Diagn Interv Radiol 2024; DOI: 10.4274/dir.2024.222088

ARTIFICIAL INTELLIGENCE AND INFORMATICS

ORIGINAL ARTICLE

Copyright@Author(s) - Available online at dirjournal.org.
Content of this journal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Application of deep learning and radiomics in the prediction of hematoma expansion in intracerebral hemorrhage: a fully automated hybrid approach

Mengtian Lu[1]
Yaqi Wang[2]
Jiaqiang Tian[1]
Haifeng Feng[3]

[1]The First Affiliated Hospital of Hubei University of Science and Technology, Xianning Central Hospital, Department of Radiology, Xianning, China

[2]The Second Affiliated Hospital of Hubei University of Science and Technology, Department of Radiology, Xianning, China

[3]The First Affiliated Hospital of Hubei University of Science and Technology, Xianning Central Hospital, Department of Ultrasound, Xianning, China

PURPOSE

Spontaneous intracerebral hemorrhage (ICH) is the most severe form of stroke. The timely assessment of early hematoma enlargement and its proper treatment are of great significance in curbing the deterioration and improving the prognosis of patients with ICH. This study aimed to develop an automated hybrid approach to predict hematoma expansion in ICH.

METHODS

The transfer learning method was applied to build a hybrid model based on a convolutional neural network (CNN) to predict the expansion of hematoma. The model integrated (1) a CNN for automated hematoma segmentation and (2) a CNN-based classifier for hematoma expansion prediction that incorporated both 2-dimensional images and the radiomics features of the 3-dimensional hematoma shape.

RESULTS

The radiomics feature module had the highest area under the receiver operating characteristic curve (AUC) of 0.58, a precision of 0, a recall of 0, and an average precision (AP) of 0.26. The ResNet50 and Inception_v3 modules had AUCs of 0.79 and 0.93, a precision of 0.56 and 0.86, a recall of 0.42 and 0.75, and an AP of 0.51 and 0.85, respectively. Radiomic with Inception_v3 and Radiomic with ResNet50 had AUCs of 0.95 and 0.81, a precision of 0.90 and 0.57, a recall of 0.79 and 0.17, and an AP of 0.87 and 0.69, respectively.

CONCLUSION

A model using deep learning and radiomics was successfully developed. This model can reliably predict the hematoma expansion of ICH with a fully automated process based on non-contrast computed tomography imaging. Furthermore, the radiomics fusion with the Inception_v3 model had the highest accuracy.

KEYWORDS

Convolutional neural network, radiomics, intracerebral hemorrhage, hematoma expansion, non-contrast computed tomography

Corresponding author: Haifeng Feng

E-mail: haifengfeng_xn@yeah.net

Received 15 June 2023; revision requested 10 July 2023; last revision received 23 November 2024; accepted 10 February 2024.

Epub: 24.04.2024

Publication date: 09.09.2024

DOI: 10.4274/dir.2024.222088

Spontaneous intracerebral hemorrhage (ICH), the most severe form of stroke, accounts for 10%–15% of strokes in high-income nations and 20%–50% in developing nations.[1] Compared with ischemic stroke, ICH is associated with poor prognostic outcomes, with a mortality rate of 40% at 1 month and a disability rate of 80% in survivors.[2-4] Generally, ICH prognosis is influenced by many factors, including baseline volume and hematoma location, Glasgow Coma scale score, intraventricular hemorrhage, and age.[5-7] Of these factors, hematoma volume is the only one that is controllable and dynamic.[5] Clinical studies have shown that 33% of patients with ICH develop early hematoma enlargement within 24 h of ICH onset.[8-10] Early hematoma enlargement in patients with ICH is independently correlated with poor prognostic outcomes and death.[3,8,11,12] Therefore, the timely assessment of early hemato-

You may cite this article as: Lu M, Wang Y, Tian J, Feng H. Application of deep learning and radiomics in the prediction of hematoma expansion in intracerebral hemorrhage: a fully automated hybrid approach. *Diagn Interv Radiol.* 2024;30(5):299-312.

ma enlargement and appropriate treatment are crucial in improving the outcomes of patients with ICH.

Recent studies have identified numerous radiological features for predicting hematoma enlargement after ICH. For example, the computed tomography angiography (CTA) spot sign is a powerful predictive method of hematoma enlargement.[13-15] However, clinical applications have shown that some patients are allergic to the contrast agents used in CTA, and patients with renal insufficiencies are relatively contraindicated.[16-18] By contrast, non-contrast computed tomography (NCCT) serves as the first choice for diagnosis of patients with acute stroke. New markers based on NCCT, such as blend,[19] swirl,[20] black hole,[21] and island,[22] as well as satellite signs,[23] can reflect the density and shape heterogeneity of hematomas with high specificity. Therefore, they have been proposed as alternative predictors of hematoma enlargement in clinical settings.[24-26] However, these predictive indicators are only used for qualitative or semi-quantitative analyses and have relatively low sensitivity and accuracy in predicting hematoma enlargement.[25] Alternative quantitative methods with automated execution, which may have superior predictive power, are still needed.

Radiomics is a promising quantitative method that has performed excellently in various biomedical fields, where it has been used to extract large numbers of quantitative characteristics from conventional medical imaging.[27,28] As well as being extensively applied in oncology studies, radiomics has recently been used for the prediction of hematoma enlargement after ICH based on NCCT imaging. Although these models exhibit positive predictive performance,[29-33] only traditional or feature-based machine learning (ML) was utilized for the studies. Deep learning is an automatic method that can skip object segmentation, feature selection, and extraction from segmented objects to identify "effective features." Since it allows the whole process to be mapped from raw input images to final classifications and can exclude the requirement for hand-crafted features, deep learning is also referred to as end-to-end ML.[27]

Main points

- A model using deep learning and radiomics was successfully developed.
- The model was based on non-contrast computed tomography imaging.
- The model predicts hematoma expansion of an intracerebral hemorrhage with a fully automated process.

Convolutional neural networks (CNNs) based on deep learning are increasingly being used worldwide with promising outcomes. However, developing CNN-based methods requires large training datasets, which is challenging and laborious in clinical settings. Two approaches partially overcome this challenge. The first involves data augmentation, which utilizes affine transformations, including translation, scaling, and rotation, to produce more data from the available data. The other approach involves transfer learning, which is promising in medical image analysis.[34] In this study, we developed a CNN-based hybrid model using a data augmentation and transfer learning method for predicting hematoma expansion by integrating the following: (1) a CNN for automated hematoma segmentation, and (2) a CNN-centered classifier for hematoma expansion prediction that incorporates 2-dimensional (2D) images as well as radiomics features for the 3-dimensional (3D) hematoma shape.

Methods

This was a retrospective study permitted by the Medical Ethics Committee of Xianning Central Hospital (no: 20211126011), and the informed consent was waived.

