-edge machines learning approaches, including Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), Gradient Boosting, and Random Forest, are applied. We identify the shapely values (SHAP) and describe the key components of the ICH risk scoring system in order to harmonize clinical judgement with model learning. 20 of the patients, were judged to be in critical condition. With a precision of 82.9% and accuracy of 78.3%, the stability of the ICH patients was predicted. The mean square error of regression for expansion rate of the hemorrhage was 0.46. According to the SHAP analysis, the most important factors defining the stability of the stroke lesion are the CTA spot sign, age, solidity, position, and length of the initial axis of the ICH volume. An ablation study was conducted to reach the conclusion that by predicting long-term results, the integration of significant geometric elements with clinical features can enhance the ICH progression rating.

Keywords: Intracerebral hemorrhage, segmentation, risk prediction, growth rate estimation, 3D CNN, attention network

INTRODUCTION

Spontaneous ICH is the most common form of hemorrhagic stroke and a major cause of morbidity and mortality worldwide. Unlike ischemic stroke, few interventions have been proven to improve the clinical outcome of ICH. Hematoma expansion is associated with poor functional outcome and is a potential therapeutic target in acute ICH treatment [1]. Therefore, it is important to differentiate patients who have a stable hematoma from those with hematomas that are at risk of expansion. Extracted geometric features from the segmented bleeding are included in the dataset [2]. Then, it is fused with clinical features to train an MLP for stability estimation. Finally, the collected meaningful features are fused to estimate the growth rate of the hemorrhage and improve the ICH progression scoring system.

METHODS

In this section, the extraction and selection of features and regression techniques that are exploited for growth rate prediction are discussed.

***Author for Correspondence**
Balaji Ganesh Rajagopal
E-mail: balajiganesh.r@ist.srmtrichy.edu.in

Assistant Professor, Department of Computer Science and Engineering, Sri Ramaswamy Memorial Institute of Science and Technology, Tiruchirappalli Campus, Tiruchirapppali, Tamil Nadu, India

Received Date: July 12, 2022
Accepted Date: July 30, 2022
Published Date: August 18, 2022

Citation: Balaji Ganesh Rajagopal. Hybrid Feature Selection Approach for Stability Prediction for Intracerebral Hemorrhage Patients. Journal of Artificial Intelligence Research & Advances. 2022; 9(2): 9–17p.

Feature Extraction

The hemorrhage geometry and its location in the brain holds a vital role in reducing the risk associated with the hemorrhage tissue's growth. It is quite evident that the closer the hemorrhage tissue is to the center of the brain, the higher are the chances of the patient becoming a victim to acute hemorrhage. The smaller the size of the hemorrhage tissue, the greater are the chances of the patient not being subject to acute hemorrhage. Initially, the most significant geometrical features, which include eigenvalues, diagonal axis coordinates, major axis length, minor axis length, first axis length, second axis length, third axis length, 3D coordinates of the first axis, 3D coordinates of the second axis, 3D coordinates of the third axis coordinates, centroid coordinates, coordinates of the voxel corners, extent and solidity for the segmented hemorrhage are extracted. Besides these geometrical features, the volume of hemorrhage and the proportion of the segmented volume to the total volume are computed. The interpretation of extraction of geometric features from the segmented hemorrhage is shown in Figure 1. The image features of the hemorrhage, mean, variance, standard deviation, entropy, skewness, kurtosis, and histogram feature intensities are also taken into account. The four most important clinical features, Age, GCS, onset to CT, and presence of CTA spot sign are evaluated. To compute the first, second, and the third axis coordinates in the 3D Euclidean space, the inertia tensors were found, which take the form of a real symmetric three-dimensional matrix. In Figure 2, we compare the 3D visualization of the segmented hemorrhage for two cases: stable and unstable, deduced from the CT before and after the 24-h monitoring of the patient.

