European Journal of Radiology Open 13 (2024) 100618

Contents lists available at ScienceDirect

European Journal of Radiology Open

journal homepage: www.elsevier.com/locate/ejro

Enhancing mortality prediction in patients with spontaneous intracerebral hemorrhage: Radiomics and supervised machine learning on non-contrast computed tomography

Antonio López-Rueda [a,b,*,1,2], María-Ángeles Rodríguez-Sánchez [c], Elena Serrano [d], Javier Moreno [b], Alejandro Rodríguez [e], Laura Llull [e], Sergi Amaro [e], Laura Oleaga [b]

[a] Clinical Informatics Department, Hospital Clínic de Barcelona, Barcelona, Spain
[b] Radiology Department, Hospital Clínic de Barcelona, Barcelona, Spain
[c] Valencian Research Institute for Artificial Intelligence (VRAIN), Universitat Politècnica de València, Valencia, Spain
[d] Radiology Department, Hospital Universitario de Bellvitge, Barcelona, Spain
[e] Neurology Department, Hospital Clínic de Barcelona, Barcelona, Spain

ARTICLE INFO

Keywords:
Radiomics
Machine learning
Intracerebral hemorrhage
Computed tomography
Stroke

ABSTRACT

Purpose: This study aims to develop a Radiomics-based Supervised Machine-Learning model to predict mortality in patients with spontaneous intracerebral hemorrhage (sICH).

Methods: Retrospective analysis of a prospectively collected clinical registry of patients with sICH consecutively admitted at a single academic comprehensive stroke center between January-2016 and April-2018. We conducted an in-depth analysis of 105 radiomic features extracted from 105 patients. Following the identification and handling of missing values, radiomics values were scaled to 0–1 to train different classifiers. The sample was split into 80–20 % training-test and validation cohort in a stratified fashion. Random Forest(RF), K-Nearest Neighbor(KNN), and Support Vector Machine(SVM) classifiers were evaluated, along with several feature selection methods and hyperparameter optimization strategies, to classify the binary outcome of mortality or survival during hospital admission. A tenfold stratified cross-validation method was used to train the models, and average metrics were calculated.

Results: RF, KNN, and SVM, with the "DropOut+SelectKBest" feature selection strategy and no hyperparameter optimization, demonstrated the best performances with the least number of radiomic features and the most simplified models, achieving a sensitivity range between 0.90 and 0.95 and AUC range from 0.97 to 1 on the validation dataset. Regarding the confusion matrix, the SVM model did not predict any false negative test (negative predicted value 1).

Conclusion: Radiomics-based Supervised Machine Learning models can predict mortality during admission in patients with sICH. SVM with the "DropOut+SelectKBest" feature selection strategy and no hyperparameter optimization was the best simplified model to detect mortality during admission in patients with sICH.

1. Introduction

Cerebrovascular disease is the second leading cause of death worldwide. After ischemic stroke, spontaneous intracerebral hemorrhage (sICH) is the second most common subtype accounting for 10–20 % of all cases [1,2].

Non-contrast head computed tomography (NCCT) is the first-line diagnostic test for the emergency evaluation of acute stroke [3]. Multiple radiological signs, representing irregularity and/or heterogeneity of the hematoma, had been defined on NCCT as predictors of expansion or clinical outcome of the patients with sICH [4–10].The limitations of these radiological signs include their low inter- and intra-observer

* Correspondence to: 170 Villarroel Street, Barcelona 08036, Spain.
E-mail address: alrueda81@hotmail.com (A. López-Rueda).
[1] Twitter: @AntonioLR81
[2] ORCID: 0000–0001-7914–9948

https://doi.org/10.1016/j.ejro.2024.100618
Received 17 October 2024; Received in revised form 17 November 2024; Accepted 28 November 2024

agreement with overlapping definitions between them [11]. On the other hand, low sensitivity rates have been described in the prognostic prediction of these signs (less than 40 % for the functional outcome) [12].

The acknowledged limitations of the previously described radiological signs highlight the need to develop reproducible quantitative parameters (biomarkers) able to minimize the subjective component of qualitative radiological assessment and improve performance in predicting the prognosis of patients with sICH.

Radiomics-based machine learning models are a promising solution to detect patients with sICH at risk of mortality during admission. This study aims to develop different models of machine learning algorithms based on radiomics features derived from the NCCT images of sICH sufferers and to compare their performance to find the best-recommended model to discriminate mortality in patients with sICH.

2. Materials and methods

The manuscript has been structured according to the Checklist for Artificial Intelligence in Medical Imaging (CLAIM initiative)[13], and the level of evidence of the study is Level 5B[14].

2.1. Study design

Retrospective analysis of a prospectively collected clinical registry of patients with sICH consecutively admitted at a single academic comprehensive stroke center between January 2016 and April 2018. The study aims to develop a screening model to detect patients with sICH with a higher mortality risk. The proposed role of the AI algorithm is to triage patients based on the mortality risk to decide which patients would benefit from a more aggressive treatment plan. The clinical outcome to predict is mortality.