Patients and image acquisition

Patients with spontaneous ICH diagnosed by the radiologist and admitted to the hospital between September 2017 and September 2021 were enrolled in this study. The inclusion criteria were as follows: (1) patients aged >18 years; (2) patients who had received baseline NCCT within 24 h of symptomatic onset and follow-up computed tomography (CT) at ≤72 h. In this study, 506 patients met the inclusion criteria. The exclusion criteria were as follows: (1) patients diagnosed with trauma, aneurysm, vascular malformation, venous sinus embolism, or tumor-induced cerebral hemorrhage (151 cases); (2) patients with emergency surgery before CT review (53 cases); (3) patients with surgical intervention during the 72-h observation period, even in the absence of hematoma expansion (15 cases); (4) patients whose CT images revealed artifacts. In this study, two different CT scanners were used to minimize variability in image acquisition parameters (7 cases). Following the exclusion criteria, 226 cases were excluded from the study. Finally, 280 cases were included in this study. The original CT image of the patients with ICH was reconstructed into a standardized image with a thickness of 1.0 mm and a spacing of 0.7 mm before being exported from the picture archiving and communication system (PACS).

Study design

The model architecture is shown in Figure 1. Briefly, the hybrid model comprised two deep learning models (models 1 and 2), with an automated pipeline between the two models. Model 1 was developed first for the automated generation of hematoma masks based on NCCT images. A hybrid classifier model (model 2) was then trained for the prediction of hemorrhage expansion based on radiomics and CNN features extracted from the NCCT images and hematoma masks in model 1.

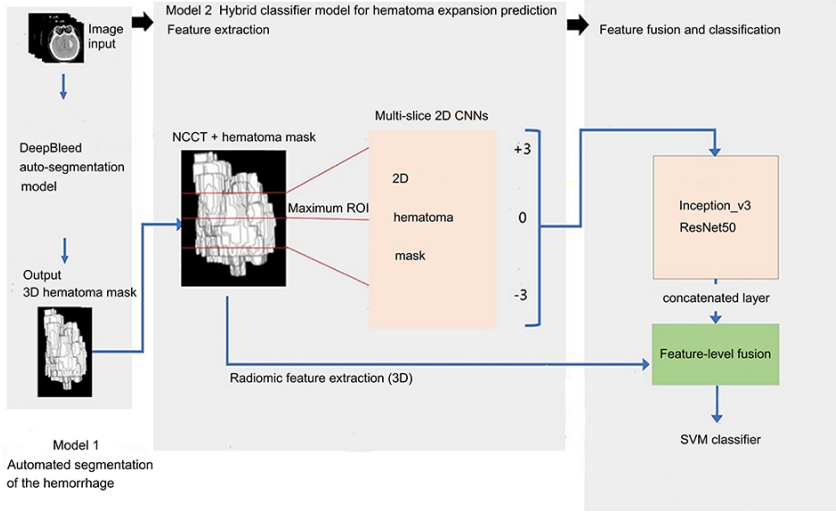

Figure 1. Fully automated hybrid model for predicting ICH expansion. ICH, intracerebral hemorrhage; NCCT, non-contrast computed tomography; ROI, region of interest; CNN, convolutional neural network; SVM, support vector machine.

Image processing

The Siemens 64-slice CT (SOMATOM go.Top, Henkestr., Erlangen, Germany) scan parameters were as follows: positioning image, 140 kV, 60 mA; spiral scanning parameters, 120 kV, effective mAs 230, pitch 0.55, layer thickness 5 mm, layer spacing 5 mm. Before exporting the image from the PACS (USA), the original image was reconstructed into a standardized image with a thickness of 1.0 mm and a spacing of 0.7 mm. The CT images were exported from PACS in the Digital Imaging and Communications in Medicine format and were transformed into the Neuroimaging Informatics Technology Initiative format. Images were set to values of -1,024 and 3,052 HU. Image analysis was performed using the Ubuntu 18.04 operating system (London, UK) and Python (USA). First, all input sequences were acquired in a 512 × 512 field. Next, the network output was set to only whole hematoma segmentation. After determining the hematoma volume and matching the original residual network, the original NCCT image and segmentation file (128 × 128 voxel) were reconstructed.

Automated segmentation of the hemorrhage (model 1)

DeepBleed is an open-source tool for quick hemorrhage segmentation.[35] This automated deep neural network model preprocesses NCCT scans (including skull stripping), segments the hemorrhage area, and outputs a binary hemorrhage mask. Training and validation of this model using data from the Minimally Invasive Surgery Plus Alteplase for Intracerebral Hemorrhage Evacuation trial showed that the dice similarity coefficient, which evaluates spatial overlap extent between ground-truth segmentation by radiologists and the automatic model, was 0.919.[35,36] Contrary to manual segmentation, this model can rapidly and precisely segment ICH with a high level of agreement. To adapt it to this study, the code was partially modified and an interface was added for the automated calculation of hematoma volume.

Labels

Previous studies have defined hematoma expansion as an outright volumetric ICH growth of >6 mL or an increase of >33% from the initial CT scan to the follow-up CT scan.[3,12] The binary label (negative or positive ICH expansion) was masked for each included study.

Hybrid classifier model for hematoma expansion prediction (model 2)

The hybrid model for ICH hematoma expansion combines well-known radiomics characteristic analysis and CNNs. Radiomics characteristic analysis was used to determine a wide range of researcher-defined quantitative features including shape, intensity, and texture of regions of interest on images. The CNN features were extracted based on CNN training.

The radiomics characteristic was extracted using the Python "pyradiomics" package (https://www.radiomics.io/pyradiomics.html).[37] The transformations were included with gradient, wavelet, original, square, exponential, logarithm, square root, and local binary pattern 3D. After image transformation, radiomics features were extracted.

The CNN model was executed using the well-known Inception_v3[38] and ResNet50 modules to extract comprehensive 2D image information from the hematoma. Axial slices with maximum hematoma areas were automatically selected as "maximum ICH image", and the other two images were extracted from three upper (+3) and three lower (−3) slices from the maximum ICH image. For small ICHs with a volume <1 cm, the upper and lower slices may not be within the hematoma range; the extracted features after separation were therefore the same and did not affect the data results. Hematomas <1 cm are often negative. Typically, CNN features are extracted from the output of fully connected last hidden layers.[39] During CNN training, "warm-up training" of the convolutional residual network was achieved using image inputs to identify relevant imaging features. The weights of the pretrained conventional ResNet50 and Inception_v3 modules were imported into model 2 and set as "non-trainable". Meanwhile, data augmentation was performed by rotating, flipping, and resizing the training images to enrich the training data. This regular operation reduced the overfitting risk and boosted a classifier's performance. The network was optimized using the Adam optimizer (beta1: 0.9, beta2: 0.999, initial learning rate: 0.0001), with an L2 penalty of 0.01, batch size of 50, and cross-entropy cost function. The maximum training epoch was set to 100, and the model was saved when the maximum accuracy was achieved on the testing set. The CNN models were then implemented using the "keras" and "tensorflow" packages and trained on an AMD Ryzen Threadripper 2970WX processor with 48 Gb RAM (USA).