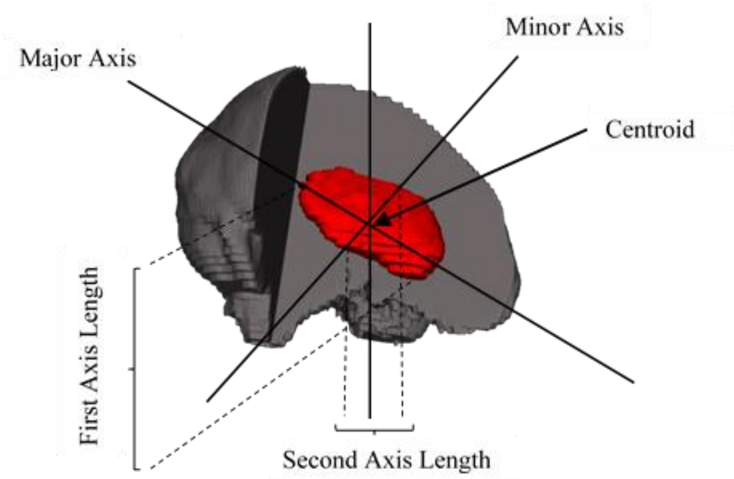
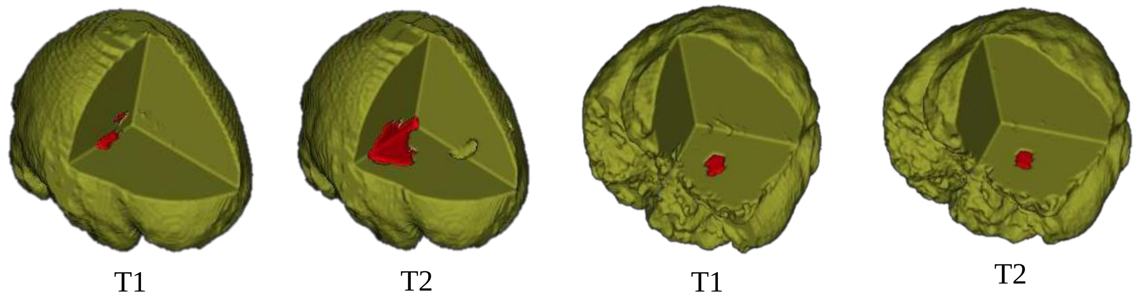


Figure 1. Visualization and interpretation of the geometric features extracted from the segmented hemorrhage. The red blob represents the segmentation of the lesion.



A) Unstable B) Stable

Figure 2. Visualization of the segmented hemorrhage output in 3D space.

The red blob represents the segmentation of the lesion. The first pair of images represents a malignant case of a hematoma, where the growth rate exceeds the threshold. The second pair of images represent a benign subject where the size of the lesion subsides by time T2.

Machine Learning Models

The probability of a target variable is predicted using the supervised learning classification algorithm known as logistic regression. There are only two viable classes since the goal or dependent variable is dichotomous in nature. The dependent variable is binary in nature, with data coded as either 1 (stands for success/yes) or 0 (stands for failure/no), to put it simply (Figure 3).

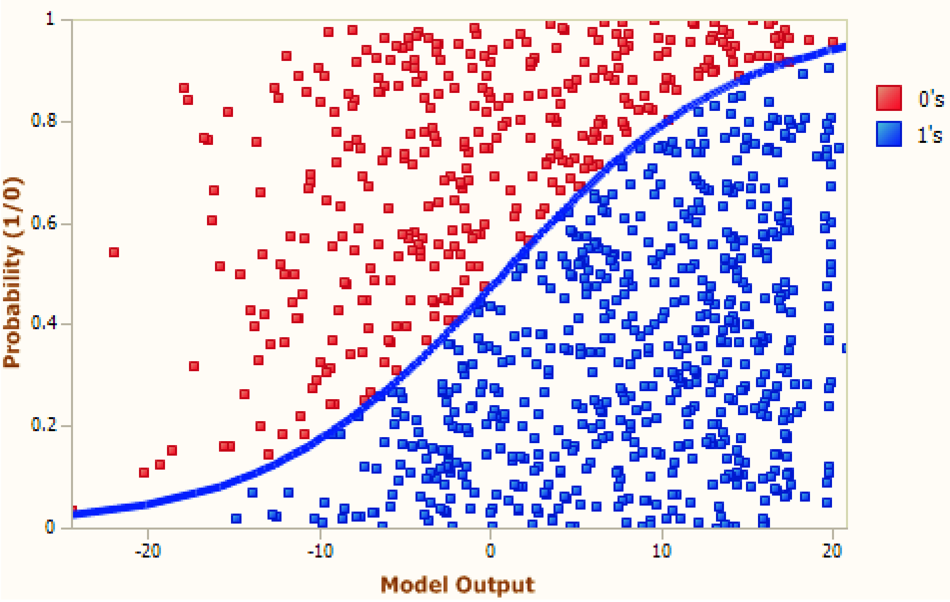


Figure 3. Logistic regression for classification.