The study protocol was approved by the local Clinical Research Ethics Committee (registration number HCB/2020/0180) under the requirements of Spanish legislation in the field of biomedical research, the protection of personal data (15/1999), and the standards of Good Clinical Practice, as well as with the 1964 Helsinki Declaration. The data were properly coded to ensure data protection and patient traceability if needed. Due to the retrospective nature of the study, informed consent was not required, but all patients signed informed consent to participate in the prospective registry. The data that supports the findings of this study are available from the corresponding author, upon reasonable request.

2.2. Data

The analyzed data was obtained from a prospectively collected clinical registry of patients with sICH consecutively admitted at a single academic comprehensive stroke center, adequate for the resolution of the clinical question. This dataset has been used previously to detect poor clinical outcomes (3–6 modified Rankin Scale) at hospital discharge with a sensitivity rate in the validation cohort of 0.897 (0778–1;95 %IC), mainly probably due to the imbalanced sample (85 % poor clinical outcomes), therefore due to a high pretest probability[15].

The main inclusion criteria for this retrospective analysis were consecutive patients older than 18 years old and NCCT acquisition within the first 24 hours after symptom onset (focal neurological deficits such as hemiparesis/hemiplegia, aphasia, facial paralysis, etc…). Exclusion criteria were ICH secondary to cranioencephalic traumatism, arteriovenous malformation, brain tumor, primary intraventricular hemorrhage, cerebral venous thrombosis, or hemorrhagic transformation of an acute ischemic stroke. We also excluded patients for whom clinical information or radiomic features were not available.

Demographics (age, sex), toxic habits (alcohol, smoking), cerebrovascular risk factors (hypertension, dyslipidemia, diabetes mellitus, atrial fibrillation, ischemic heart disease), medical history of previous stroke and concomitant therapies (antiplatelet or anticoagulant drug treatment) were prospectively collected.

On admission, glycemia (mmol/l) and initial neurological assessment were recorded using the "National Institutes of Health Stroke Scale" (NIHSS)[16]. Functional outcome was quantified using the Modified Rankin Scale (mRS) score at discharge, including mortality rate[17].

A sequential NCCT was performed on two Multislice CT scanners (Somatom Definition Flash and Somatom Sensation 64, Siemens Healthcare, Erlangen, Germany). Axial sequential images were performed parallel to the orbitomeatal line from the skull base to the vertex using standard parameters (140 kV, 230 mAs, and 5-mm slice thickness). No resampling or other image preprocessing was done.

The qualitative interpretation was made in a blinded fashion to clinical information by a neuroradiologist, with more than 10 years of experience. The qualitative assessment included an evaluation of hemorrhage location (basal ganglia, lobar, brainstem, and cerebellum), presence of intraventricular hemorrhage, and hematoma volume. The volume of the hematoma was calculated according to the validated AxBxC/2 method [18]. Finally, the interpretation of the Hematoma Maturity Score (HMS) [10] was made by consensus between two experienced radiologists (both with over 10 years of experience in acute brain NCCT in the emergency setting), without access to clinical information.

DICOM (Digital imaging and communications in medicine) files were transferred to an external computing station for processing. A single neuroradiologist (10 years' experience) segmented the hematoma using the "Segment Editor" module of 3DSlicer software version 4.10.2[19]. Contours of all hematomas were manually drawn slice by slice and three-dimensional volumes of interest (VOIs) were extracted from each sICH.

From the "Radiomics" module of the 3DSlicer Software, a total of 105 features of each of the VOIs were automatically obtained. Features were related to intensity (19 features), shape (10 features), and texture (76 features). 3DSlicer "Radiomics" module is based on the pyRadiomics library [20], which meets the "Image Biomarker Standardization Initiative" (IBSI) standard [21].

In this quantitative analysis, we used radiomic features to train the models and predict patient mortality with sICH. There were no missing values. No outlier elimination strategy was performed given the low number of instances to carry out the training. Instead, radiomics features were scaled from 0 to 1 as a normalization process to mitigate the possible effects of outliers or extreme values. Mortality during hospital admission was determined as the target variable, with a relatively balanced sample (mortality rate 33,3 %).

All data preprocessing, analysis, and visualization were performed on the Google Cloud computing service "Google Colab" (colab.research.google.com) using Python 3.0 programming language (Python Software Foundation; http://www.python.org). For the statistical analysis involved in machine learning and classification, Python packages Scikit-learn 1.0.2 (Scikit-learn: Machine Learning in Python; https://scikit-learn.org) and Seaborn 0.11.2 (Seaborn: statistical data visualization; https://seaborn.pydata.org) were used.

2.3. Ground truth

The target of the study was to evaluate whether supervised learning classifiers based on the radiomic features of the NCCT were able to predict mortality during hospital admission in patients with sICH. The main limitation regarding the target variable definition was that the sample may not fit the whole population of sICH patients, since those who died before reaching the comprehensive stroke center were not included in the dataset. This limitation or bias does not invalidate the study since the model will be used on a population similar to the one that has been trained.

The proposed role of the AI algorithm is to triage patients with a

mortality risk, which may benefit from more aggressive treatment regimens, such as intensive blood pressure reduction, early surgery approach in lobar sICH, or minimally invasive surgery in deep sICH.