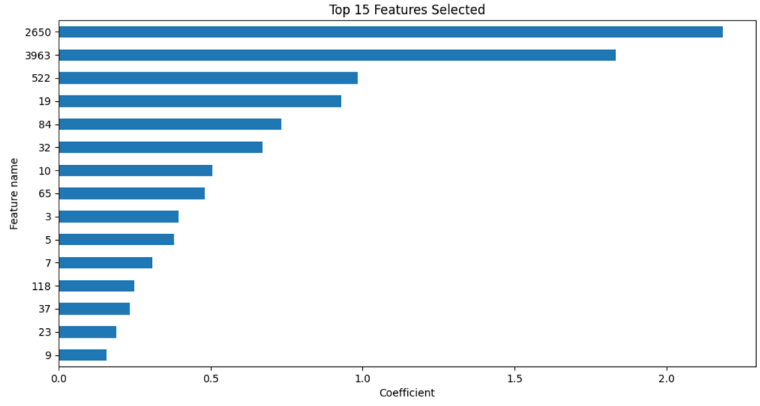

Figure 2. Top features selected.

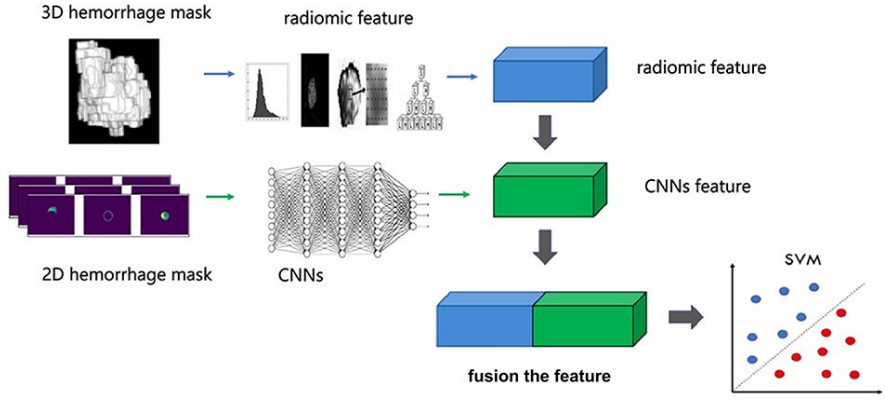

Figure 3. Schematic of 2D and 3D image feature fusion. CNN, convolutional neural network; SVM, support vector machine.

The features selected are shown in Figure 2. The feature-level fusion approaches were used to fuse the features to collect complementary information from radiomics and CNNs and improve diagnostic accuracy (Figure 3).[40] Following the fusion pipeline, sequential forward feature elimination was chosen for feature selection, which was able to automatically identify the subset of features that are highly appropriate to the problem. The grid search parameters consisted of many features that account for data complexity and separability. Next, the "support vector machine" was used as the ML model for classification. The training set was divided into two stratified sets for cross-validation, and the estimator with the best average accuracy in the cross-validation was chosen as the best estimator. Subsequently, 100 non-expanding (NEG-ICH) and expanding (POS-ICH) cases were randomly selected as the test set, and the remaining cases were used as training and validation sets. The radiomics-specific reporting [checklist for evaluation of radiomics research (CLEAR)[41] and radiomics quality score (RQS)[42]] are presented as Supplementary Materials S1 and S2. Thirty-seven items were addressed in the CLEAR checklist, and the RQS score was 10 (27.78%).

Statistical analysis

Model performance was evaluated by metrics. The area under the receiver operating characteristic curve (AUC) was determined to assess the ability of a classifier to distinguish between classes. The confusion matrix (CM) denotes instances in a predicted class, whereas columns denote instances in an actual class. Precision indicates the accuracy of the classifier in identifying positive samples. Recall indicates the classifier's ability to find all positive samples. The area under the precision-recall curve is the average precision (AP). Python software with the "Scikit-Learn" package (https://scikit-learn.org/stable/) was used for statistical analyses.

Results

The schematic of patient recruitment is shown in Figure 4. There were 280 cases included in the analysis, of which 180 were randomly grouped into the training and validation sets (43 POS-ICH cases and 137 NEG-ICH cases) and 100 were randomly grouped into the test sets (24 POS-ICH cases and 76 NEG-ICH cases). Model 1, which automatically labeled the hematoma area and calculated the hematoma volume (Figure 5), was able to efficiently mask the hematoma. The 213 NEG-ICH cases and 67 POS-ICH cases were then classified and labeled. The NEG-ICH group had a male/female ratio of 153:60, an average age of 64.4 years (range, 50–85 years), and a hematoma volume ranging from 0.271 to 79.6 mL. The POS-ICH group had a male/female ratio of 37:30, an average age of 68.7 years (range, 53–89 years), and a hematoma volume ranging from 0.464 to 107.01 mL (Table 1a). These features were not markedly different between the training and testing datasets. The baseline features of the training, validation, and testing sets are shown in Table 1b.

In model 2, the "PyRadiomics" package was used for the extraction of shape-based characteristics from the hematoma masks, with some of the features extracted from the library being automatically deprecated. For each case, 107 features belonging to "Shape Features", "First Order", "Gray Level Dependence Matrix", "Gray Level Co-occurrence Matrix", "Gray Level Size Zone Matrix", "Gray Level Run-Length Matrix", and "Neighborhood Gray Tone Difference Matrix" were obtained using PyRadiomics. The results in the test dataset with the optimal estimator are presented in Figure 6 and Table 2. The radiomics feature module had the highest

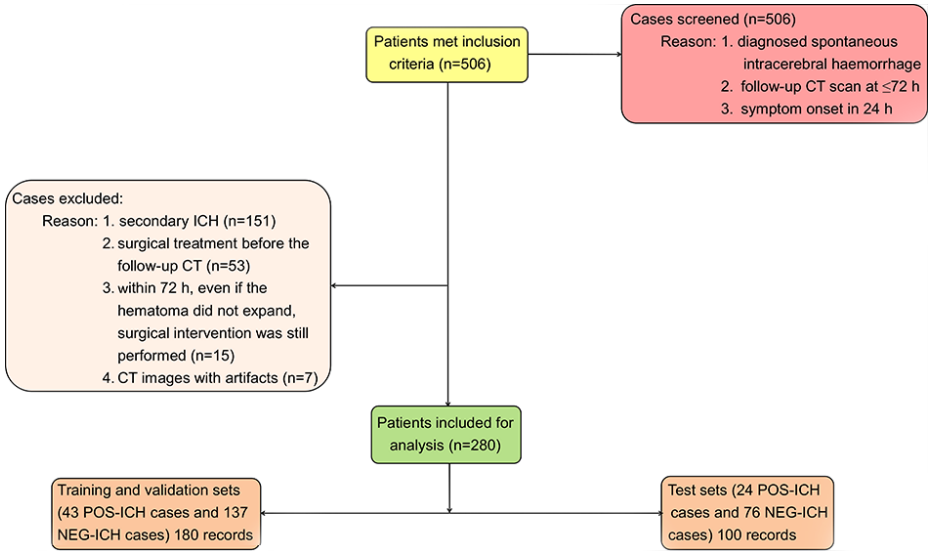

Figure 4. Flow diagram of ICH patient selection. ICH, intracerebral hemorrhage; CT, computed tomography; POS-ICH, positive hematoma expansion in intracerebral hemorrhage; NEG-ICH, negative hematoma expansion in intracerebral hemorrhage.

