Iranian Journal of Medical Physics

ijmp.mums.ac.ir

A Machine Learning Approach for Differential Diagnosis of Vascular and Non-Vascular Intracranial Hemorrhage in Non-Contrast CT Images

Sahar Faraji[1], Bita Abbasi[2], Amin Amiri Tehranizadeh[3], Zeinab Naseri[3], Lida Jarahi[4], Fatemeh Khojasteh Rahimi[2], Shahrokh Nasseri[1], Azadeh Hashemi[2*]

1. Department of Medical Physics, Faulty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.
2. Department of Radiology, Faculty of Medicine, Mashhad University of Medical sciences, Mashhad, Iran.
3. Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.
4. Department of Community Medicine, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.

ARTICLE INFO

Article type:
Original Paper

Article history:
Received: Apr 04, 2024
Accepted: Sep 14, 2024

Keywords:
Intracranial Hemorrhages
Machine Learning
Computed X ray
Tomography

ABSTRACT

Introduction: Accurate diagnosis of acute Intracranial Hemorrhage (ICH) involving vascular and non-vascular bleeding has proven to be challenging due to the visual complexities in non-contrast Computed Tomography images (NCCT). Consequently, there has been a necessity for the adoption of novel techniques to address this issue, recently. This study aims to develop a new framework for automatic and accurate diagnosis of ICH and the ability of machine learning to differentiate vascular and non-vascular causes of Intracranial hemorrhages based on CT scan images without contrast material. Determining whether intracranial hemorrhage is vascular or non-vascular is clinically significant as it influences treatment decisions.

Material and Methods: In this retrospective study, NCCT images were gathered from a group of 370 patients, comprising 67 subjects with vascular bleeding and 303 with non-vascular bleeding. Radiomics features encompassing morphological, texture, and intensity-related characteristics, were extracted for every image slice. Subsequently, the effectiveness of five classification methods—namely, Support Vector Classifier (SVC), Logistic Regression (LR), Random Forest (RF), Decision Tree (DT) and K-Nearest Neighbors (KNN) was evaluated.

Results: Metrics for evaluating classification methods, sensitivity, specificity and accuracy for the Logistic Regression were 55%, 65% and 63%, respectively. The AUC-ROC in this model was 0.66, which is better than other methods with large margin.

Conclusion: In this study, an evaluation of five different classification methods revealed that all of them exhibited sufficient level of specificity. However, when it comes to classification sensitivity and accuracy, the Logistic Regression approach outperformed the others.

► *Please cite this article as:*
Faraji S, Abbasi B, Amiri Tehranizadeh A, Naseri Z, Jarahi L, Khojasteh Rahimi F, Nasseri Sh, Hashemi A. A Machine Learning Approach for Differential Diagnosis of Vascular and Non-Vascular Intracranial Hemorrhage in Non-Contrast CT Images.. Iran J Med Phys 2025; 22: 65-76. 10.22038/ijmp.2025.79158.2400.

Introduction

Stroke is the second most frequent cause of death and one of the main causes of disability worldwide [1]. Due to improved life expectancy, the burden of stroke will increase [2]. An essential type of stroke, Intracranial hemorrhage (ICH), accounts for approximately 10% to 20% of all strokes. The morbidity and mortality resulted from ICH is higher than ischemic stroke [3]. Hemorrhagic stroke caused by the rupture of a blood vessel or an abnormal vascular structure while ischemic stroke is resulted from interruption of the blood supply to the brain [4]. According to the location of bleeding, ICH can be divided into two subgroups, vascular and non-vascular ICH. Determining whether intracranial hemorrhage is vascular or non-vascular is clinically significant as it influences treatment decisions.

Understanding the risk factors for ICH is crucial due to its increasing incidence with age. Studies have shown that the risk of ICH rises significantly in older populations, with individuals over 85 years old being nearly 10 times more likely to experience ICH compared to those aged 45 to 54 years [5]. Major risk factor for ICH include high blood pressure and amyloidosis [6,7], while other contributing factors are smoking [8], alcoholism [9], low cholesterol [10], Diabetes mellitus [11], ICH-related genes [12], and chronic kidney disease [13]. Identifying these risk factors helps in understanding the underlying cause of the hemorrhage, whether vascular or non-vascular,

*Corresponding Author: Tel: +98-: 9150062310; Email: HashemiAZ981@mums.ac.ir

which is essential for effective management and treatment.

The signs and symptoms of ICH and ischemic stroke are similar, usually involving the sudden onset of focal neurological deficits. Decreased level of consciousness, vomiting, headache, seizures and very high blood pressure may indicate the presence of ICH. However, none of these signs are specific enough to differentiate hemorrhagic from ischemic stroke at diagnosis, and therefore the diagnosis of ICH always relies on neuroimaging [14,15]. The neuroimaging methods that used in diagnosis of ICH include non-contrast computerized tomography (NCCT) [16], CT angiography (CTA) [17] and magnetic resonance angiography (MRA) [18]. NCCT is considered the gold standard for the diagnosis of ICH due to excellent sensitivity for the detection of acute ICH. Also, NCCT can provide useful elements such as location, intraventricular extension of ICH, hydrocephalus, presence and extent of edema, and midline shift or brainstem compression secondary to a mass effect caused by hematoma. CT angiography is the most accessible and non-invasive method to detect vascular abnormalities as secondary causes of ICH.

In a retrospective study, W. Strub et al. [19] interpreted 22,590 head CT images. Among these images, a total of 1037 discrepancies were observed. 141 cases of bleeding causes that was not mentioned in the initial report or was incorrectly explained. The most common pattern of bleeding that was wrongly observed in the initial report (55 cases) was subdural bleeding. Of these 55 cases, 8 cases (15%) were diagnosed as false positives. In addition, there were 2 cases in which are correctly described as presence of bleeding but incorrectly described it as subdural hematoma instead of epidural (1 case) or subarachnoid hemorrhage (1 case).

A major barrier to critical imaging results is related to the delay in radiologist report with an immediate significant finding that must be reviewed from a large number of patients. Approximately, 40% of all imaging of hospitalized patients are determined as urgent cases according to the clinician's order [20], but currently there is no automatic method for active triage of these examinations based on the real urgency of the imaging findings. Many methods have already been developed to classify critical imaging findings in head NCCT images. However, the complexity of constructing and implementing these algorithms, along with their limited clinical focus, may have caused their limitations. Recently, artificial intelligence using different techniques has attracted wide attention in medical fields [21, 22,23]. Artificial intelligence with machine learning (ML) algorithms can be very accurate in detecting vital findings such as bleeding, pressure effect and hydrocephalus in head NCCT [24]. Deep learning (DL) is a computing model with multi-processing layers, in which raw input is used for progressively learning. Some DL architectures include Convolutional Neural Networks (CNN) and Recurrent Neural Network (RNN) which are mostly used to solve processing problems [25].