2.4. Data partitions

The appropriateness of the sample size was tested by the reference value of Rajput et al., i.e., they regarded sample size of a dataset as adequate when the dataset meets both two criteria (i) prediction accuracy $>$ 80 % and (ii) Cohen's d $>$ 0.5 [22]. A stratified random patient-level split of the dataset was performed, allocating 80 % for cross-validation (training/test) and 20 % for a held-out validation set, maintaining the same mortality rate across all subsets.

2.5. Model

Radiomics features were scaled from 0 to 1 as a normalization process to ensure the correct performance of the classifiers. Different feature selection and hyperparameter optimization strategies were performed in the training/test dataset.

Considering the large number of radiomics features and the high correlation between them, we chose several feature selection methods, which are described below:

- "DropOut" Feature Selection: In this method, we removed features with Pearson`s correlation between them greater than 0.6 in absolute terms, so we reduced from 105 to 67 radiomics features for modeling.
- "DropOut + SelectKBest" Feature Selection: In this method, we first remove the same features as in the previous method, and then we select the 10 features that have the highest association with mortality in a univariate analysis using "*SelectKBest*" from the Sklearn library (ANOVA F-value).
- "DropOut + L1" Feature Selection: In this method, we first remove the same features as in the previous method, and then we perform a feature selection based on the Lasso regularization using "*SelectFromModel*" from the Sklearn library. This process does not perform any additional feature selection, so we do not train the classifier with this model.
- "DropOut + L2" Feature Selection: In this method, we first remove the same features as in the previous method, and then we perform a feature selection based on the Ridge regularization using "*SelectFromModel*" from the Sklearn library. From this process, 27 features were selected and used to train the classifiers.

Radiomic features were selected for modeling with supervised learning algorithms such as Random Forest (RF), K-Nearest Neighbors (KNN), and Support Vector Machines (SVM) because they show high performances in the medical literature[23–25].

2.6. Training and evaluation

Regarding the hyperparameters of these supervised learning algorithms, we trained the models with different hyperparameter optimization strategies as follows:

- With no hyperparameter optimization.
- Hyperparameter optimization by grid search using "*GridSearchCV*" from the Sklearn library.
- Hyperparameter optimization by randomized search using "*RandomizedSearchCV*" from the Sklearn library.

Supervised learning algorithms were trained with the "*cross_val_score*" function from the Sklearn library in the training/test dataset. From this process, the mean values over 10 training/test iterations were obtained for Accuracy, Recall/Sensitivity, Precision, and F1 score.

Once training was completed, predictions were made with trained algorithms in the validation dataset. Accuracy, Recall/Sensitivity, Precision, and F1 score were obtained, as well as ROC curve (Receiver Operating Characteristic curve) and AUC (Areas Under the Curve) values of the algorithms and confusion matrix with the best performances. As a screening method, we established the sensitivity, false negative, and negative predictive values (NPV) as the most important metrics to measure the model's performance, and therefore to select the best model in validation dataset.

3. Results

3.1. Data

One hundred and five patients met the inclusion and exclusion criteria and were included in the analysis, the patient's flowchart was shown in Fig. 1. The patient's demographic and clinical data are summarized in Table 1. Functional outcome and qualitative radiological features are summarized in Tables 1 and 2 respectively. The correlation matrix of all radiomic parameters is visualized in a heat map showing a high positive and negative correlation between multiple features (Fig. 2). The dataset was stratified (random split) in training/test and validation samples with an 8:2 ratio, so the training/test dataset was composed of 84 patients, and the validation dataset was composed of 21 patients.

3.2. Model performance

Mean values over 10 iterations in training/test dataset were obtained for Accuracy, Recall/Sensitivity, Precision, and F1 score are shown in Table 3. Next, we evaluated the supervised learning models already trained over the validation dataset, as shown in Table 4.

"DropOut" and "DropOut + SelectKBest" were the feature selections with the best performances on the validation dataset, with a sensitivity range between 0.90 and 0.95. No performance improvement with the hyperparameter optimization strategies compared to no optimization strategy. Considering that the "DropOut + SelectKBest" strategy selected 10 features and has similar results as the "DropOut" strategy with 67 features, we performed the ROC curve and the confusion matrix with the strategy with the fewest number of variables to simplify the model as much as possible. In the same direction, we performed the ROC curve and the confusion matrix with the no hyperparameter-optimized models to select the most simplified models.

Therefore, the ROC curve and confusion matrix were performed with the 3 trained models (RF, KNN, SVM) with the "DropOut + SelectKBest" feature selection strategy and with no hyperparameter optimization on the validation cohort. Fig. 3 shows areas under the curve ranging from 0.97 to 1 with the 3 trained models on the validation cohort.

Table 5 showed a confusion matrix with 2 false negative predictions in the RF and the KNN models (NPV 0.88) and no false negative prediction in the SVM model (NPV 1) on the validation cohort.

4. Discussion

In our study, we evaluated different radiomic-based supervised learning models to predict mortality during admission in patients with sICH. RF, KNN, and SVM, with "DropOut + SelectKBest" feature selection strategy without hyperparameter optimization demonstrated the best performances with the least number of radiomic features. These methods achieved a sensitivity range between 0.90 and 0.95 and an AUC range from 0.97 to 1 on the validation dataset. Regarding the confusion matrix, the SVM model did not predict any false negative test (NPV 1), compared to the RF and the KNN models which predicted 2 false negative tests (NPV 0.88).