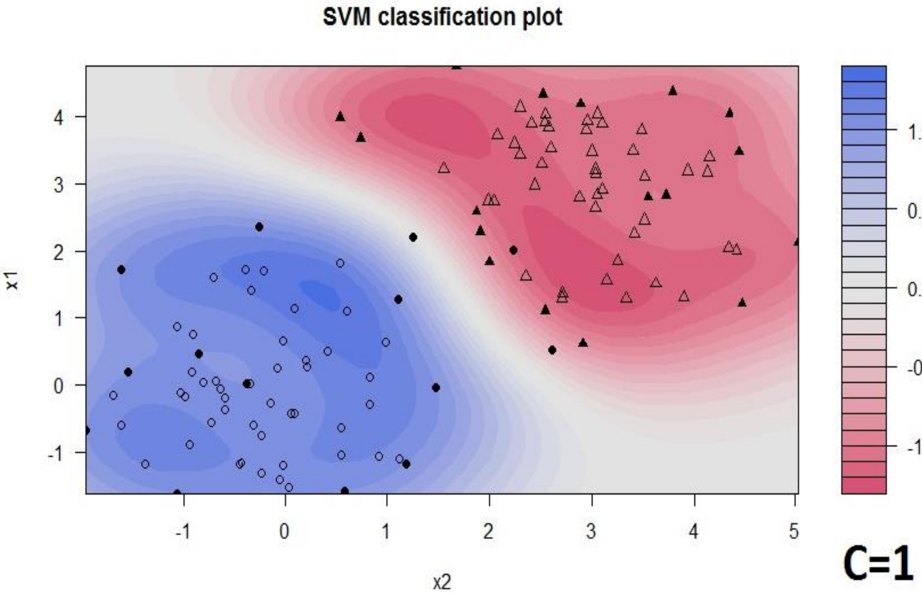


Figure 4. SVM for classification.

Support Vector Machine

SVMs, also known as support-vector networks in machine learning, are supervised learning models with corresponding learning algorithms that examine data used for regression and classification analysis [1]. An SVM training technique creates a model that categorizes fresh examples into one of two categories given a set of training examples, making it a non-probabilistic binary linear classifier (although there are ways to apply SVM in a probabilistic classification scenario, as Platt scaling) (Figure 4) [3].

Random Forest

A technique for supervised learning called random forest is utilized for both classification and regression. But it is primarily utilized for classification issues. Since trees make up a forest, more trees indicate a more vigorous forest, as is common knowledge. Similar to this, a random forest algorithm builds decision trees using data samples, obtains predictions from each one, and then votes to determine which is the best answer as shown in Figure 5.

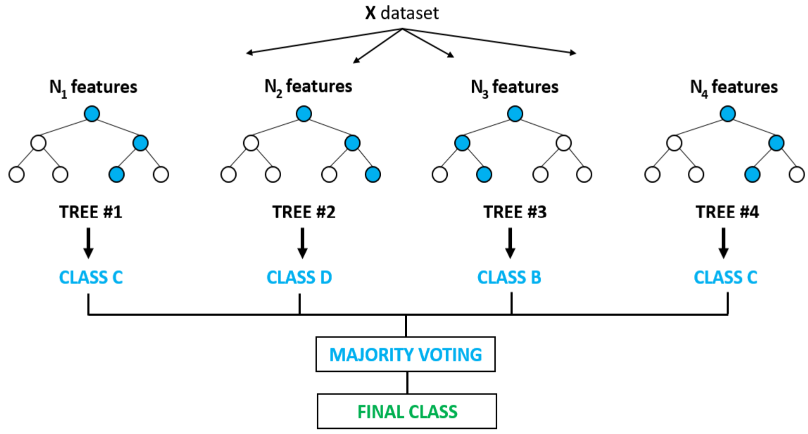


Figure 5. Random Forest for classification.

MLP

An MLP is a class of feed-forward artificial neural network (ANN). MLP is a vague word that is sometimes used to refer to any feedforward ANN and other times just to networks made up of many layers of convolution layers (with threshold activation).