Table 1. (a, b) Baseline clinical features of included participants

a. NEG-ICH vs. POS-ICH

Variables	NEG-ICH	POS-ICH	*P* value*
Gender (male/female)	153/60	37/30	0.24
Age (years) [median (range)]	64.4 (50–85)	68.7 (53–89)	0.58
Hematoma volume (mL)	0.271–79.6	0.464–107.01	0.48

b. Training and validation vs. test sets

Variables	Training and validation sets		Test sets		*P* value*	
	NEG-ICH	POS-ICH	NEG-ICH	POS-ICH	NEG-ICH	POS-ICH
Gender (male/female)	98/39	24/19	55/21	13/11	0.27	0.21
Age (years) [median (range)]	64.3 (50–86)	68.4 (53–86)	64.5 (52–88)	68.9 (56–89)	0.61	0.56
Hematoma volume (mL)	0.271–79.3	0.464–106.02	0.273–79.6	0.466–107.01	0.50	0.45

*$P < 0.05$. NEG-ICH, negative hematoma expansion in intracerebral hemorrhage; POS-ICH, positive hematoma expansion in intracerebral hemorrhage.

AUC of 0.58, a precision of 0, a recall of 0, and an AP of 0.26. The CM showed that all POS-ICH cases were identified as NEG-ICH cases, indicating that this module could not detect hematoma expansion.

For CNN features, the ResNet50 and Inception_v3 modules were compared using various CNNs as feature extractors for the optimization of predictive performance, and the results of the multi-slice CNN performances in test sets are shown in Figures 7 and 8 and Table 2. The ResNet50 and Inception_v3 modules had AUCs of 0.79 and 0.93, a precision of 0.56 and 0.86, a recall of 0.42 and 0.75, and an AP of 0.51 and 0.85, respectively.

Next, the radiomics and CNN features were fused to assess if the predictive capability could be improved. The CNN features were extracted from the last hidden layer of images before outputting (dimensional feature vectors, 4,096 for Inception_v3, 6,144 for ResNet50). The results of the two fusion models are shown in Figures 9 and 10 and Table 2.

Table 2. Predictive performance of each module

Module	AUC	Precision	Recall	AP
Radiomics	0.58	0	0	0.26
ResNet50	0.79	0.56	0.42	0.51
Inception_v3	0.93	0.86	0.75	0.85
Radiomic + Inception_v3	0.95	0.90	0.79	0.87
Radiomic + ResNet50	0.81	0.57	0.17	0.69

AUC, area under the receiver operating characteristic curve; AP, average precision.

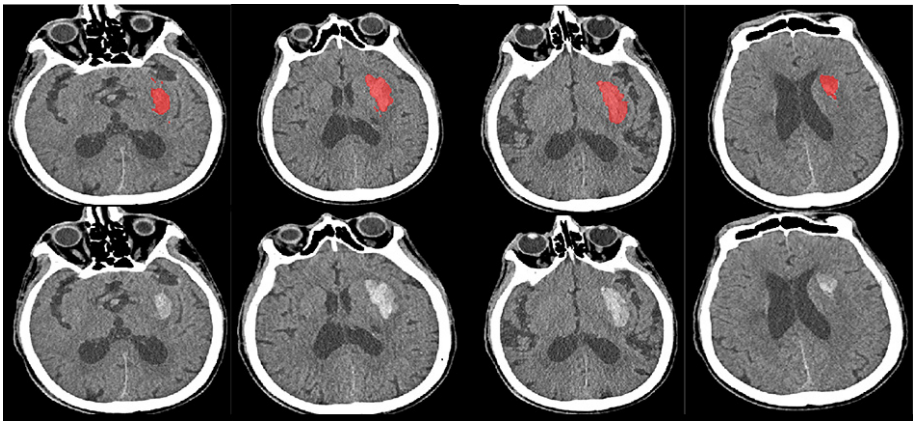

Figure 5. Regions of interest were auto-segmented (indicated in red).

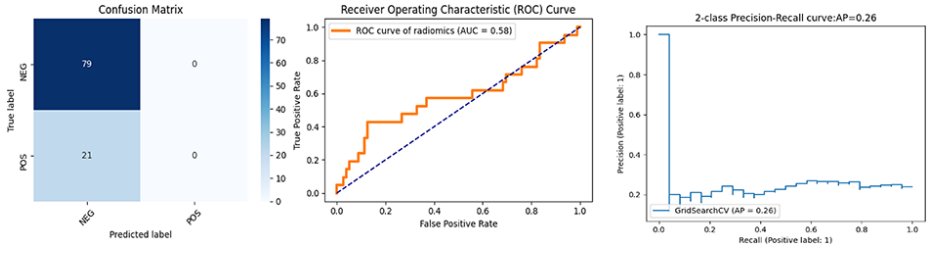

Figure 6. Predictive performance of the radiomics feature module. AP, average precision; AUC, area under the receiver operating characteristic curve.

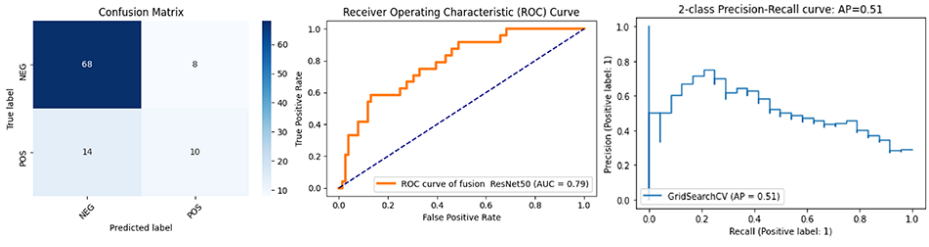

Figure 7. Predictive performance of the CNNs feature module with ResNet50. CNN, convolutional neural network; AP, average precision; AUC, area under the receiver operating characteristic curve.