The intended use of the classifiers is to serve as a screening tool, facilitating the identification of patients with sICH at high mortality risk

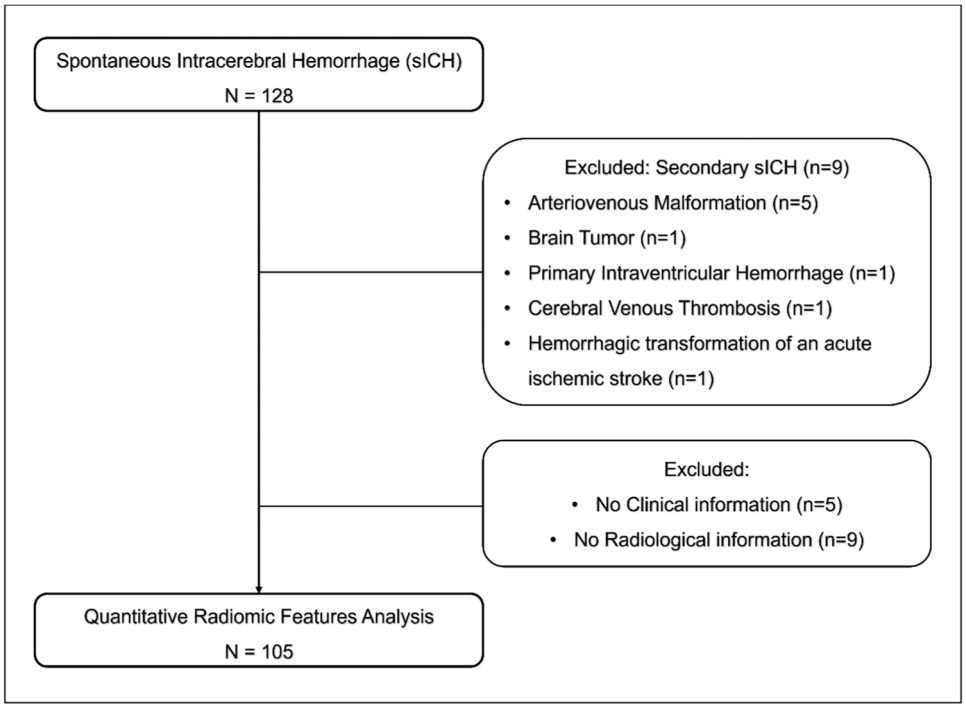

Fig. 1. Flowchart of the patient selection.

Table 1
Demographic and clinical data of the patients.

		Patients (n = 105)
Age, Mean (SD)		74 (13,3)
Sex (Male), n (%)		59 (56,2)
Alcohol, n (%)		7 (6,7)
Tobacco, n (%)		7 (6,7)
Hypertension, n (%)		66 (62,9)
Dyslipidemia, n (%)		41 (39)
Diabetes Mellitus, n (%)		23 (21,9)
Atrial Fibrillation, n (%)		22 (21)
Ischemic cardiopathy, n (%)		8 (7,6)
Previous stroke, n (%)		14 (13,3)
Anticoagulants, n (%)		25 (23,8)
Antiplatelets, n (%)		26 (24,8)
Antiplatelets and/or Anticoagulants, n (%)		50 (47,6)
Glycemia – mmol/l, Mean (SD)		155 (60,8)
NIHSS, Mean (SD)		14,7 (10,3)
mRS at discharge, n (%)	0	4 (3,8)
	1	5 (4,8)
	2	7 (6,7)
	3	18 (17,1)
	4	20 (19)
	5	16 (15,2)
	6	35 (33,3)
Mortality during hospital admission, n (%)		35 (33,3)

SD, Standard Deviation; NIHSS, National Institutes of Health Stroke Scale, mRS, Modified Ranking Scale

Table 2
Qualitative radiological features of the sICH of the patients.

		Patients (n = 105)
Location of sICH	Lobar, n (%)	47 (44,8)
	Deep, n (%)	47 (44,8)
	Cerebellum, n (%)	8 (7,6)
	Brainstem, n (%)	3 (2,9)
Ventricular extension, n (%)		47 (44,8)
Hematoma maturity score, n (%)		87 (82,9)
Hematoma mean volume - ml (SD)		37,9 (49,64)

sICH, Spontaneous intracerebral hemorrhage; SD, Standard Deviation

during hospital admission to decide on more aggressive treatment planning. For that purpose, the proposed models should have a high sensitivity rate with as few false negatives as possible. In this context, the goal of this study was to create a model based on radiomic features that eventually could serve to triage patients with sICH at high mortality risk during hospital admission. This goal has been achieved with the results of this study in this cohort, but these results may not be generalized due to the absent of external validation, which is the main limitation of this study. So, the generalization of the classifiers is uncertain and therefore further evaluation of the topic in prospective multicenter cohorts is warranted

To our knowledge, this is the first study of radiomic-based supervised learning algorithms, in which the main objective is to predict mortality during admission in patients with sICH. There are 2 similar studies, in which they build CT radiomic-based supervised learning algorithms to predict poor clinical outcomes (3–6 modified Rankin Scale) at discharge [15] and at 6 months[26].

