



## Research article

## Magnetic resonance imaging-based deep learning imaging biomarker for predicting functional outcomes after acute ischemic stroke

Tzu-Hsien Yang<sup>a,1</sup>, Ying-Ying Su<sup>a,1</sup>, Chia-Ling Tsai<sup>b</sup>, Kai-Hsuan Lin<sup>c</sup>, Wei-Yang Lin<sup>c,d,\*</sup>, Sheng-Feng Sung<sup>e,f,\*</sup><sup>a</sup> Department of Radiology, Ditmanson Medical Foundation Chia-Yi Christian Hospital, Chiayi City, Taiwan<sup>b</sup> Computer Science Department, Queens College, City University of New York, Flushing, NY, USA<sup>c</sup> Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi, Taiwan<sup>d</sup> Advanced Institute of Manufacturing with High-Tech Innovations, National Chung Cheng University, Chiayi, Taiwan<sup>e</sup> Division of Neurology, Department of Internal Medicine, Ditmanson Medical Foundation Chia-Yi Christian Hospital, Chiayi City, Taiwan<sup>f</sup> Department of Beauty & Health Care, Min-Hwei Junior College of Health Care Management, Tainan, Taiwan

## ARTICLE INFO

## Keywords:

Acute ischemic stroke  
Deep learning  
Prognosis  
Radiomics  
Risk score

## ABSTRACT

**Purpose:** Clinical risk scores are essential for predicting outcomes in stroke patients. The advancements in deep learning (DL) techniques provide opportunities to develop prediction applications using magnetic resonance (MR) images. We aimed to develop an MR-based DL imaging biomarker for predicting outcomes in acute ischemic stroke (AIS) and evaluate its additional benefit to current risk scores.**Method:** This study included 3338 AIS patients. We trained a DL model using deep neural network architectures on MR images and radiomics to predict poor functional outcomes at three months post-stroke. The DL model generated a DL score, which served as the DL imaging biomarker. We compared the predictive performance of this biomarker to five risk scores on a holdout test set. Additionally, we assessed whether incorporating the imaging biomarker into the risk scores improved the predictive performance.**Results:** The DL imaging biomarker achieved an area under the receiver operating characteristic curve (AUC) of 0.788. The AUCs of the five studied risk scores were 0.789, 0.793, 0.804, 0.810, and 0.826, respectively. The imaging biomarker's predictive performance was comparable to four of the risk scores but inferior to one ( $p = 0.038$ ). Adding the imaging biomarker to the risk scores improved the AUCs ( $p$ -values) to 0.831 (0.003), 0.825 (0.001), 0.834 (0.003), 0.836 (0.003), and 0.839 (0.177), respectively. The net reclassification improvement and integrated discrimination improvement indices also showed significant improvements (all  $p < 0.001$ ).**Conclusions:** Using DL techniques to create an MR-based imaging biomarker is feasible and enhances the predictive ability of current risk scores.

## 1. Introduction

Stroke is one of the leading causes of death and adult disability worldwide, and its incidence and prevalence have increased over the past 30 years [1]. About 25 percent of adults aged 25 or older will experience a stroke during their lifetime [2]. Predicting functional outcomes after a stroke is essential for clinicians to make informed treatment decisions. It also provides patients and their families with valuable information regarding the expected recovery process.

Numerous clinical risk scores have been developed using regression analysis based on a combination of patient characteristics and clinical data to predict functional outcomes following acute ischemic stroke (AIS) [3–5]. In addition to conventional regression-based methods, machine learning algorithms have become popular in constructing prognostic models for AIS [6,7]. Moreover, infarct volume and location significantly predict functional outcomes in AIS [8,9]. Therefore, imaging features derived from non-contrast computed tomography (CT) [10,11], CT perfusion [12], CT angiography [11], and magnetic

\* Corresponding authors at: Department of Computer Science and Information Engineering, National Chung Cheng University, 168 University Road, Minhsiung, Chiayi 621301, Taiwan (W.-Y.L.); and Division of Neurology, Department of Internal Medicine, Ditmanson Medical Foundation Chia-Yi Christian Hospital, 539 Zhongxiao Road, East District, Chiayi City 60002, Taiwan.

E-mail addresses: [wylin@cs.ccu.edu.tw](mailto:wylin@cs.ccu.edu.tw) (W.-Y. Lin), [sfsung@cych.org.tw](mailto:sfsung@cych.org.tw) (S.-F. Sung).

<sup>1</sup> Tzu-Hsien Yang and Ying-Ying Su contributed equally to this work.

<https://doi.org/10.1016/j.ejrad.2024.111405>

Received 22 December 2023; Received in revised form 5 February 2024; Accepted 1 March 2024

Available online 2 March 2024

0720-048X/© 2024 Elsevier B.V. All rights reserved.

resonance (MR) imaging [13–17] are increasingly used alongside clinical data in the prognostication of AIS.

Although clinical risk scores can be easily calculated or derived, researchers constantly search for new biomarkers to enhance or refine risk prediction [18,19]. With the advancements in artificial intelligence, deep learning (DL) techniques have shown impressive capabilities in medical diagnoses and prognostication. Image biomarkers directly extracted from histology or radiographic images using DL have shown potential for predicting survival and therapy response in cancer patients [20,21].

Motivated by the studies mentioned above, this study aimed to investigate whether a DL-based imaging biomarker could help predict functional outcomes of AIS. Specifically, we constructed a DL model that generates a DL score, representing the probability of poor recovery, and assessed its added value to clinical risk scores.