Radiomic with Inception_v3 and Radiomic with ResNet50 had AUCs of 0.95 and 0.81, a precision of 0.90 and 0.57, a recall of 0.79 and 0.17, and an AP of 0.87 and 0.69, respectively. A summary of the analysis is included in doi: 10.5281/zenodo.10570452. The current analyses showed that the radiomics fusion with the Inception_v3 model had the highest accuracy, indicating that improved performance was obtained relative to the radiomics and CNN modules alone.

Discussion

In this study, we developed a hybrid model that incorporates data from 2D hematoma signal intensities and 3D hematoma shapes into one CNN, along with CNN-based automated hematoma segmentation and a fully automated pipeline, without any operator-dependent processes. The developed automated hybrid model can predict hematoma expansion prediction with high AUC (0.95), precision (0.90), recall (0.79), and AP (0.87) values. This strategy would also improve understanding of the synergistic nature of fusion classification using CNN-based transfer learning and radiomics features. The data obtained in this study highlight improved classification performance when using fusion strategies, radiomics-related features, or features extracted from CNN transfer learning alone. The performance of the CNN model using the Inception_v3 algorithm based on a training dataset of 180 ICH cases for the prediction of hematoma expansion achieved an AUC of 0.93. With the addition of radiomics features, the hybrid model demonstrated a small improvement in the prediction performance. The performance was moderately effective based on a relatively small case series. The model had both radiological features and internal CNN features, and, therefore, the results were better.

Previous studies on the prediction of ICH expansion can be divided into two categories. The first category is a prediction model based on radiological characteristics, which are subjectively judged by image signs, including blend signs,[19] swirl signs,[20] black hole signs,[21] island signs,[22] and satellite signs.[23] These signs were combined with clinical factors to develop a predictive score table, such as the BAT score table,[43] NAG score,[44] HEAVN score,[45] HEP score,[46] the 9-point score,[47] and the BRAIN score.[48] Generally, these studies differ in indicators in their predictive scoring systems, but their radiological features may be associated with low sensitivity and low-quality results,

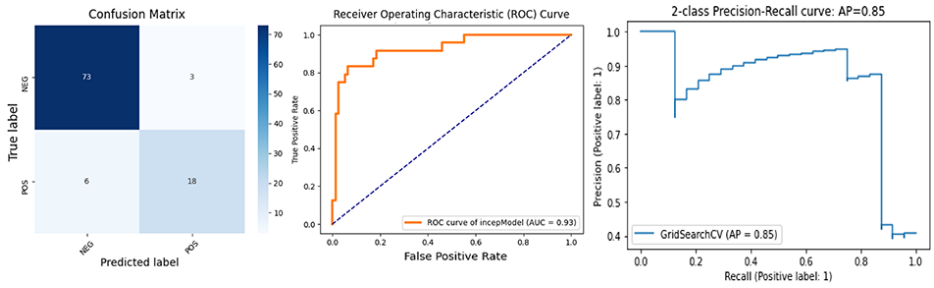

Figure 8. Predictive performance of the CNNs feature module with Inception_v3. CNN, convolutional neural network; AP, average precision; AUC, area under the receiver operating characteristic curve.

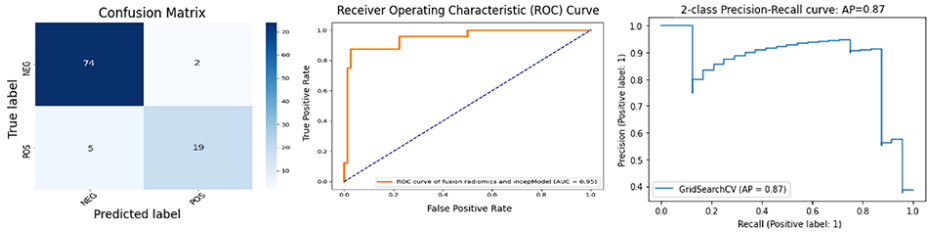

Figure 9. Predictive performance of the fusion model combining Radiomic and Inception_v3. AP, average precision; AUC, area under the receiver operating characteristic curve.

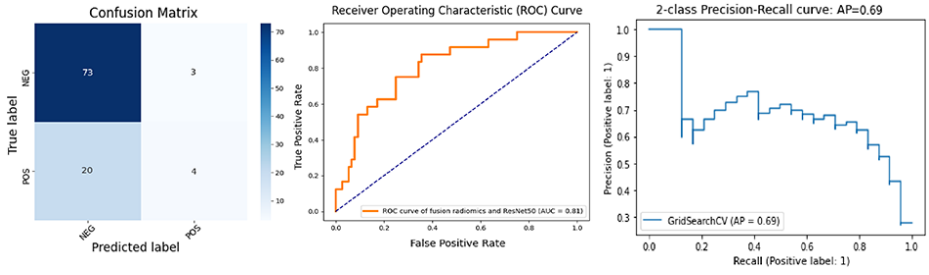

Figure 10. Predictive performance of the fusion model combining Radiomic and ResNet50. AP, average precision; AUC, area under the receiver operating characteristic curve.

as these visual assessments are susceptible to inter-observer variability. Furthermore, previous evaluation methods for ICH expansion were separated but not mixed. However, the relationship between predictors and hematoma enlargement is complex, which would make accurate predictions of hematoma enlargement challenging. As ML can potentially overcome these challenges, another prediction model based on radiomics has emerged. Previously, most prediction models using radiomics were based on traditional ML, which requires manual hematoma segmentation, feature extraction, screening, and reduction.[30-33,49,50] Although these models have effective predictive performance, the execution is not fully automated. Moreover, operator-dependent processes, including manual segmentation, are laborious and could cause interrater variability, which limits their clinical applications. Deep learning strategies, particularly CNN, have been used to solve this problem. CNNs and radiomics analyses are typical quantitative methods for image analysis and can extract high-dimensional as well as abstract numeric data beyond what is perceivable through visual image analysis. Thus, in this study, we first used deep learning to develop a prediction model in which hematoma segmentation and feature extraction were automated, while the radiomics and CNN features were merged into a hybrid prediction model. This fusion model can combine complementary features and improve robustness by reducing the uncertainty of each feature or mode.

Although the radiomics feature module had a high AUC (0.58), 0 accuracy, and 0 recall, the predictive performance of feature-based radiomics in this study was not as effective as that reported in previous studies. This was probably because of a lack of clinical information in the current study. Generally, previous studies obtained various patients' features, including demographic factors, neurological status, and laboratory test parameters. However, these characteristics may not be accessible at all hospitals and may not cover the full spectrum of predictive information that can be acquired from patients. Notably, the mixed model in this study has effective predictive performance without clinical information, indicating that it has greater clinical feasibility.