The previous study of our group by Serrano et al.[15], that predicts poor clinical outcomes at discharge, was training with the same dataset as this study. The sensitivity rate in the validation cohort was 0.897 (0778–1;95 %IC), but these results were mainly due to the imbalanced sample (85 % poor clinical outcomes), hence a high pretest probability, so maybe a CT radiomic-based supervised learning algorithms is not useful to predict poor clinical outcome at discharge.

On the other hand, Xu et al. [26] develop a study with CT radiomic-based supervised learning algorithms to predict poor clinical outcomes at 6 months, with a balanced sample (67.8 % poor clinical outcomes) and excellent results in sensitivity (>90 %) and AUC (0.92) in validation cohort. So, those results, in combination with our results, show that CT radiomic-based supervised learning algorithms can predict mortality at discharge and poor clinical outcomes at 6 months in patients with sICH.

There are 2 other studies, similar to these previous ones, in which they build models to detect the expansion of sICH with radiomic biomarkers, which is one of the reasons why patients have poor clinical outcomes. In the first previous study, Shen and Cols. achieved an AUC of 0.92 in the training dataset and, sensitivity, specificity, and precision in

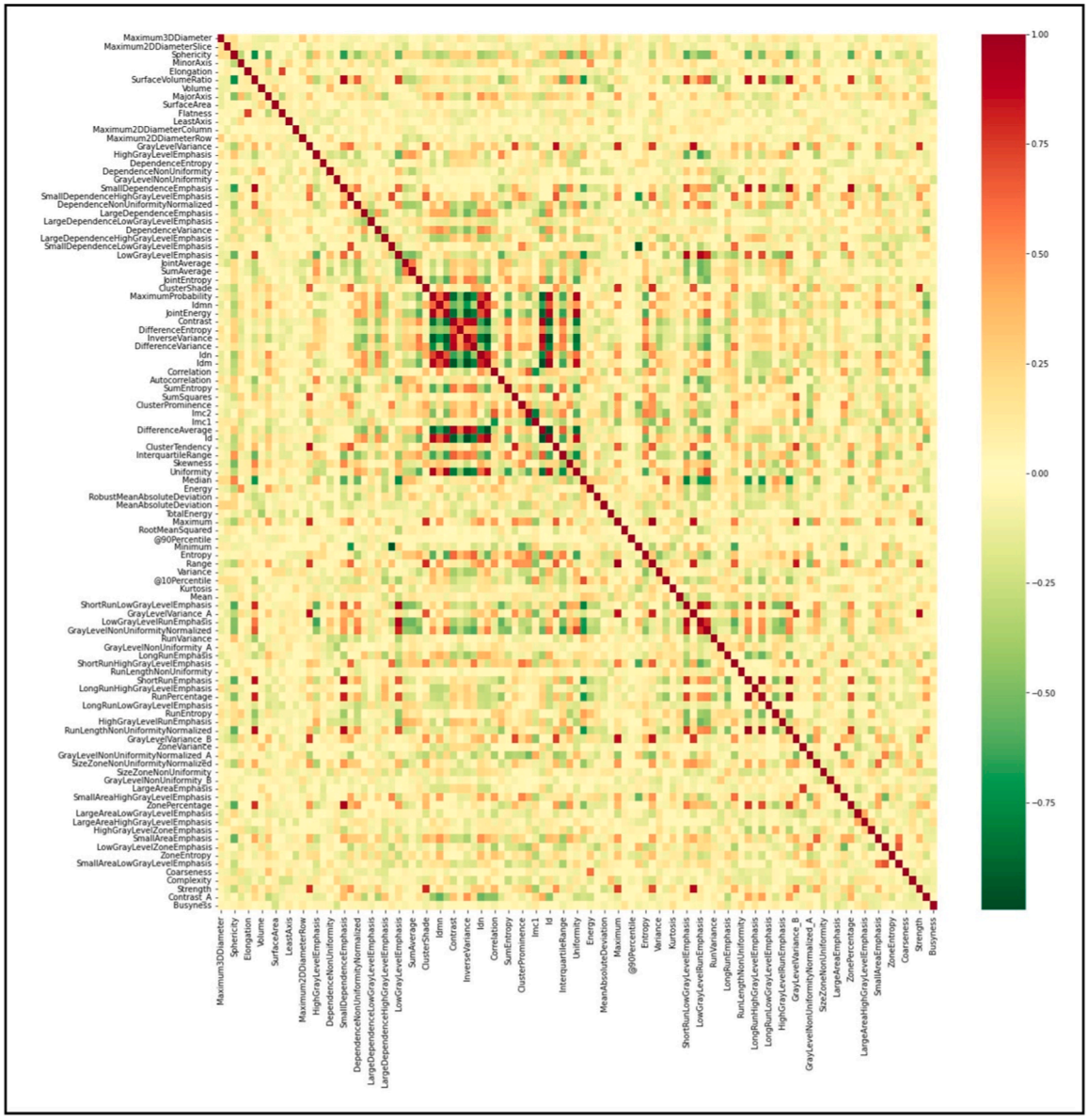

Fig. 2. Heatmap correlation matrix of the sICHs radiomics features.

the validation dataset of 0.86, 0.85, and 0.85 respectively. [27] On the other hand, Song and Cols. achieved an AUC between 0.728 and 0.829 for the detection of sICH expansion in the validation cohort. In this latter study, model performance did not improve substantially adding clinical-radiological variables to the classifiers.[23] These two previously published papers aimed to detect sICH expansion, but sICH patients without expansion are also at risk of mortality during admission.