Identifying the location and type of any bleeding in CT scan images is an important step in patient treatment. Diagnosis of this disease requires immediate intervention that is complex and often time-consuming [26]. In order to help clinicians and radiologists save precious time and diagnose more accurately and rapidly, computer-aided diagnostic systems have been developed to process images and diagnose ICH. In recent years, many methods have been published using deep learning models to diagnose ICH [27,32].

In Altuve et al. [33], the authors compared the head scans of 100 patients with intracranial hemorrhage and 100 healthy individuals using a model based on the residual neural network and the ability of this network to detect Intracranial bleeding was checked. Finally, it was found that this model had 96% sensitivity and 95% specificity in diagnosing ICH.

Some studies have made efforts to diagnose head CT abnormalities including ICH, using ML methods. However, it remains unknown whether such simple approaches (2D, hybrid or simple 3D) are capable of making reliable predictions. In addition, none of these studies attempted to differentiate the causes of ICH. Accurate differentiation between vascular and non-vascular hemorrhages in non-contrast CT images can be challenging. Subtle differences in bleeding patterns and structural features can be difficult to discern, especially when the hemorrhage is extensive and complex. CT angiography (CTA) and magnetic resonance angiography (MRA) can be helpful, but they also have limitations and may require more detailed analysis and integration with other diagnostic methods. Therefore, this study aims to develop a new framework for automatic and accurate diagnosis of ICH and the ability of ML to differentiate vascular and non-vascular causes of ICH based on CT scan images without contrast material. The development of such automated frameworks could improve diagnostic accuracy and reduce human error.

Materials and Methods

Study population

In this retrospective study, CT scan images of 67 patients with vascular bleeding and 303 patients with non-vascular bleeding were used. The selection of these patients was done by a radiologist. Images in DICOM format were extracted from the PACS system of Imam Reza Hospital, Mashhad, Iran. The demographic information of the patients is presented in Table 1.

Table 1. Demographic Characteristics of Study Patients

	Vascular	Non-Vascular
Number of Male	34	172
Number of Female	33	131
Averge Age	51	56
STD. Age	14	15

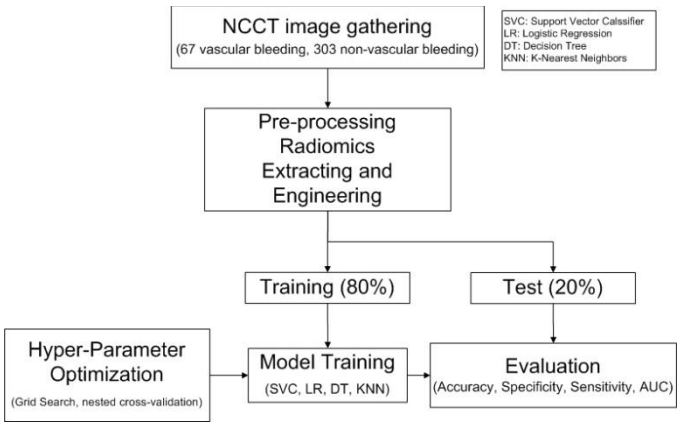

Figure 1. Analysis flow for the development and evaluation of models.

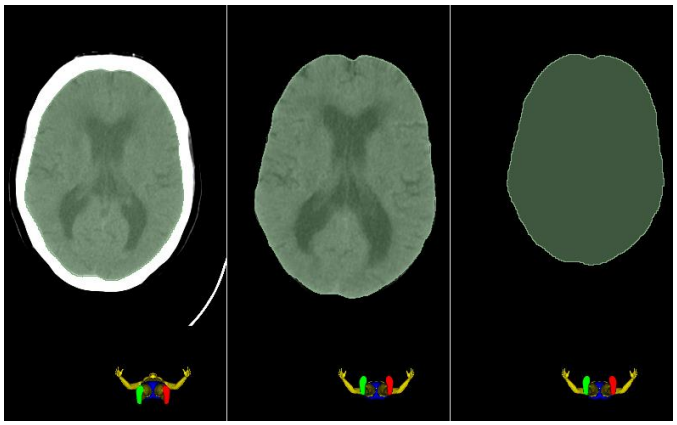

Figure 2. Axial sections of CT images, extracted brain tissue and corresponding segment

In figure (1), analysis flow for the development and evaluation of models can be seen.

Pre-processing

We performed the following pre-processing steps to prepare our images for feature engineering, each of which will be explained in detail.

- Format conversion and motion artifact removal: dcm2niix software (version v1.0.20220720) was used to process the images, including correcting the head position during imaging and removing the skull bone.
- Removal of artifacts: In order to eliminate potential artifacts, such as metallic artifacts, the pixel range of all images was set to [-1024, 3071]. For this purpose, fslmaths was used in the FSL comprehensive library.
- Removal of the skull bone: In order to extract the brain tissue or remove the skull, the thresholding method, interval [0, 100], was used. The tool used was fslmaths. Then the BET tool was used in FSL. The fractional intensity threshold value in the BET tool was set to 0.5. In order to extract the features in the whole brain, the volume of the whole brain tissue needs to be volumetrically labeled in the output of the previous step.
- Creating a CSV file containing the list of patients: Extracting the characteristics of patients, one by one and separately, is an exhausting and time-consuming task. In such cases, the batch mode method can be used, in which the name and data path of the patients, including the CT file and label Map Volume in .nii.gz format, are stored in a CSV file. In order to prepare this file, the Pandas library was used.
- Implementation of the file containing the list of types of features and pre-processing of images

The list of the type of features to be extracted from the images was stored in a text file, with the extension yaml.

- Run Pyradiomics in Batch Mode to extract features.

The result of Pyradiomics is a file containing features. This file contains the extracted characteristics of patients with vascular bleeding. In this way, 1327 features were extracted by applying wavelet transform. These features were saved in a CSV file. It should be emphasized that this number of features was made possible with the help of wavelet transformation and the extraction of features in different spatial resolutions.

In Figure (2), the axial sections of the CT images, the extracted brain tissue and the corresponding segment can be visualized using 3DSlicer.

Features engineering

In the study, various Python libraries were utilized for machine learning tasks, including NumPy, scikit-learn, imbalanced-learn, pandas, boruta, seaborn, mrmr, scipy, and matplotlib. These libraries were employed for tasks such as data manipulation, model implementation, evaluation metrics, and visualization in the software environment.