A minimum of three layers of nodes makes up an MLP: the input layer, the hidden layer, and the output layer. Each node, except for the input nodes, is a neuron that employs a nonlinear activation function [4]. Backpropagation is a supervised learning method that is used by MLP during training. MLP is distinguished from linear perceptron by its multiple layers and non-linear activation [5]. It can distinguish data that cannot be separated linearly (Figure 6).

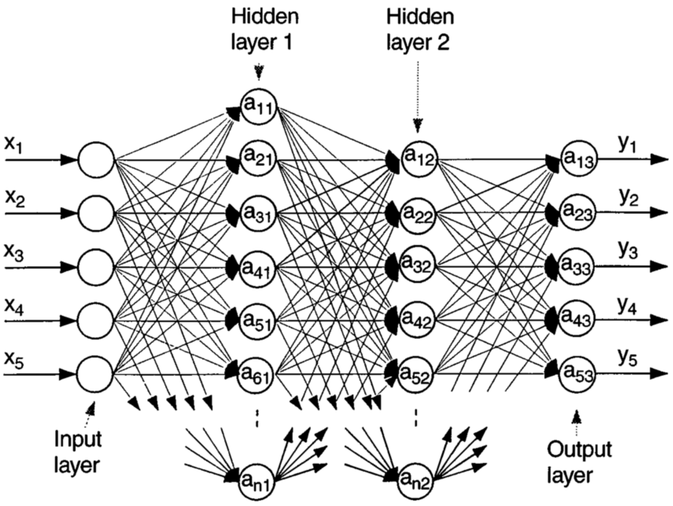


Figure 6. Multilayer Perceptron for classification.

Feature Selection and Prediction Model

To optimize prediction by the classification model, we need to feed the model with the most decisive feature. We extract 39 geometrical and clinical features; thus, for feature ranking, we utilize the technique of recursive feature elimination (RFE). A multilayer perceptron (MLP) of two fully connected layers is used for risk prediction. In addition to the MLP, risk estimation has been performed using other state-of-the-art models such as SVM, LR and RF.

EXPERIMENTS AND RESULTS

Dataset

This study has been performed on datasets of Brain CT scans and their corresponding segmentation of hemorrhage affected tissues. We use raw CT images obtained from the Government Hospital, Chennai. These CT images are for two-time different time frames, T1 and T2, where T2 is 24 h ahead of T1. We have paired non-contrasted CT scans of 79 patients (158 CT images) and additionally segregate these 79 CT image pairs into 59 pairs for training images and 20 pairs for validation.

The hemorrhage growth rate is quantified by computing the difference in the volumes of the hemorrhage tissues present at the two-time instances. The rate of expansion of hemorrhage is defined as [6]:

$$\delta = \frac{V_{T_2} - V_{T_1}}{V_{T_1}} \tag{1}$$

The blue line indicates the threshold for volume expansion rate (0.33). The red and green lines denote the volume expansion rates of critical patients and benign patients, respectively.

Where, δ, V_{T_2} and V_{T_1} indicate the Growth Rate, volume of hemorrhage at time T_2 (24 h ahead of T_1) and volume of hemorrhage at time T_1 respectively. From the Eq. (1), we evaluate the value of δ. If the value of the growth rate (δ) is greater than 0.333, the subject is said to be a severe victim of hemorrhage. We assign the patient a label of '1' for critical cases and a label of '0' for mild cases. Out of the 79 CT image pairs that constitute the dataset, we have a total of 17 cases where the patient's condition is critical. These cases are distributed into three different folds to conduct cross-validation experiments for the ICH estimation. Figure 7 exhibits the volume expansion rates (δ) of all the patients [7].

Stability Estimation

Various geometric features are extracted from segmented tumor, suppressing the non-significant features are fused with clinical feature to train our MLP for the stability prediction and risk score regression. The model performance is measured by using different evaluation metrics such as accuracy, sensitivity, specificity, precision, and Area Under Curve (AUC). Table 1 represents the prediction evaluation of the individual fold in 3-fold cross-validation. Our model achieves 78.33% stability prediction accuracy with the precision of 82.9%. The models are assessed with other state- of-the-art classification models given in Table 2. The table infers that SVM produces similar performance as MLP. Furthermore, the growth rate is subjected to regression analysis, and performance is compared. The Mean Square Error (MSE) loss and standard deviation is computed for all these models and visualized in Figure 8.

Table 1. 3-Fold cross-validation on 15 selected features for ICH stability estimation.