The clinical outcome in sICH patients is not only due to the expansion of the hematoma but may also be due to edema produced in the surrounding brain parenchyma. For this reason, we believe that clinical outcome should be the prognostic imaging goal and not just hematoma expansion. Radiomic features related to ICH irregularity and heterogeneity may be important for predicting hematoma expansion and edema in the surrounding brain parenchyma. We hypothesize that ICH irregularity causes greater insult to the adjacent parenchyma, which could lead to an increased risk of edema in the adjacent parenchyma.

The main strength of this study was the use of a well characterized prospective cohort of consecutive sICH subjects admitted in a Comprehensive Stroke Center. Nevertheless, this study has several limitations. A limitation of the study is the small sample, especially for the validation set. However, the main limitation is the lack of external cohort validation, so the generalization of the classifiers is uncertain and therefore further evaluation of the topic in prospective multicenter cohorts is warranted. If the performance of the classifiers in the external validation cohort is lower than that obtained in the internal validation cohort, clinical-demographic data or additional radiological signs could be added to the models.

5. Conclusions

Radiomics-based Supervised Machine Learning models can predict mortality during admission in patients with spontaneous intracerebral hemorrhage in our cohort. Generalization of the classifiers is uncertain and therefore further evaluation of the topic in prospective multicenter cohorts is warranted.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Table 3
Mean values over 10 iterations as results of the algorithm's training on the training/test dataset.

			Accuracy	Recall/Sensitivity	Precision	F1 Score
DropOut	RF	No Optimization	0.753	0.433	0.717	0.387
		GridSearchCV	0.764	0.4	0.633	0.497
		RandomizedSearchCV	0.751	0.417	0.483	0.487
	KNN	No Optimization	0.728	0.5	0.708	0.566
		GridSearchCV	0.728	0.5	0.708	0.566
		RandomizedSearchCV	0.728	0.5	0.708	0.566
	SVM	No Optimization	0.726	0.283	0.65	0.38
		GridSearchCV	0.726	0.283	0.65	0.38
		RandomizedSearchCV	0.726	0.283	0.65	0.38
DropOut + SelectKBest	RF	No Optimization	0.753	0.567	0.757	0.567
		GridSearchCV	0.741	0.499	0.757	0.603
		RandomizedSearchCV	0.764	0.533	0.793	0.628
	KNN	No Optimization	0.751	0.499	0.733	0.567
		GridSearchCV	0.751	0.499	0.733	0.567
		RandomizedSearchCV	0.751	0.499	0.733	0.567
	SVM	No Optimization	0.715	0.5	0.576	0.492
		GridSearchCV	0.715	0.5	0.576	0.492
		RandomizedSearchCV	0.715	0.5	0.576	0.492
DropOut + L2	RF	No Optimization	0.764	0.4	0.683	0.420
		GridSearchCV	0.729	0.25	0.567	0.397
		RandomizedSearchCV	0.717	0.4	0.817	0.437
	KNN	No Optimization	0.772	0.467	0.833	0.587
		GridSearchCV	0.772	0.467	0.833	0.587
		RandomizedSearchCV	0.772	0.467	0.833	0.587
	SVM	No Optimization	0.761	0.533	0.8	0.598
		GridSearchCV	0.761	0.533	0.8	0.598
		RandomizedSearchCV	0.761	0.533	0.8	0.598

RF, Random Forest; KNN, K-Nearest Neighbors; SVM, Support Vector Machines

Table 4
Evaluation results of the trained models on the validation dataset.

			Accuracy	Recall/Sensitivity	Precision	F1 Score
DropOut	RF	No Optimization	0.95	0.95	0.96	0.95
		GridSearchCV	0.95	0.95	0.96	0.95
		RandomizedSearchCV	0.95	0.95	0.96	0.95
	KNN	No Optimization	0.90	0.90	0.92	0.90
		GridSearchCV	0.90	0.90	0.92	0.90
		RandomizedSearchCV	0.90	0.90	0.92	0.90
	SVM	No Optimization	0.90	0.90	0.92	0.90
		GridSearchCV	0.90	0.90	0.92	0.90
		RandomizedSearchCV	0.90	0.90	0.92	0.90
DropOut + SelectKBest	RF	No Optimization	0.90	0.90	0.92	0.90
		GridSearchCV	0.90	0.90	0.92	0.90
		RandomizedSearchCV	0.90	0.90	0.92	0.90
	KNN	No Optimization	0.90	0.90	0.92	0.90
		GridSearchCV	0.90	0.90	0.92	0.90
		RandomizedSearchCV	0.90	0.90	0.92	0.90
	SVM	No Optimization	0.95	0.95	0.96	0.95
		GridSearchCV	0.95	0.95	0.96	0.95
		RandomizedSearchCV	0.95	0.95	0.96	0.95
DropOut + L2	RF	No Optimization	0.81	0.81	0.81	0.80
		GridSearchCV	0.81	0.81	0.81	0.80
		RandomizedSearchCV	0.81	0.81	0.81	0.80
	KNN	No Optimization	0.81	0.81	0.85	0.78
		GridSearchCV	0.81	0.81	0.85	0.78
		RandomizedSearchCV	0.81	0.81	0.85	0.78
	SVM	No Optimization	0.86	0.86	0.86	0.85
		GridSearchCV	0.86	0.86	0.86	0.85
		RandomizedSearchCV	0.86	0.86	0.86	0.85