There are some limitations to this study. First, it is a single-center retrospective study with a small sample size. Further investigations with large sample sizes from multiple centers are required to verify the current results. In this study, we used CNNs to extract high-level features using two different modules (Inception_v3 and ResNet50). Three axial slices were selected to yield three individual samples per patient. Therefore, the convolutional residual network and the CNN classifier were trained on 540 (180 × 3) samples and tested on 300 (100 × 3) samples. Second, even though CNNs exclude the processes of feature computation as well as selection using convolutions and can directly capture key characteristics from the images, signal intensity-based CNNs cannot directly capture 3D shapes of hematoma. Furthermore, model 2 can integrate a hematoma image and shape it into one CNN, but the developed model is not end to end. Model 1 for hematoma segmentation and model 2 for ICH status prediction were separately trained and pooled. However, the image preprocessing, models 1 and 2, and the pipeline in between are completely based on open-source modules that can be integrated into a Python-based pipeline, and a test sample can be automatically run through each step of our model.

In conclusion, we developed a model using deep learning and radiomics that can reliably predict hematoma expansion in ICH in a fully automated process based on NCCT imaging. Hematoma expansion was the only indicator of ICH outcome, and prognostic factors, such as neurological deterioration and mortality, were not included. Due to its high reproducibility and generalizability, this model can be applied more widely in predicting prognosis in acute ICH.

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Krishnamurthi RV, Feigin VL, Forouzanfar MH, et al. Global and regional burden of first-ever ischaemic and haemorrhagic stroke during 1990-2010: findings from the Global Burden of Disease Study 2010. *Lancet Glob Health*. 2013;1(5):259-281. [CrossRef]

2. 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;9(2):167-176. [CrossRef]
3. Dowlatshahi D, Demchuk AM, Flaherty ML, et al. Defining hematoma expansion in intracerebral hemorrhage: relationship with patient outcomes. *Neurology*. 2011;76(14):1238-1244. [CrossRef]
4. Poon MT, Fonville AF, Al-Shahi Salman R. Long-term prognosis after intracerebral haemorrhage: systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry*. 2014;85(6):660-667. [CrossRef]
5. Davis SM, Broderick J, Hennerici M, et al. Hematoma growth is a determinant of mortality and poor outcome after intracerebral hemorrhage. *Neurology*. 2006;66(8):1175-1181. [CrossRef]
6. Broderick JP, Brott TG, Duldner JE, Tomsick T, Huster G. Volume of intracerebral hemorrhage. A powerful and easy-to-use predictor of 30-day mortality. *Stroke*. 1993;24(7):987-993. [CrossRef]
7. Hemphill JC 3rd, Bonovich DC, Besmertis L, Manley GT, Johnston SC. The ICH score: a simple, reliable grading scale for intracerebral hemorrhage. *Stroke*. 2001;32(4):891-897. [CrossRef]
8. Delcourt C, Huang Y, Arima H, et al. Hematoma growth and outcomes in intracerebral hemorrhage: the INTERACT1 study. *Neurology*. 2012;79(4):314-319. [CrossRef]
9. Liotta EM, Prabhakaran S, Sangha RS, et al. Magnesium, hemostasis, and outcomes in patients with intracerebral hemorrhage. *Neurology*. 2017;89(8):813-819. [CrossRef]
10. Brouwers HB, Greenberg SM. Hematoma expansion following acute intracerebral hemorrhage. *Cerebrovasc Dis*. 2013;35(3):195-201. [CrossRef]
11. Hall AN, Weaver B, Liotta E, et al. Identifying modifiable predictors of patient outcomes after intracerebral hemorrhage with machine learning. *Neurocrit Care*. 2021;34(1):73-84. [CrossRef]
12. Yogendrakumar V, Ramsay T, Fergusson DA, et al. Redefining hematoma expansion with the inclusion of intraventricular hemorrhage growth. *Stroke*. 2020;51(4):1120-1127. [CrossRef]
13. Peng WJ, Reis C, Reis H, Zhang J, Yang J. Predictive value of CTA spot sign on hematoma expansion in intracerebral hemorrhage patients. *Biomed Res Int*. 2017;2017:4137210. [CrossRef]
14. Xu X, Zhang J, Yang K, Wang Q, Xu B, Chen X. Accuracy of spot sign in predicting hematoma expansion and clinical outcome: a meta-analysis. *Medicine (Baltimore)*. 2018;97(34):e11945. [CrossRef]
15. Orito K, Hirohata M, Nakamura Y, et al. Leakage sign for primary intracerebral hemorrhage: a novel predictor of hematoma growth. *Stroke*. 2016;47(4):958-963. [CrossRef]
16. Huang Y, Zhang Q, Yang M. A reliable grading system for prediction of hematoma expansion in intracerebral hemorrhage in the basal ganglia. *Biosci Trends*. 2018;12(2):193-200. [CrossRef]
17. Dowlatshahi D, Brouwers HB, Demchuk AM, et al. Predicting intracerebral hemorrhage growth with the spot sign: the effect of onset-to-scan time. *Stroke*. 2016;47(3):695-700. [CrossRef]
18. Caplan LR. Recognizing and preventing intracerebral hematoma expansion. *JAMA Neurol*. 2016;73(8):914-915. [CrossRef]
19. Li Q, Zhang G, Huang YJ, et al. Blend sign on computed tomography: novel and reliable predictor for early hematoma growth in patients with intracerebral hemorrhage. *Stroke*. 2015;46(8):2119-2123. [CrossRef]
20. Yu Z, Zheng J, He M, et al. Accuracy of swirl sign for predicting hematoma enlargement in intracerebral hemorrhage: a meta-analysis. *J Neurol Sci*. 2019;399:155-160. [CrossRef]
21. Li Q, Zhang G, Xiong X, et al. Black hole sign: novel imaging marker that predicts hematoma growth in patients with intracerebral hemorrhage. *Stroke*. 2016;47(7):1777-1781. [CrossRef]
22. Li Q, Liu QJ, Yang WS, et al. Island sign: an imaging predictor for early hematoma expansion and poor outcome in patients with intracerebral hemorrhage. *Stroke*. 2017;48(11):3019-3025. [CrossRef]
23. Shimoda Y, Ohtomo S, Arai H, Okada K, Tominaga T. Satellite sign: a poor outcome predictor in intracerebral hemorrhage. *Cerebrovasc Dis*. 2017;44(3-4):105-112. [CrossRef]
24. Boulouis G, Morotti A, Charidimou A, Dowlatshahi D, Goldstein JN. Noncontrast computed tomography markers of intracerebral hemorrhage expansion. *Stroke*. 2017;48(4):1120-1125. [CrossRef]
25. Morotti A, Boulouis G, Dowlatshahi D, et al. Standards for detecting, interpreting, and reporting noncontrast computed tomographic markers of intracerebral hemorrhage expansion. *Ann Neurol*. 2019;86(4):480-492. [CrossRef]
26. Zhang D, Chen J, Xue Q, et al. Heterogeneity signs on noncontrast computed tomography predict hematoma expansion after intracerebral hemorrhage: a meta-analysis. *Biomed Res Int*. 2018;2018:6038193. [CrossRef]
27. Suzuki K. Overview of deep learning in medical imaging. *Radiol Phys Technol*. 2017;10(3):257-273. [CrossRef]
28. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278(2):563-577. [CrossRef]
29. Zhu DQ, Chen Q, Xiang YL, et al. Predicting intraventricular hemorrhage growth with a machine learning-based, radiomics-clinical model. *Aging (Albany NY)*. 2021;13(9):12833-12848. [CrossRef]
30. Pszczolkowski S, Manzano-Patrón JP, Law ZK, et al. Quantitative CT radiomics-based models for prediction of haematoma expansion and poor functional outcome in primary intracerebral haemorrhage. *Eur Radiol*. 2021;31(10):7945-7959. [CrossRef]
31. Li H, Xie Y, Wang X, Chen F, Sun J, Jiang X. Radiomics features on non-contrast computed tomography predict early enlargement of spontaneous intracerebral hemorrhage. *Clin Neurol Neurosurg*. 2019;185:105491. [CrossRef]
32. Ma C, Zhang Y, Niyazi T, et al. Radiomics for predicting hematoma expansion in patients with hypertensive intraparenchymal hematomas. *Eur J Radiol*. 2019;115:10-15. [CrossRef]
33. Xie H, Ma S, Wang X, Zhang X. Noncontrast computer tomography-based radiomics model for predicting intracerebral hemorrhage expansion: preliminary findings and comparison with conventional radiological model. *Eur Radiol*. 2020;30(1):87-98. [CrossRef]
34. Guyon I, Dror G, Lemaire V, Silver DL, Taylor G, Aha DW. Analysis of the IJCNN 2011 UTL challenge. *Neural Netw*. 2012;32:174-178. [CrossRef]
35. Sharrock MF, Mould WA, Ali H, et al. 3D Deep neural network segmentation of intracerebral hemorrhage: development and validation for clinical trials. *Neuroinformatics*. 2021;19(3):403-415. [CrossRef]
36. Hanley DF, Thompson RE, Rosenblum M, et al. Efficacy and safety of minimally invasive surgery with thrombolysis in intracerebral haemorrhage evacuation (MISTIE III): a randomised, controlled, open-label, blinded endpoint phase 3 trial. *Lancet*. 2019;393(10175):1021-1032. [CrossRef]
37. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77(21):104-107. [CrossRef]
38. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. *Cornell University*. 2015. [CrossRef]
39. Paul R, Hawkins SH, Balagurunathan Y, et al. Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma. *Tomography*. 2016;2(4):388-395. [CrossRef]
40. Afshar P, Mohammadi A, Plataniotis KN, Oikonomou A, Benali H. From hand-crafted to deep learning-based cancer radiomics. Challenges and opportunities. *Cornell University*. 2019. [CrossRef]