These features are divided into the following classes (four classes):

- Shape-based features
- First-order statistical features
- Second-order statistical features
- Higher order statistical features

Before extracting the features determined by PyRadiomics, the images were processed. This step represents an attempt to homogenize the images from which features are extracted according to properties such as pixel spacing and gray level intensities. In the text file, with the YAML extension, in addition to determining the type of features, the settings for image processing were also determined, which includes applying Wavelet and LoG filters to the images. LoG (Laplacian of Gaussian) filter was used to smooth the images. Also, wavelet transform was used to extract features in surfaces with different spatial resolutions.
The CSV file created in the previous step contains the characteristics of two groups vascular, non-vascular and a column called Outcome (zero and one). The last column is the target for each patient. A part of this data is considered as the input of machines. In order to learn machines, this data was divided into two parts X (characteristics of two groups) and Y (Outcome column). As mentioned, the total number of extracted features was 1327, which, after removing non-scalar features, resulted in 1288 remaining features (columns of matrix X). The number of features in each class are as follows: Shape-based features (14 features), First-order statistical features (18 features), Second-order statistical features (73 features), and Higher order statistical features (1183 features).

The nested cross-validation method was used in order to create training, validation, and test data, as well as to prevent overfitting [34]. In the nested cross-validation method, an internal cross-validation loop was used to perform data transformation and hyperparameter optimization. Next, this loop is placed inside an external cross-validation loop to evaluate the transformed data and the optimized model in different test sets. This outer loop is for the purpose of testing the model and provides the possibility of approximating the performance of selected models. In this study, the outer loop was divided into 3 parts with 3 repetitions, which are 2 parts for training and 1 part for testing in each repetition. So, it covers the entire data set. The inner loop was divided into 3 parts with 3 iterations, which are 2 parts for training and 1 part for validation in each iteration. In each external validation loop, the training data was first used to normalize the features, select the effective features, and balance the classes with increasing data. Figure 3 shows a diagram of cross-validation.

After these steps, five classification algorithms were called and the performance of five algorithms was evaluated in each outer loop. In order to check the performance of algorithms and choose the best model with the best hyperparameters, GridSearchCV (Grid Search Cross-Validation) method was used. In this method, the internal validation loop is applied to the training data of the external validation loop. Then it evaluates the performance of the best model on external validation loop test data with accuracy, sensitivity, specificity and AUC (area under the Receiver-Operating Characteristic curve) indicators. Finally, five models were presented for each part of the external loop. The final selection of the best model is based on the highest average AUC-Score (Area Under the Curve Score) which is calculated from the three external loop sections. In the following, more explanations are given for each of the stages of machine learning.

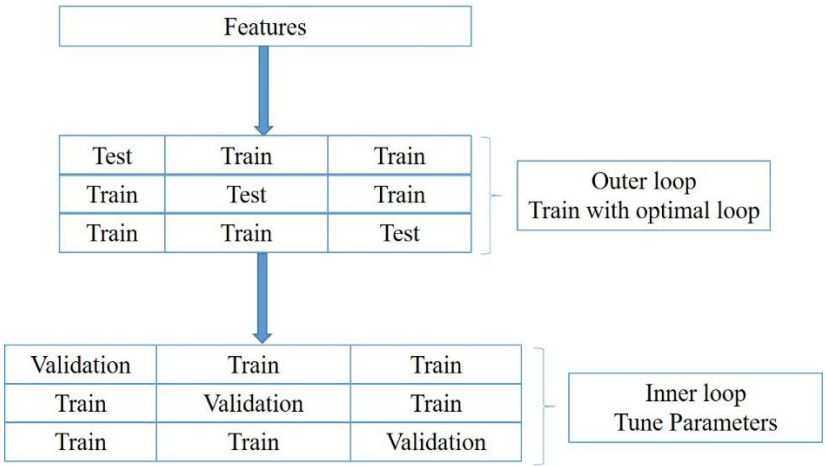

Figure 3. Diagram of Cross-validation

The range of values in different features is not the same. When the dataset contains features (x) that vary greatly in magnitude and unit, it becomes important to standardize the feature values. Otherwise, this issue will cause instability of the model and decrease the accuracy of model estimation. To avoid this problem and to normalize the data, the MinMaxScaler (Min-max normalization) method was used. In this method, the features are transferred to the range of zero to one.

Many extracted features can be redundant or irrelevant to the purpose of the study. Reducing the number of features increases the learning speed and improves machine performance. In other words, this feature reduction also facilitates the understanding of the classified problem and improves the performance of the model. In this study, the selection of the most effective features was done using two methods, Minimum Redundancy and Maximum Relevance (MRMR) and Recursive Feature Elimination Cross validation (RFECV). The MRMR algorithm usually selects a subset of features that have the highest correlation with a class (relevance) and the lowest correlation among themselves (redundancy) [35, 36]. At first, RFECV fits using all the features then the less important features gradually remove until the desired number of features remains [37, 38]. In this study, the optimal features were selected according to the following steps:

a) Using the MRMR method, first 37 optimal features were selected out of 1288 features.

b) Then to select the most optimal features using the RFECV method among 37 features, the minimum number of features was 30 and the SVC selector algorithm was considered. Also, in this method, ROC-AUC index was used to select features.

The number of patients in both vascular and non-vascular groups is not the same. The challenge of working with such an unbalanced dataset is that in most machine learning techniques, the minority class is ignored during training. This will lead to poor performance of cars in the minority class. Synthetic Minority Oversampling Technique (SMOTE) for short was used in this research. This algorithm creates additional artificial samples by linear combination of existing samples to balance the minority class [39, 40]. In this technique, the number of neighbors of the minority class was selected equal to 5 (k_neighbors = 5).

Algorithms Model

A classification algorithm can be a supervised learning technique that is used to identify new patterns based on training data. A classification algorithm learns from a given data set or samples and then classifies the new samples into a number of classes or groups. Five classification models were used in this study:

1- Support Vector Classifier (SVC)
2- Logistic Regression (LR)
3- Random Forest (RF)
4- Decision Tree (DT)
5- K-NearestNeighbors (KNN)

Figure 4 shows the machine learning data flow.

The performance of each of these models depends on parameters that are called metaparameters. The values of these parameters have a great impact on the performance of the machine. One of the common methods of obtaining the best hyperparameters for a machine learning model is the GridSearchCV method.