RF, Random Forest; KNN, K-Nearest Neighbors; SVM, Support Vector Machines

Ethical statement

The work described has not been published previously.
The article is not under consideration for publication elsewhere.
The article's publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out.

CRediT authorship contribution statement

Antonio López-Rueda: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation,

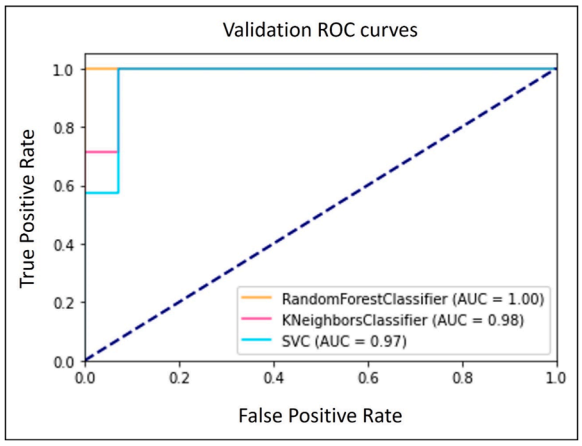

Fig. 3. ROC curve and AUC values with the "DropOut + SelectKBest" feature selection strategy and with no hyperparameter optimization models on the validation cohort.

Table 5
Confusion matrix of models with "DropOut + SelectKBest" feature selection strategy and with no hyperparameter optimization for mortality prediction on the validation dataset.

		RF Predicted		KNN Predicted		SVM Predicted	
		0	1	0	1	0	1
Ground Truth	0	14	0	14	0	13	1
	1	**2**	5	**2**	5	**0**	7

RF, Random Forest; KNN, K-Nearest Neighbors; SVM, Support Vector Machines; False negative values are in **bold**.