41. Kocak B, Baessler B, Bakas S, et al. CheckList for evaluation of radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII. *Insights Imaging*. 2023;14(1):75. [CrossRef]
42. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749-762. [CrossRef]
43. Morotti A, Dowlatshahi D, Boulouis G, et al. Predicting intracerebral hemorrhage expansion with noncontrast computed tomography: the BAT score. *Stroke*. 2018;49(5):1163-1169. [CrossRef]
44. Visser WE, Vrijmoeth P, Visser FE, Arts WF, van Toor H, Visser TJ. Identification, functional analysis, prevalence and treatment of monocarboxylate transporter 8 (MCT8) mutations in a cohort of adult patients with mental retardation. *J Stroke Cerebrovasc Dis*. 2018;27(10):2606-2612. [CrossRef]
45. Miyahara M, Noda R, Yamaguchi S, et al. New prediction score for hematoma expansion and neurological deterioration after spontaneous intracerebral hemorrhage: a hospital-based retrospective cohort study. *J Stroke Cerebrovasc Dis*. 2018;27(9):2543-2550. [CrossRef]
46. Yao X, Xu Y, Siwila-Sackman E, Wu B, Selim M. The HEP score: a nomogram-derived hematoma expansion prediction scale. *Neurocrit Care*. 2015;23(2):179-187. [CrossRef]
47. Brouwers HB, Chang Y, Falcone GJ, et al. Predicting hematoma expansion after primary intracerebral hemorrhage. *JAMA Neurol*. 2014;71(2):158-164. [CrossRef]
48. Wang X, Arima H, Al-Shahi Salman R, et al. Clinical prediction algorithm (BRAIN) to determine risk of hematoma growth in acute intracerebral hemorrhage. *Stroke*. 2015;46(2):376-381. [CrossRef]
49. Chen Q, Zhu D, Liu J, et al. Clinical-radiomics nomogram for risk estimation of early hematoma expansion after acute intracerebral hemorrhage. *Acad Radiol*. 2021;28(3):307-317. [CrossRef]
50. Song Z, Guo D, Tang Z, et al. Noncontrast computed tomography-based radiomics analysis in discriminating early hematoma expansion after spontaneous intracerebral hemorrhage. *Korean J Radiol*. 2021;22(3):415-424. [CrossRef]

Supplementary Material 1.

CLEAR Checklist v1.0

Note: Use the checklist in conjunction with the main text for clarification of all items.
Yes, details provided; No, details not provided; n/e, not essential; n/a, not applicable; Page, page number

Section	No.	Item	Yes	No	n/a	Page
Title						
	1	Relevant title, specifying the radiomic methodology	☑	☐	☐	1
Abstract						
	2	Structured summary with relevant information	☑	☐	☐	1
Keywords						
	3	Relevant keywords for radiomics	☑	☐	☐	1
Introduction						
	4	Scientific or clinical background	☑	☐	☐	1-2
	5	Rationale for using a radiomic approach	☑	☐	☐	1-2
	6	Study objective(s)	☑	☐	☐	2
Method						
Study Design	7	Adherence to guidelines or checklists (e.g., CLEAR checklist)	☑	☐	☐	2
	8	Ethical details (e.g., approval, consent, data protection)	☑	☐	☐	2
	9	Sample size calculation	☐	☐	☑	
	10	Study nature (e.g., retrospective, prospective)	☑	☐	☐	2
	11	Eligibility criteria	☑	☐	☐	2
	12	Flowchart for technical pipeline	☑	☐	☐	2
Data	13	Data source (e.g., private, public)	☑	☐	☐	2
	14	Data overlap	☐	☐	☑	
	15	Data split methodology	☑	☐	☐	4
	16	Imaging protocol (i.e., image acquisition and processing)	☑	☐	☐	2-3
	17	Definition of non-radiomic predictor variables	☐	☐	☑	
	18	Definition of the reference standard (i.e., outcome variable)	☑	☐	☐	3
Segmentation	19	Segmentation strategy	☑	☐	☐	3
	20	Details of operators performing segmentation	☐	☐	☑	
Pre-processing	21	Image pre-processing details	☑	☐	☐	3
	22	Resampling method and its parameters	☐	☐	☑	
	23	Discretization method and its parameters	☐	☑	☐	

Supplementary Material 1.