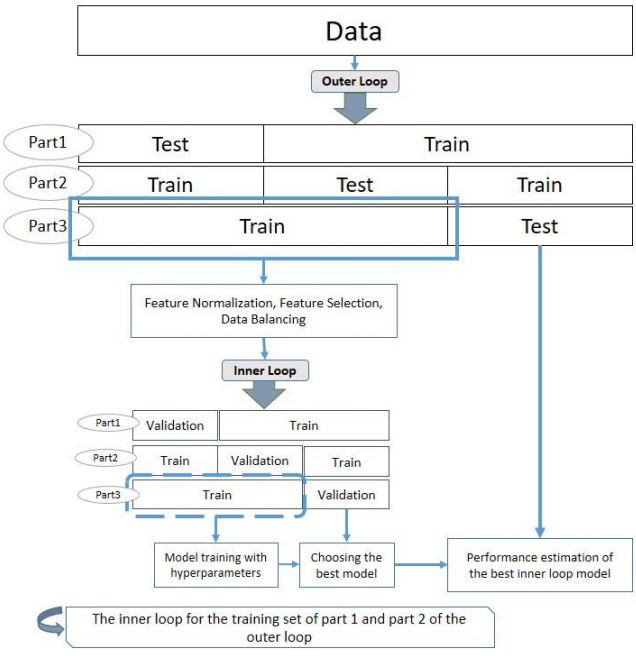

Figure 4. Machine Learning Data Flow

In this study, the input of this algorithm includes the five machine learning models mentioned above, the type of hyperparameters of these models, cross-validation and the use of accuracy index to evaluate the performance of the model. The optimal parameters of the model used by applying this method were obtained through cross-validation. Finally, the performance of the best models with the best meta-parameters was evaluated on the outer loop test data. The evaluated indicators for the models are given in the next section.

Model performance evaluation criteria

After identifying the final model in each outer loop, the following criteria were measured to evaluate the prediction performance of the supervised models. These criteria include: Confusion Matrix, Sensitivity, Specificity, Accuracy, Receiver Operating Characteristic Curve (ROC) and Area Under Curve (AUC).

Algorithms evaluation

To achieve optimal model performance and fine-tune its hyperparameters, we employed a nested cross-validation approach. The model's evaluation involved a 10-fold cross-validation, and in the final step, we predicted the model performances on the test dataset to evaluate the generalizability and the potential over-fitting problems by calculating key metrics including Accuracy, Specificity, and Sensitivity. To further assess the model's performance, we determined the AUC based on the ROC curve.

Results

Data normalization

The following curves, Figure (5), show the effect of normalization on the training characteristics of each part of the outer loop. As can be seen, before normalization, the values of the features are in different intervals and the number of features in each of these values is very scattered, after applying this method, the values of the features will be in the range of zero and one and the number of features in this range are concentrated. This concentration indicates improved stability and consistency in the training process, as normalization ensures that all features contribute equally to the model, enhancing overall accuracy and reducing model instability. In other words, by standardizing the feature values to a uniform scale, the MinMaxScaler effectively mitigates issues related to varying magnitudes and units, leading to a more reliable and precise model estimation.

Choosing the optimal feature

Among 1288 features, at firstly 37 features were selected by MRMR method. Then, with the RFECV method, at least 30 features were selected from among 37 features. Table 2 shows the number of features selected from the training features of each part of the outer loop. The outer loop was divided into 3 parts. In this table, it means of each section of the outer loop. As shown in Figure 4, both the outer and inner loops have 3 parts, folds, or loops. So, in Table 2, the term "part" refers to the parts associated with the outer loop. In each part of the outer loop, the best radiomics features were selected, and the number of these selected features in each part of the outer loop is listed in Table 2.