Writing – original draft, Writing – review & editing, Visualization, Project administration. **María-Ángeles Rodríguez-Sánchez:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Elena Serrano:** Conceptualization, Methodology, Software, Validation, Investigation, Resources, Writing – original draft, Writing – review & editing, Visualization. **Javier Moreno:** Conceptualization, Methodology, Software, Validation, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Alejandro Rodríguez:** Conceptualization, Methodology, Validation, Investigation, Resources, Writing – review & editing, Visualization. **Laura Llull:** Conceptualization, Methodology, Validation, Investigation, Resources, Writing – review & editing, Visualization. **Sergi Amaro:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Writing – review & editing, Visualization. **Laura Oleaga:** Conceptualization, Methodology, Validation, Investigation, Resources, Writing – review & editing, Visualization, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M.A. Ikram, R.G. Wieberdink, P.J. Koudstaal, International epidemiology of intracerebral hemorrhage, Curr. Atheroscler. Rep. 14 (4) (2012) 300–306.
[2] G.S. Collaborators, Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019, Lancet Neurol. 20 (10) (2021) 795–820.
[3] J.C. Hemphill, S.M. Greenberg, C.S. Anderson, K. Becker, B.R. Bendok, M. Cushman, et al., Guidelines for the Management of Spontaneous Intracerebral Hemorrhage: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association, Stroke 46 (7) (2015) 2032–2060.
[4] Q. Li, Q.J. Liu, W.S. Yang, X.C. Wang, L.B. Zhao, X. Xiong, et al., Island Sign: An Imaging Predictor for Early Hematoma Expansion and Poor Outcome in Patients With Intracerebral Hemorrhage, Stroke 48 (11) (2017) 3019–3025.
[5] E. Selariu, E. Zia, M. Brizzi, K. Abul-Kasim, Swirl sign in intracerebral haemorrhage: definition, prevalence, reliability and prognostic value, BMC Neurol. 12 (2012) 109.
[6] Q. Li, G. Zhang, X. Xiong, X.C. Wang, W.S. Yang, K.W. Li, et al., Black Hole Sign: Novel Imaging Marker That Predicts Hematoma Growth in Patients With Intracerebral Hemorrhage, Stroke 47 (7) (2016) 1777–1781.
[7] Q. Li, W.S. Yang, X.C. Wang, D. Cao, D. Zhu, F.J. Lv, et al., Blend sign predicts poor outcome in patients with intracerebral hemorrhage, PLoS One 12 (8) (2017) e0183082.
[8] G. Boulouis, A. Morotti, H.B. Brouwers, A. Charidimou, M.J. Jessel, E. Auriel, et al., Noncontrast Computed Tomography Hypodensities Predict Poor Outcome in Intracerebral Hemorrhage Patients, Stroke 47 (10) (2016) 2511–2516.
[9] G. Boulouis, A. Morotti, A. Charidimou, D. Dowlatshahi, J.N. Goldstein, Noncontrast Computed Tomography Markers of Intracerebral Hemorrhage Expansion, Stroke 48 (4) (2017) 1120–1125.
[10] E. Serrano, A. López-Rueda, J. Moreno, A. Rodríguez, L. Llull, C. Zwanzger, et al., The new Hematoma Maturity Score is highly associated with poor clinical outcome in spontaneous intracerebral hemorrhage, Eur. Radio. 32 (1) (2022) 290–299.
[11] A. Morotti, G. Boulouis, D. Dowlatshahi, Q. Li, C.D. Barras, C. Delcourt, et al., Standards for Detecting, Interpreting, and Reporting Noncontrast Computed Tomographic Markers of Intracerebral Hemorrhage Expansion, Ann. Neurol. 86 (4) (2019) 480–492.
[12] Z.K. Law, A. Ali, K. Krishnan, A. Bischoff, J.P. Appleton, P. Scutt, et al., Noncontrast Computed Tomography Signs as Predictors of Hematoma Expansion, Clinical Outcome, and Response to Tranexamic Acid in Acute Intracerebral Hemorrhage, Stroke 51 (1) (2020) 121–128.
[13] J. Mongan, L. Moy, C.E. Kahn, Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers, Radio. Artif. Intell. 2 (2) (2020) e200029.
[14] N. Pham, V. Hill, A. Rauschecker, Y. Lui, S. Niogi, C.G. Fillipi, et al., Critical Appraisal of Artificial Intelligence-Enabled Imaging Tools Using the Levels of Evidence System, AJNR Am. J. Neuroradiol. 44 (5) (2023) E21–E28.
[15] E. Serrano, J. Moreno, L. Llull, A. Rodríguez, C. Zwanzger, S. Amaro, et al., Radiomic-based nonlinear supervised learning classifiers on non-contrast CT to predict functional prognosis in patients with spontaneous intracerebral hematoma, Radiol. (Engl. Ed.) 65 (6) (2023) 519–530.
[16] T. Brott, H.P. Adams, C.P. Olinger, J.R. Marler, W.G. Barsan, J. Biller, et al., Measurements of acute cerebral infarction: a clinical examination scale, Stroke 20 (7) (1989) 864–870.
[17] J.L. Saver, N. Chaisinanunkul, B.C.V. Campbell, J.C. Grotta, M.D. Hill, P. Khatri, et al., Standardized Nomenclature for Modified Rankin Scale Global Disability Outcomes: Consensus Recommendations From Stroke Therapy Academic Industry Roundtable XI, Stroke 52 (9) (2021) 3054–3062.
[18] R.U. Kothari, T. Brott, J.P. Broderick, W.G. Barsan, L.R. Sauerbeck, M. Zuccarello, et al., The ABCs of measuring intracerebral hemorrhage volumes, Stroke 27 (8) (1996) 1304–1305.
[19] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.C. Fillion-Robin, S. Pujol, et al., 3D Slicer as an image computing platform for the Quantitative Imaging Network, Magn. Reson Imaging 30 (9) (2012) 1323–1341.
[20] J.J.M. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, et al., Computational Radiomics System to Decode the Radiographic Phenotype, Cancer Res 77 (21) (2017) e104–e107.
[21] A. Zwanenburg, M. Vallières, M.A. Abdalah, H.J.W.L. Aerts, V. Andrearczyk, A. Apte, et al., The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping, Radiology 295 (2) (2020) 328–338.
[22] D. Rajput, W.J. Wang, C.C. Chen, Evaluation of a decided sample size in machine learning applications, BMC Bioinforma. 24 (1) (2023) 48.
[23] Z. Song, D. Guo, Z. Tang, H. Liu, X. Li, S. Luo, et al., Noncontrast Computed Tomography-Based Radiomics Analysis in Discriminating Early Hematoma Expansion after Spontaneous Intracerebral Hemorrhage, Korean J. Radio. 22 (3) (2021) 415–424.
[24] M.E. Mayerhoefer, A. Materka, G. Langs, I. Häggström, P. Szczypiński, P. Gibbs, et al., Introduction to Radiomics, J. Nucl. Med 61 (4) (2020) 488–495.
[25] F. Al-Mufti, A.M. Thabet, T. Singh, M. El-Ghanem, K. Amuluru, C.D. Gandhi, Clinical and Radiographic Predictors of Intracerebral Hemorrhage Outcome, Inter. Neurol. 7 (1-2) (2018) 118–136.
[26] X. Xu, J. Zhang, K. Yang, Q. Wang, X. Chen, B. Xu, Prognostic prediction of hypertensive intracerebral hemorrhage using CT radiomics and machine learning, Brain Behav. 11 (5) (2021) e02085.
[27] Q. Shen, Y. Shan, Z. Hu, W. Chen, B. Yang, J. Han, et al., Quantitative parameters of CT texture analysis as potential markersfor early prediction of spontaneous intracranial hemorrhage enlargement, Eur. Radio. 28 (10) (2018) 4389–4396.