Section	No.	Item	Yes	No	n/a	Page
	24	Image types (e.g., original, filtered, transformed)	☑	☐	☐	3
Feature extraction	25	Feature extraction method	☑	☐	☐	3
	26	Feature classes	☑	☐	☐	4
	27	Number of features	☐	☑	☐	
	28	Default configuration statement for remaining parameters	☐	☑	☐	
Data preparation	29	Handling of missing data	☐	☑	☐	
	30	Details of class imbalance	☑	☐	☐	3
	31	Details of segmentation reliability analysis	☐	☑	☐	
	32	Feature scaling details (e.g., normalization, standardization)	☑	☐	☐	3
	33	Dimension reduction details	☑	☐	☐	3
Modeling	34	Algorithm details	☑	☐	☐	3
	35	Training and tuning details	☑	☐	☐	3-4
	36	Handling of confounders	☐	☑	☐	
	37	Model selection strategy	☑	☐	☐	2
Evaluation	38	Testing technique (e.g., internal, external)	☑	☐	☐	4
	39	Performance metrics and rationale for choosing	☑	☐	☐	4
	40	Uncertainty evaluation and measures (e.g., confidence intervals)	☐	☑	☐	
	41	Statistical performance comparison (e.g., DeLong's test)	☐	☑	☐	
	42	Comparison with non-radiomic and combined methods	☐	☐	☑	
	43	Interpretability and explainability methods	☐	☑	☐	
Results						
	44	Baseline demographic and clinical characteristics	☑	☐	☐	4
	45	Flowchart for eligibility criteria	☑	☐	☐	4
	46	Feature statistics (e.g., reproducibility, feature selection)	☑	☐	☐	4-5
	47	Model performance evaluation	☑	☐	☐	4-5
	48	Comparison with non-radiomic and combined approaches	☐	☐	☑	
Discussion						
	49	Overview of important findings	☑	☐	☐	5-6
	50	Previous works with differences from the current study	☑	☐	☐	5-6
	51	Practical implications	☑	☐	☐	5-6
	52	Strengths and limitations (e.g., bias and generalizability issues)	☑	☐	☐	6

Supplementary Material 1.

Section	No.	Item	Yes	No	n/a	Page
Open Science						
Data availability	53	Sharing images along with segmentation data [n/e]	☑	☐	☐	4-5
	54	Sharing radiomic feature data	☐	☑	☐	
Code availability	55	Sharing pre-processing scripts or settings	☐	☑	☐	
	56	Sharing source code for modeling	☐	☑	☐	
Model availability	57	Sharing final model files	☐	☑	☐	
	58	Sharing a ready-to-use system [n/e]	☐	☑	☐	

Kocak B, Baessler B, Bakas S, Cuocolo R, Fedorov A, Maier-Hein L, Mercaldo N, Müller H, Orlhac F, Pinto Dos Santos D, Stanzione A, Ugga L, Zwanenburg A. CheckList for EvaluAtion of Radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII. Insights Imaging. 2023 May 4;14(1):75. doi: 10.1186/s13244-023-01415-8

Supplementary Material 2.

Image protocol quality - well-documented image protocols (for example, contrast, slice thickness, energy, etc.) and/or usage of public image protocols allow reproducibility/replicability

- [x] protocols well documented
- [x] public protocol used
- [ ] none

Multiple segmentations - possible actions are: segmentation by different physicians/algorithms/software, perturbing segmentations by (random) noise, segmentation at different breathing cycles. Analyse feature robustness to segmentation variabilities

- ○ yes
- ◉ no

Phantom study on all scanners - detect inter-scanner differences and vendor-dependent features. Analyse feature robustness to these sources of variability

- ○ yes
- ◉ no

Imaging at multiple time points - collect images of individuals at additional time points. Analyse feature robustness to temporal variabilities (for example, organ movement, organ expansion/shrinkage)

- ○ yes
- ◉ no

Feature reduction or adjustment for multiple testing - decreases the risk of overfitting. Overfitting is inevitable if the number of features exceeds the number of samples. Consider feature robustness when selecting features

- ◉ Either measure is implemented
- ○ Neither measure is implemented

Multivariable analysis with non radiomics features (for example, EGFR mutation) - is expected to provide a more holistic model. Permits correlating/inferencing between radiomics and non radiomics features

- ○ yes
- ◉ no

Detect and discuss biological correlates - demonstration of phenotypic differences (possibly associated with underlying gene–protein expression patterns) deepens understanding of radiomics and biology

- ○ yes
- ◉ no

Cut-off analyses - determine risk groups by either the median, a previously published cut-off or report a continuous risk variable. Reduces the risk of reporting overly optimistic results

- ○ yes
- ◉ no

Supplementary Material 2.

Discrimination statistics - report discrimination statistics (for example, C-statistic, ROC curve, AUC) and their statistical significance (for example, p-values, confidence intervals). One can also apply resampling method (for example, bootstrapping, cross-validation)

- [ ] a discrimination statistic and its statistical significance are reported
- [x] a resampling method technique is also applied
- [ ] none

Calibration statistics - report calibration statistics (for example, Calibration-in-the-large/slope, calibration plots) and their statistical significance (for example, P-values, confidence intervals). One can also apply resampling method (for example, bootstrapping, cross-validation)

- [ ] a calibration statistic and its statistical significance are reported
- [x] a resampling method technique is applied
- [ ] none

Prospective study registered in a trial database - provides the highest level of evidence supporting the clinical validity and usefulness of the radiomics biomarker

- ( ) yes
- (•) no

Validation - the validation is performed without retraining and without adaptation of the cut-off value, provides crucial information with regard to credible clinical performance

- [ ] No validation
- [x] validation is based on a dataset from the same institute
- [ ] validation is based on a dataset from another institute
- [ ] validation is based on two datasets from two distinct institutes
- [ ] the study validates a previously published signature
- [ ] validation is based on three or more datasets from distinct institutes

Comparison to 'gold standard' - assess the extent to which the model agrees with/is superior to the current 'gold standard' method (for example, TNM-staging for survival prediction). This comparison shows the added value of radiomics

- ( ) yes
- (•) no

Potential clinical utility - report on the current and potential application of the model in a clinical setting (for example, decision curve analysis).

- ( ) yes
- (•) no

Cost-effectiveness analysis - report on the cost-effectiveness of the clinical application (for example, QALYs generated)

- ( ) yes
- (•) no

Supplementary Material 2.

Open science and data - make code and data publicly available. Open science facilitates knowledge transfer and reproducibility of the study

- [ ] scans are open source
- [x] region of interest segmentations are open source
- [ ] the code is open sourced
- [ ] radiomics features are calculated on a set of representative ROIs and the calculated features and representative ROIs are open source

Total score **10** (27.78%)