Table 2. Number of the most effective features selected

part	Number of selected effective features
1	33
2	30
3	31

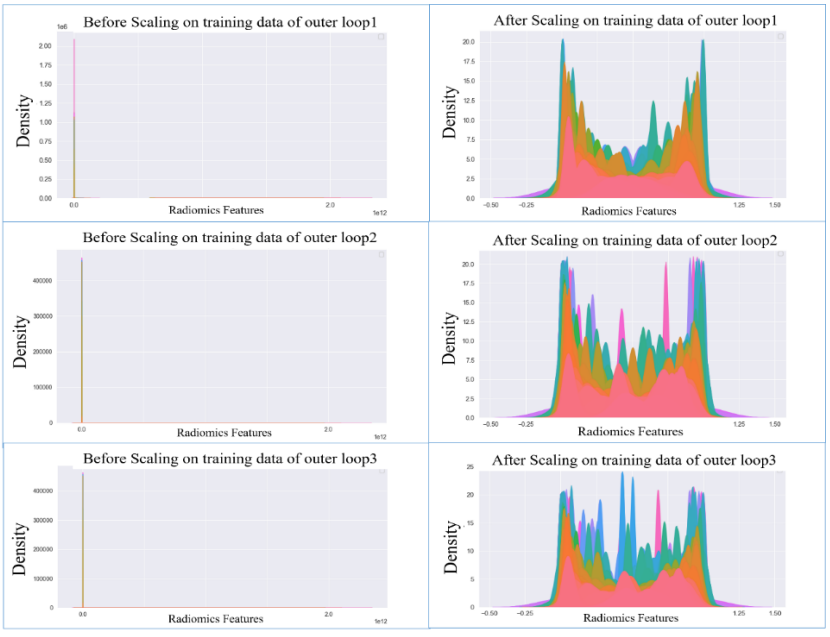

Figure 5. Impact of Normalization on Feature Distribution

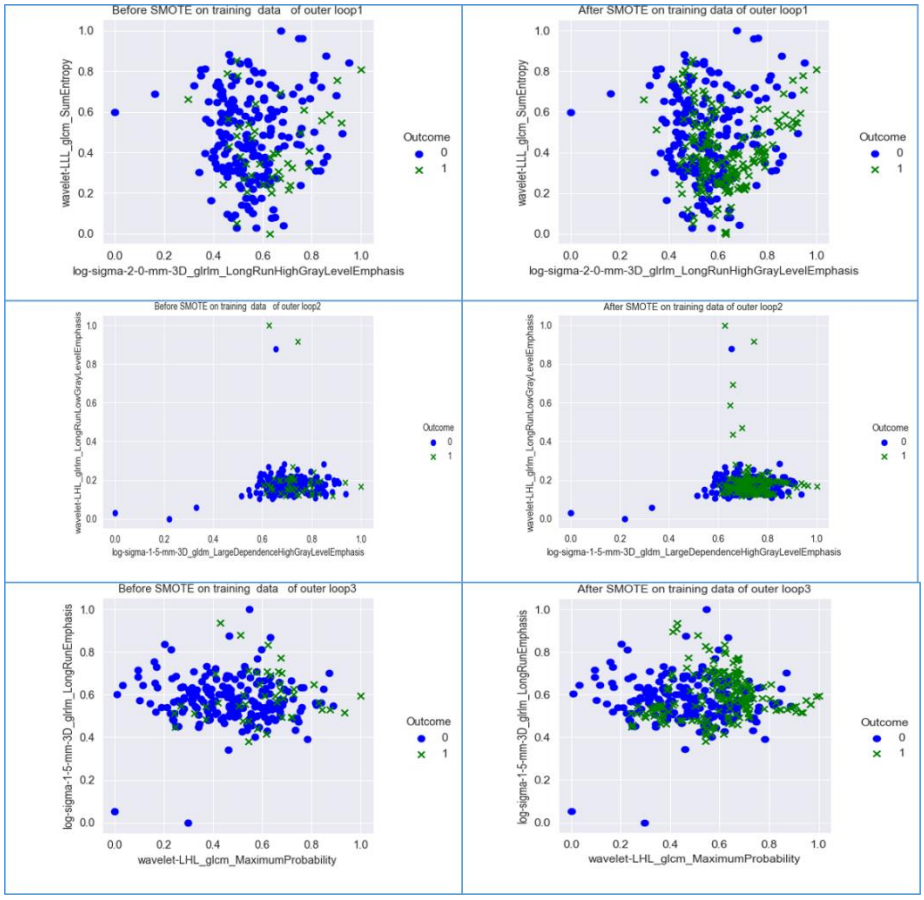

Figure 6. The effect of SMOTE method on the data

Data balancing

Training data in the first, second, and third parts of the external loop have 246 patients. Of these, 202 patients are in class zero (non-vascular) and 44 patients are in class one (vascular). After applying the SMOTE method with the number of neighbors (k) 5, the number of training target data for the one, two and three outer loop segments reached 404 patients; That is, with this method, 158 patients have been added to the target data with class one. The effect of the SMOTE method on the training target data of each part of the outer loop can be seen in the following image, Figure (6).

Modeling and evaluation of model performance

In this part, firstly, the results related to the average AUC calculated from all three parts of the outer loop for all machine learning algorithms predicting two vascular and non-vascular groups are expressed. Then the best predictive algorithm that had the highest AUC average value compared to other algorithms will be selected. Also, the details of the indicators evaluated for this algorithm (including sensitivity, accuracy, specificity, etc.), the best hyperparameters used for this algorithm, and a schematic of the ROC curve and the disturbance matrix in this model are presented. Table 3 shows information about the average AUC of all three parts of the outer loop of predictive classification models that were made from CT features.

Table 3. Average AUC of all three parts of the outer loop for predictive models

Features	Model	AUC ± SD
CT	LR	0.62 ± 0.04
	KNN	0.57 ± 0.06
	DT	0.52 ± 0.06
	SVC	0.60 ± 0.02
	RF	0.61 ± 0.02

Table 4.The interpretation of the AUC [41]

Model Performance	Value
Poor	0.5 - 0.6
Fair	0.6 – 0.7
Good	0.7 – 0.8
Very Good	0.8 – 0.9
Excellent	0.9 – 1.0

As can be seen in Table 3, the LR algorithm showed the highest value of the CT features from the point of view of the average AUC value, with an average AUC equal to 0.62 ± 0.04 among the models compared with it. AUC can rank models based on overall performance, so AUC is

considered in evaluating models. The information related to the interpretation of the AUC obtained for the models [41] is listed in Table 4.

As a result, this model will be selected as the best predicting model for both vascular and non-vascular groups. Next, the results related to this best model will be shown.

Table 5 shows the information related to the measured indicators, including accuracy, sensitivity, specificity and AUC related to each part of the outer loop for the best predicting algorithm for vascular and non-vascular groups.

In Table 6, the information related to the best predictor classification model, with other models, obtained from Table 3 is presented with the best hyperparameters. The specifications of the hyperparameters that achieved the best performance in the models are as follows in Table 6.

In the LR model, the parameters C and solver respectively correspond to the inverse of regularization strength and the optimization algorithm. In the KNN model, n_neighbors is a parameter that specifies the number of neighbors. In the DT model, the parameters max_depth, criterion and splitter correspond to the maximum depth of the tree, the function used to measure the quality of a split, and the strategy employed to choose the split at each node, respectively. In the SVC model, the parameters C and probability correspond to the regularization parameter and enabling probability estimates, respectively. In the RF model, the parameters criterion and max_depth correspond to the function to measure the quality of a split and the maximum depth of the tree, respectively.

Figure (7) shows the schematic curves of ROC and AUC of each of the outer loop sections along with the average AUC of all three outer loop sections for the best prediction algorithm of vascular and non-vascular groups.

Table 5. The measured indices of each part of the outer loop for the best predictive algorithm

Features-Model	Part	AUC	Accuracy	Sensitivity	Specificity
CT-LR	1	0.66	0.63	0.55	0.65
	2	0.56	0.6	0.27	0.67
	3	0.63	0.68	0.50	0.72

Table 6. The best hyperparameters for the best model

Model	Hyperparameters
LR	C=1000.0, solver='liblinear'
KNN	n_neighbors=2
DT	max_depth=9, criterion='entropy', splitter='random'
SVC	C=1000.0, probability='True'
RF	criterion='entropy', max_depth=8

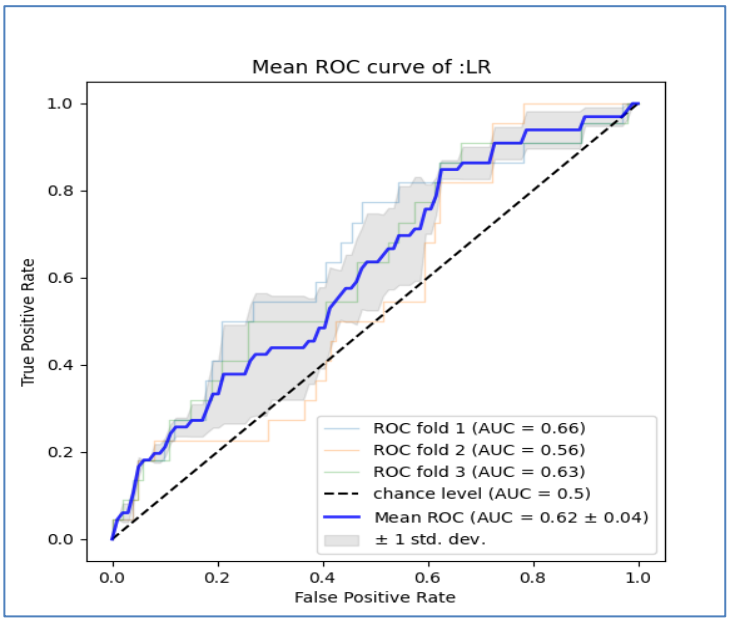

Figure 7. Schematic curves of ROC and AUC of each outer loop segment along with the average AUC

Discussion

Differential diagnosis of vascular and non-vascular bleeding in patients with ICH based on NCCT scan images is critical to reduce brain damage and guide the physicians in selecting the most suitable treatment approach for the specific type of stroke.