	Accuracy	**Specificity**	**Sensitivity**	**Precision**	**AUC**
Fold 1	0.80	0.933	0.400	0.789	0.53
Fold 2	0.75	0.933	0.200	0.777	0.52
Fold 3	0.80	1	0.7895	1	0.65
Average	0.7833	0.9533	0.4631	0.829	0.56

The variations of accuracy and precision after increasing the numbers of features are illustrated in the Figure 9. The figure shows that adding number of the features in the model training can improve the prediction accuracy. It is observed that the accuracy saturated or degraded after feature number is more than 15.

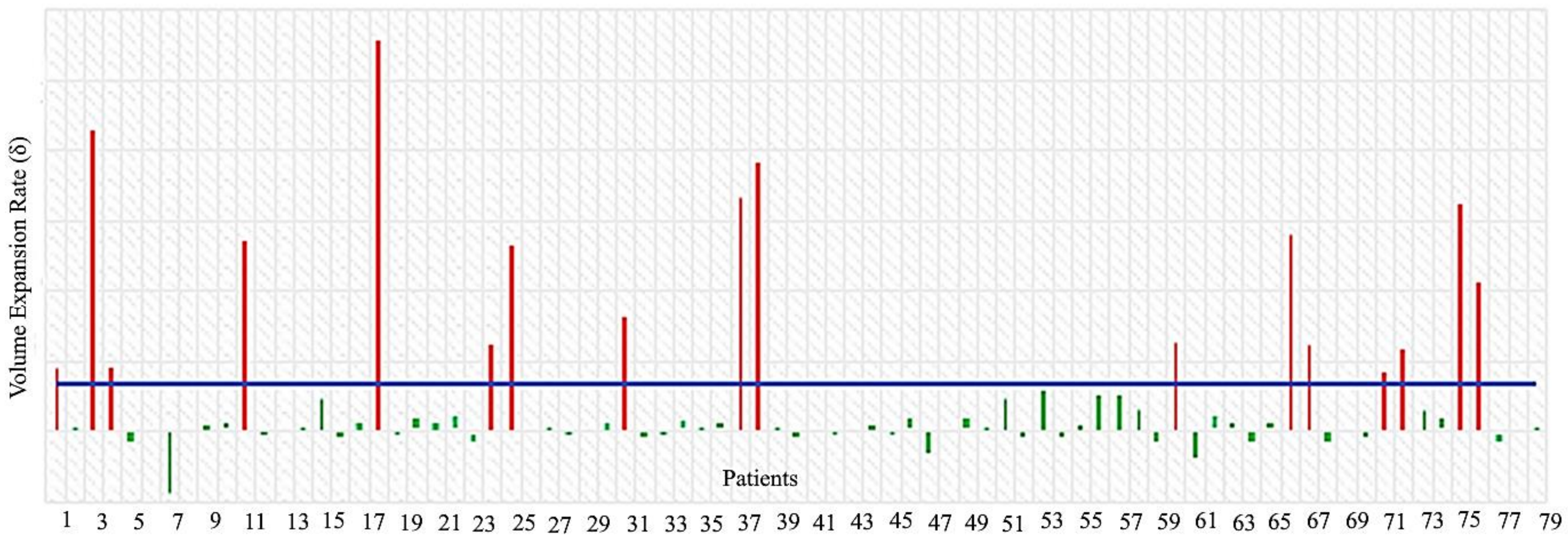


Figure 7. Plot of the volume expansion rates of all the patients used for Risk Estimation in our dataset.

Table 2. Performance comparison among different classification models.

	Accuracy	**Specificity**	**Sensitivity**	**Precision**
MLP	0.783	0.955	0.463	0.829
SVM	0.783	0.822	0.666	0.584
RF	0.750	0.883	0.350	0.562
LR	0.700	0.777	0.466	0.444

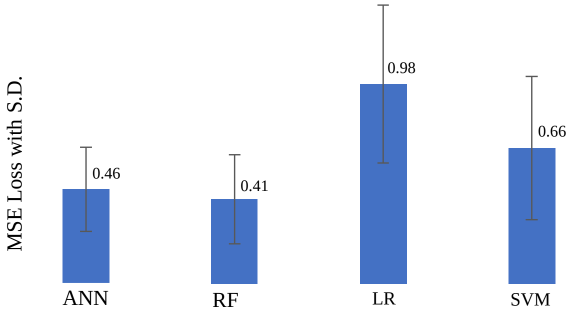


Figure 8. Representation of the variation of MSE loss and standard deviation for various models used in Regression Analysis.