To do so, we curated and analyzed CT scan images of 370 patients with ICH, of which 67 had vascular, and the rest with non-vascular causes. To study the discriminative ability of the machine learning algorithms for the automatic prediction of the type of bleeding, five standard binary classification methods are employed. In total, over 1300 radiomics features were identified in the images, and 37 of them were selected in the final model. Among the five models, the logistic regression model was chosen as the best model, as it had the highest AUCROC. Ultimately, the best sensitivity and specificity were 55% and 65%, respectively, and the accuracy was also measured at 63%. The AUCROC in this model was 0.66, which is considered relatively good but indicates that this model still falls short of an ideal model that can accurately differentiate between vascular and non-vascular bleeding. To the best of our knowledge, there has been no study conducted to differentiate between vascular and non-vascular causes of ICH using machine learning techniques on whole brain tissue features in CT images. Therefore, we evaluated the results using some experiments on various machine learning and feature selection techniques. Alfaer et. al. [29] proposed a deep learning model named AICH-FDLSI in detecting ICH. Their model achieved a sensitivity ranging from 86% to 88%, with a specificity of 87%. Residual Neural network is utilized by Altuve et. al. [33] to recognize the bleeding type in ICH patients. The results showed that this model had a sensitivity of 96% and a specificity of 95%.

In Table 7, we present a list of the published results on the ICH classification task using machine learning methods, to recognize brain bleeding and also types of ICH. To ensure a fair comparison, we have also included some studies based on machine learning algorithms in Table 8. To compare the results of the proposed ICH classification method with those of other published studies, we were faced with a limited number of studies on the classification of vascular and non-vascular bleeding. In addition, other studies are often based on the automatic or manual segmentation of the bleeding regions, which are challenging and vulnerable to many systematic errors. These miscalculations can degrade the overall performance of the classification system. In the proposed system, the proposed machine learning models are trained using the extracted radiomics features from the segmented brain regions. Despite potential redundancy inside brain region, because of the simplicity of the utilized segmentation method, the proposed system could be more reliable and feasible in clinical settings. Deep learning models are powerful feature extractors and therefore have superior capability in medical image analysis systems. However, they require high resources and therefore the utilization of these systems is limited and challenging in current workflow. These models face challenges such as the need for large annotated datasets due to privacy concerns, the risk of overfitting with limited data, and difficulties in generalizing across diverse populations and imaging devices. Machine learning systems by using feature engineering techniques has comparable results in limited resource conditions. Therefore, in cases where we have limited data, machine learning models are better than deep learning methods in terms of Generalization and overfitting problems.

Table 7. Performance of the published studies on the classification of the ICH using Deep Learning techniques.

Method	Model	AUC	Accuracy	Sensitivity	Specificity
Schmitt [42]	CNN	0.89	0.91	NA	0.9
Hopkins [43]	DNN	0.99	0.98	NA	0.99
Seyam [44]	DL	0.939	0.872	0.93	NA
Altuve [45]	ResNet-18	NA	0.959	0.956	0.962
Tang [46]	CNN	NA	0.905	0.919	0.883
Cortes [47]	DL	0.978	0.927	0.914	0.94
Kau [48]	DL	NA	0.94	0.682	0.968
Tharek [49]	CNN	NA	0.95	0.969	0.931
Rao [50]	InceptionV3	0.988	0.989	0.974	0.986
Voter [51]	DL	NA	NA	0.923	0.977
Kuo [52]	CNN	NA	NA	1.00	0.90

Table 8. Performance of the proposed method and published results on the classification of the ICH using Machine Learning techniques.

Method	Model	AUC	Accuracy	Sensitivity	Specificity
Abe [53]	XGBoost	0.8	NA	74	74.9
Trevisi [54]	RF	0.93	83.55	77.52	86.29
Uchida [55]	LR	0.87	NA	27	97
Proposed method	LR	0.66	0.63	0.55	0.65

Conclusion

In this research, five different classification techniques including Support Vector Classifier, Logistic Regression, Random Forest, Decision Tree, and K-Nearest Neighbors were employed to identify vascular bleeding type in non-contrast CT images. Subsequently, their effectiveness was evaluated by comparing their performance. The findings indicated that logistic regression outperformed the other four methods in terms of sensitivity, specificity, and overall accuracy. Consequently, it is recommended that the logistic regression approach be adopted as the preferred classification method for a computer-assisted system designed for diagnosing vascular bleeding types.

Acknowledgment

This research was supported by Mashhad University of Medical Sciences, under grant No. 4010494.