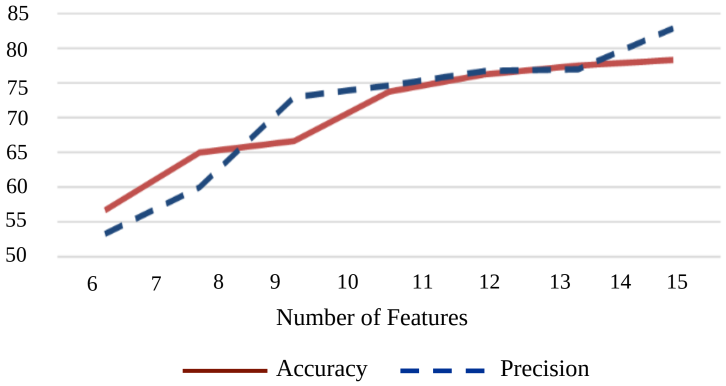


Figure 9. The variation of Accuracy and Precision with the increase in the number of features.

The SHAP analysis and ROC curve of our model prediction are demonstrated in the Figures 10 and Figure 11. The impact of the most significant features to the model output is shown in Figure 10 which shows high and low values as red and blue, respectively, e.g., High CTA spot values have a high impact for the model prediction [8]. Further, from the SHAP analysis, it is inferred that the most crucial feature for the ICH risk scoring is the clinical feature and CTA spot sign. The plot is sorted according to the sum of SHAP value magnitudes over all the samples and the most significant eleven features are shown in Figure 10. Besides, features like location and first axis length of the ICH voxel, solidity, Kurtosis, and entropy are the significant pieces of information to calculate ICH score [9].

DISCUSSION

In existing system, only clinical features are exploited to calculate the risk score for the ICH treatment. However, clinical scoring system can be improved by integrating geometric features of the segmented ICH voxel. Table 3 represents the enhancement of the accuracy from 71.6 to 78.3% in fused clinical and geometric features. To incorporate prediction, we use SHAP value analysis [10], which assigns each feature an importance value for a particular MLP prediction. The table helps in concluding that combining both clinical and geometric features, predicts better stability estimation.

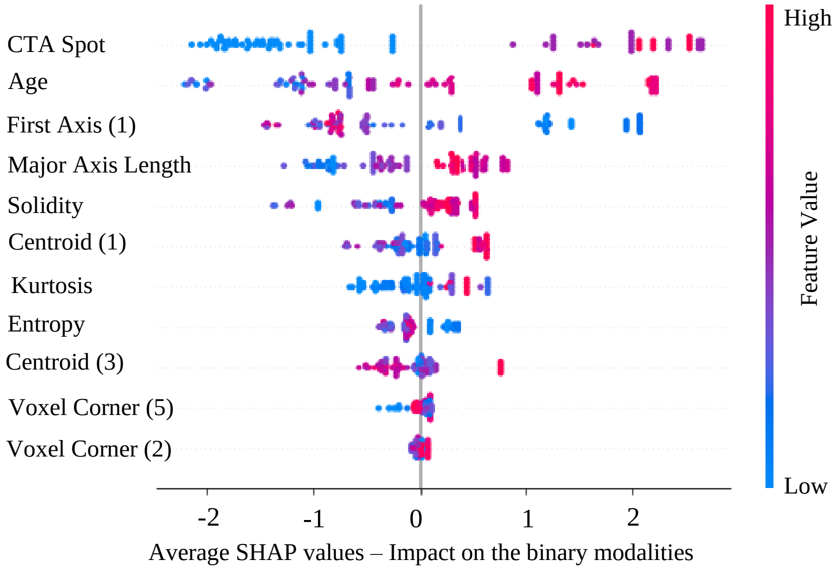


Figure 10. Importance of feature vector (Geometrical and Clinical) for stability estimation.

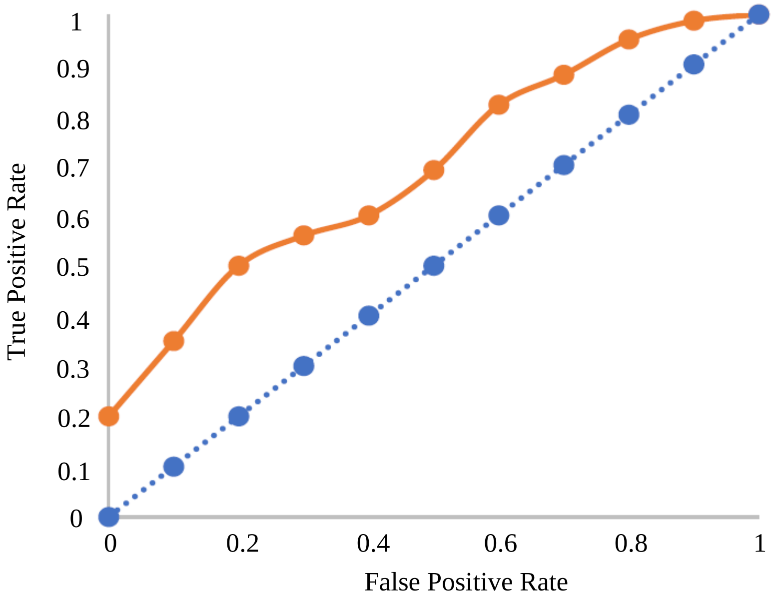


Figure 11. Receiver Operating Characteristic curve (ROC curve), illustrating the diagnostic ability of the proposed binary classifier on prediction of stability of hemorrhage.

Table 3. Ablation study between clinical and geometric features for stability estimation.

Clinical	Geometrical	Accuracy	Specificity	Sensitivity	Precision
✓	×	0.716	0.828	0.4	0.872
×	✓	0.733	0.844	0.377	0.791
✓	✓	0.783	0.955	0.463	0.829

CONCLUSION

In the risk estimation task, stability prediction has been enhanced and risk scoring models have been developed by training the MLP for the classification and regression tasks respectively. The classification models predict whether a case is stable or not while regression model estimates exact expansion rate of the ICH. To illustrate our models, the significant features using SHAP analysis are visualized. This aligns with clinical intuition and additional risk factors to calculate the ICH risk score. In future, an end-to-end multi-task learning model can be developed to segment and estimate the ICH using larger dataset.

REFERENCES

1. Brouwers HB, Greenberg SM. Hematoma expansion following acute intracerebral hemorrhage. Cerebrovasc Dis. 2013; 35(3): 195–201.
2. Karim MR, Cochez M, Zappa A, Sahay R, Rebholz-Schuhmann D, Beyan O, Decker S. Convolutional embedded networks for population scale clustering and bio-ancestry inferencing. IEEE/ACM Trans Comput Biol Bioinform. 2020 May 18; 19(1): 369–382.
3. Brouwers HB, Chang Y, Falcone GJ, Cai X, Ayres AM, Battey TW, Vashkevich A, McNamara KA, Valant V, Schwab K, Orzell SC. Predicting hematoma expansion after primary intracerebral hemorrhage. JAMA Neurol. 2014 Feb 1; 71(2): 158–64.
4. Li Q, Huang YJ, Zhang G, Lv FJ, Wei X, Dong MX, Chen JJ, Zhang LJ, Qin XY, Xie P. Intraventricular hemorrhage and early hematoma expansion in patients with intracerebral hemorrhage. Sci Rep. 2015; 5: 11357.
5. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems. 2017; 4765–4774.
6. Liu T, Fan W, Wu C. A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset. Artif Intell Med. 2019 Nov 1; 101: 101723.
7. Ng Y, Qi W, King NK, Christianson T, Krishnamoorthy V, Shah S, Divani A, Bettin M, Coleman ER, Flaherty ML, Walsh KB. Initial antihypertensive agent effects on acute blood pressure after intracerebral haemorrhage. Stroke Vasc Neurol. 2022 Apr 20: svn-2021-001101.
8. Selvaraju P. Predicting future problem gamblers using Machine Learning Algorithms. Independent thesis Advanced level (degree of Master (Two Years). Uppsala University; 2022.
9. Cao R, Fang Z, Jin M, Shang Y. Application of machine learning approaches to predict the strength property of geopolymer concrete. Materials. 2022 Mar 24; 15(7): 2400.
10. Ahmad Wan Muhammad Amir Wan, *et al*. An Integrated Biostatistical Approach To Reveal The Health Status Among Elderly People At Receiving Home Care. Int J Sci Technol Res. 2020; 9(1): 335–339.