References

1. Feigin VL. Stroke in developing countries: can the epidemic be stopped and outcomes improved? Lancet Neurol. 2007 Feb;6(2):94-7. doi: 10.1016/S1474-4422(07)70007-8.
2. Feigin VL, Lawes CM, Bennett DA, Barker-Collo SL, Parag V. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. Lancet Neurol. 2009 Apr;8(4):355-69. doi: 10.1016/S1474-4422(09)70025-0.
3. Garg R, Biller J. Recent advances in spontaneous intracerebral hemorrhage. F1000Res. 2019 Mar 18;8:F1000 Faculty Rev-302. doi: 10.12688/f1000research.16357.1.
4. Donnan GA, Fisher M, Macleod M, Davis SM. Stroke. Lancet. 2008 May 10;371(9624):1612-23. doi: 10.1016/S0140-6736(08)60694-7.
5. van Asch CJ, Luitse MJ, Rinkel GJ, van der Tweel I, Algra A, Klijn CJ. Incidence, case fatality, and functional outcome of intracerebral haemorrhage over time, according to age, sex, and ethnic origin: a systematic review and meta-analysis. Lancet Neurol. 2010 Feb;9(2):167-76. doi: 10.1016/S1474-4422(09)70340-0.
6. O'Donnell MJ, Xavier D, Liu L, Zhang H, Chin SL, Rao-Melacini P, et al. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. Lancet. 2010 Jul 10;376(9735):112-23. doi: 10.1016/S0140-6736(10)60834-3.
7. Viswanathan A, Greenberg SM. Cerebral amyloid angiopathy in the elderly. Ann Neurol. 2011 Dec;70(6):871-80. doi: 10.1002/ana.22516.
8. Andersen KK, Olsen TS, Dehlendorff C, Kammersgaard LP. Hemorrhagic and ischemic strokes compared: stroke severity, mortality, and risk factors. Stroke. 2009 Jun;40(6):2068-72. doi: 10.1161/STROKEAHA.108.540112.
9. Zhang Y, Tuomilehto J, Jousilahti P, Wang Y, Antikainen R, Hu G. Lifestyle factors on the risks of ischemic and hemorrhagic stroke. Arch Intern Med. 2011 Nov 14;171(20):1811-8. doi: 10.1001/archinternmed.2011.443.
10. Wieberdink RG, Poels MM, Vernooij MW, Koudstaal PJ, Hofman A, van der Lugt A, et al. Serum lipid levels and the risk of intracerebral hemorrhage: the Rotterdam Study. Arterioscler Thromb Vasc Biol. 2011 Dec;31(12):2982-9. doi: 10.1161/ATVBAHA.111.234948.
11. Emerging Risk Factors Collaboration; Sarwar N, Gao P, Seshasai SR, Gobin R, Kaptoge S, Di Angelantonio E, Ingelsson E, Lawlor DA, Selvin E, Stampfer M, Stehouwer CD, Lewington S, Pennells L, Thompson A, Sattar N, White IR, Ray KK, Danesh J. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet. 2010 Jun 26;375(9733):2215-22. doi: 10.1016/S0140-6736(10)60484-9. Erratum in: Lancet. 2010 Sep 18;376(9745):958. Hillage, H L [corrected to Hillege, H L]. PMID: 20609967; PMCID: PMC2904878.
12. Rost NS, Greenberg SM, Rosand J. The genetic architecture of intracerebral hemorrhage. Stroke. 2008 Jul;39(7):2166-73. doi: 10.1161/STROKEAHA.107.501650. Epub 2008 May 8. PMID: 18467649.
13. Vanent KN, Leasure AC, Acosta JN, Kuohn LR, Woo D, Murthy SB, Kamel H, Messé SR, Mullen MT, Cohen JB, Cohen DL, Townsend RR, Petersen NH, Sansing LH, Gill TM, Sheth KN, Falcone GJ. Association of Chronic Kidney Disease With Risk of Intracerebral Hemorrhage. JAMA Neurol. 2022 Sep 1;79(9):911-918. doi: 10.1001/jamaneurol.2022.2299. PMID: 35969388; PMCID: PMC9379821.
14. Lee M, Saver JL, Chang KH, Liao HW, Chang SC, Ovbiagele B. Impact of microalbuminuria on incident stroke: a meta-analysis. Stroke. 2010 Nov;41(11):2625-31. doi: 10.1161/STROKEAHA.110.581215. Epub 2010 Oct 7. PMID: 20930164.
15. Sheth KN. Spontaneous Intracerebral Hemorrhage. N Engl J Med. 2022 Oct 27;387(17):1589-1596. doi: 10.1056/NEJMra2201449. Erratum in: N Engl J Med. 2023 Apr 13;388(15):1440. PMID: 36300975.
16. Hemphill JC 3rd, Greenberg SM, Anderson CS, Becker K, Bendok BR, Cushman M, Fung GL, Goldstein JN, Macdonald RL, Mitchell PH, Scott PA, Selim MH, Woo D; American Heart Association Stroke Council; Council on Cardiovascular and Stroke Nursing; Council on Clinical Cardiology. Guidelines for the Management of Spontaneous Intracerebral Hemorrhage: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2015 Jul;46(7):2032-60. doi: 10.1161/STR.0000000000000069. Epub 2015 May 28. PMID: 26022637.
17. Macellari F, Paciaroni M, Agnelli G, Caso V. Neuroimaging in intracerebral hemorrhage. Stroke. 2014 Mar;45(3):903-8. doi: 10.1161/STROKEAHA.113.003701. Epub 2014 Jan 14. PMID: 24425128.
18. Morotti A, Goldstein JN. Diagnosis and Management of Acute Intracerebral Hemorrhage. Emerg Med Clin North Am. 2016 Nov;34(4):883-899. doi: 10.1016/j.emc.2016.06.010. Epub 2016 Sep 3. PMID: 27741993; PMCID: PMC5089075.
19. Strub WM, Leach JL, Tomsick T, Vagal A. Overnight preliminary head CT interpretations provided by residents: locations of misidentified

intracranial hemorrhage. AJNR Am J Neuroradiol. 2007 Oct;28(9):1679-82. doi: 10.3174/ajnr.A0653. Epub 2007 Sep 20. PMID: 17885236; PMCID: PMC8134182.
20. Chan KT, Carroll T, Linnau KF, Lehnert B. Expectations Among Academic Clinicians of Inpatient Imaging Turnaround Time: Does it Correlate with Satisfaction? Acad Radiol. 2015 Nov;22(11):1449-56. doi: 10.1016/j.acra.2015.06.019. Epub 2015 Aug 24. PMID: 26310727.
21. Ioffe S. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167. 2015.
22. Zhang J, Chen L. Breast Cancer Diagnosis from Perspective of Class Imbalance. Iranian Journal of Medical Physics, 2019; 16(3): 241-249. doi: 10.22038/ijmp.2018.31600.1373.
23. Jajroudi M, Enferadi M, Bagherpour Z, Reiazi R. Association Rule Mining-Based Radiomics in Breast Cancer Diagnosis. Iranian Journal of Medical Physics, 2024; 21(2): 113-120. doi: 10.22038/ijmp.2023.66728.2147.
24. Chockley K, Emanuel E. The End of Radiology? Three Threats to the Future Practice of Radiology. J Am Coll Radiol. 2016 Dec;13(12 Pt A):1415-1420. doi: 10.1016/j.jacr.2016.07.010.
25. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May 28;521(7553):436-44. doi: 10.1038/nature14539.
26. Steiner T, Rosand J, Diringer M. Intracerebral hemorrhage associated with oral anticoagulant therapy: current practices and unresolved questions. Stroke. 2006 Jan;37(1):256-62. doi: 10.1161/01.STR.0000196989.09900.f8.
27. Shinohara Y, Takahashi N, Lee Y, Ohmura T, Kinoshita T. Development of a deep learning model to identify hyperdense MCA sign in patients with acute ischemic stroke. Jpn J Radiol. 2020 Feb;38(2):112-117. doi: 10.1007/s11604-019-00894-4.
28. Sage A, Badura P. Intracranial hemorrhage detection in head CT using double-branch convolutional neural network, support vector machine, and random forest. Applied Sciences. 2020 Oct 27;10(21):7577.
29. Alfaer NM, Aljohani HM, Abdel-Khalek S, Alghamdi AS, Mansour RF. Fusion-Based Deep Learning with Nature-Inspired Algorithm for Intracerebral Haemorrhage Diagnosis. J Healthc Eng. 2022 Jan 18;2022:4409336. doi: 10.1155/2022/4409336.
30. Nafees Ahmed S, Prakasam P. A systematic review on intracranial aneurysm and hemorrhage detection using machine learning and deep learning techniques. Prog Biophys Mol Biol. 2023 Oct;183:1-16. doi: 10.1016/j.pbiomolbio.2023.07.001.
31. Maghami M, Sattari SA, Tahmasbi M, Panahi P, Mozafari J, Shirbandi K. Diagnostic test accuracy of machine learning algorithms for the detection intracranial hemorrhage: a systematic review and meta-analysis study. Biomed Eng Online. 2023 Dec 4;22(1):114. doi: 10.1186/s12938-023-01172-1.
32. Wang X, Shen T, Yang S, Lan J, Xu Y, Wang M, et al. A deep learning algorithm for automatic detection and classification of acute intracranial hemorrhages in head CT scans. Neuroimage Clin. 2021;32:102785. doi: 10.1016/j.nicl.2021.102785.
33. Altuve M, Pérez A. Intracerebral hemorrhage detection on computed tomography images using a residual neural network. Phys Med. 2022 Jul;99:113-119. doi: 10.1016/j.ejmp.2022.05.015.
34. Cearns M, Hahn T, Baune BT. Recommendations and future directions for supervised machine learning in psychiatry. Transl Psychiatry. 2019 Oct 22;9(1):271. doi: 10.1038/s41398-019-0607-2.
35. Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. J Bioinform Comput Biol. 2005 Apr;3(2):185-205. doi: 10.1142/s0219720005001004.
36. Radovic M, Ghalwash M, Filipovic N, Obradovic Z. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. BMC Bioinformatics. 2017 Jan 3;18(1):9. doi: 10.1186/s12859-016-1423-9.
37. Kohavi R, John GH. Wrappers for feature subset selection. Artificial intelligence. 1997 Dec 1;97(1-2):273-324.
38. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Machine learning. 2002 Jan;46:389-422.
39. Hatt M, Krizsan AK, Rahmim A, Bradshaw TJ, Costa PF, Forgacs A, et al. Joint EANM/SNMMI guideline on radiomics in nuclear medicine: Jointly supported by the EANM Physics Committee and the SNMMI Physics, Instrumentation and Data Sciences Council. European Journal of Nuclear Medicine and Molecular Imaging. 2023 Jan;50(2):352-75.
40. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research. 2002 Jun 1;16:321-57.
41. Bekkar M, Djemaa HK, Alitouche TA. Evaluation measures for models assessment over imbalanced data sets. J Inf Eng Appl. 2013 Apr;3(10).
42. Schmitt N, Mokli Y, Weyland CS, Gerry S, Herweh C, Ringleb PA, et al. Automated detection and segmentation of intracranial hemorrhage suspect hyperdensities in non-contrast-enhanced CT scans of acute stroke patients. European Radiology. 2022 Apr 1:1-9.
43. Hopkins BS, Murthy NK, Texakalidis P, Karras CL, Mansell M, Jahromi BS, Potts MB, Dahdaleh NS. Mass deployment of deep neural network: real-time proof of concept with screening of intracranial hemorrhage using an open data set. Neurosurgery. 2022 Apr 8;90(4):383-9.
44. Seyam M, Weikert T, Sauter A, Brehm A, Psychogios MN, Blackham KA. Utilization of artificial intelligence–based intracranial hemorrhage detection on emergent noncontrast CT images in clinical workflow. Radiology: Artificial Intelligence. 2022 Feb 9;4(2):e210168.
45. Altuve M, Pérez A. Intracerebral hemorrhage detection on computed tomography images using a residual neural network. Physica Medica. 2022 Jul 1;99:113-9.
46. Tang Z, Zhu Y, Lu X, Wu D, Fan X, Shen J, et al. Deep learning-based prediction of hematoma expansion using a single brain computed tomographic slice in patients with spontaneous intracerebral hemorrhages. World Neurosurgery. 2022 Sep 1;165:e128-36.

47. Cortés-Ferre L, Gutiérrez-Naranjo MA, Egea-Guerrero JJ, Pérez-Sánchez S, Balcerzyk M. Deep Learning Applied to Intracranial Hemorrhage Detection. Journal of Imaging. 2023 Feb 7;9(2):37.
48. Kau T, Ziurlys M, Taschwer M, Kloss-Brandstätter A, Grabner G, Deutschmann H. FDA-approved deep learning software application versus radiologists with different levels of expertise: detection of intracranial hemorrhage in a retrospective single-center study. Neuroradiology. 2022 May 1:1-0.
49. Tharek A, Muda AS, Hudi AB, Hudin AB. Intracranial hemorrhage detection in ct scan using deep learning. Asian Journal of Medical Technology. 2022 Jan 31;2(1):1-8.
50. Rao BN, Mohanty S, Sen K, Acharya UR, Cheong KH, Sabut S. Deep Transfer Learning for Automatic Prediction of Hemorrhagic Stroke on CT Images. Computational and Mathematical Methods in Medicine. 2022 Apr 16;2022.
51. Voter AF, Larson ME, Garrett JW, Yu JP. Diagnostic accuracy and failure mode analysis of a deep learning algorithm for the detection of cervical spine fractures. American Journal of Neuroradiology. 2021 Aug 1;42(8):1550-6.
52. Kuo W, Häne C, Mukherjee P, Malik J, Yuh EL. Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. Proceedings of the National Academy of Sciences. 2019 Nov 5;116(45):22737-45.
53. Abe D, Inaji M, Hase T, Takahashi S, Sakai R, Ayabe F, et al. A prehospital triage system to detect traumatic intracranial hemorrhage using machine learning algorithms. JAMA Network Open. 2022 Jun 1;5(6):e2216393-.
54. Trevisi G, Caccavella VM, Scerrati A, Signorelli F, Salamone GG, Orsini K, et al. Machine learning model prediction of 6-month functional outcome in elderly patients with intracerebral hemorrhage. Neurosurgical Review. 2022 Aug;45(4):2857-67.
55. Uchida K, Kouno J, Yoshimura S, Kinjo N, Sakakibara F, Araki H, et al. Development of machine learning models to predict probabilities and types of stroke at prehospital stage: The Japan urgent stroke triage score using machine learning (JUST-ML). Translational Stroke Research. 2022 Jun 1:1-2.