## 2. Material and methods

### 2.1. Study settings and data sources

The single-center retrospective study took place in a tertiary teaching hospital with a certified comprehensive stroke center. The study hospital employed non-contrast CT and multi-phase CT angiography for screening AIS patients for reperfusion therapies, including intravenous thrombolysis and mechanical thrombectomy. MR imaging was performed for follow-up.

The study data was obtained from the hospital stroke registry, which collected data on demographic characteristics, risk factors, comorbidities, interventions, complications, and outcomes of stroke patients. Stroke severity was assessed using the National Institutes of Health Stroke Scale (NIHSS), while functional outcomes were evaluated using the modified Rankin Scale (mRS). Brain MR images were obtained from the hospital’s picture archiving and communication system (PACS). The study protocol was approved by the Institutional Review Board of the

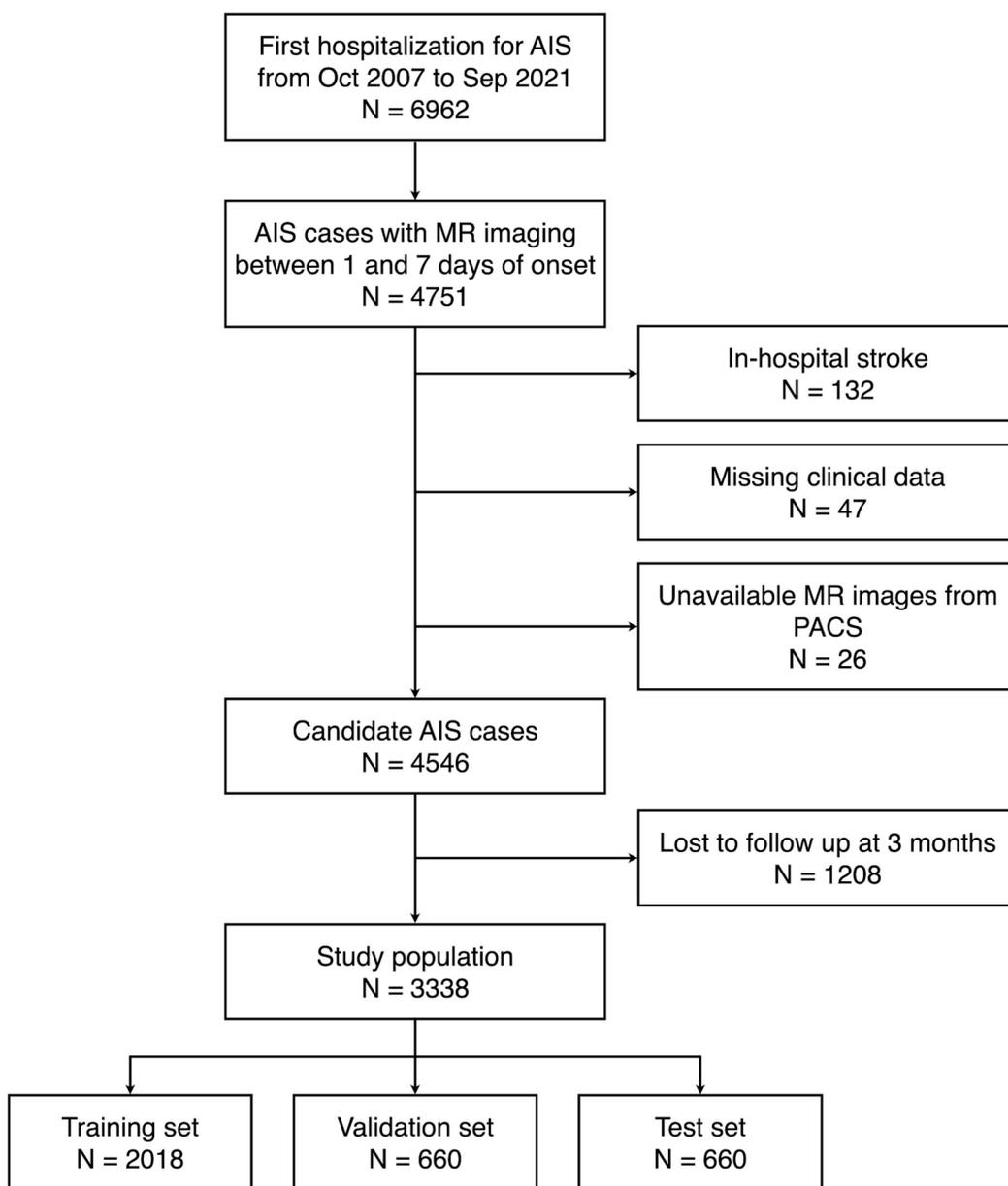


Fig. 1. Flowchart showing the derivation of the study population. AIS: acute ischemic stroke; MR: magnetic resonance; PACS: picture archiving and communication system.

study hospital.

2.2. Study population

This study focused on adult patients hospitalized for AIS within 10 days of stroke onset between October 2007 and September 2021 (Fig. 1). Only the earliest hospitalization for each patient was considered.

Inclusion criteria required patients to have undergone MR imaging between 1 and 7 days post-stroke. Exclusion criteria were in-hospital strokes, patients with missing clinical data, and those with unavailable MR images. Additionally, patients who were lost to follow-up at three months post-stroke were eliminated from the study.

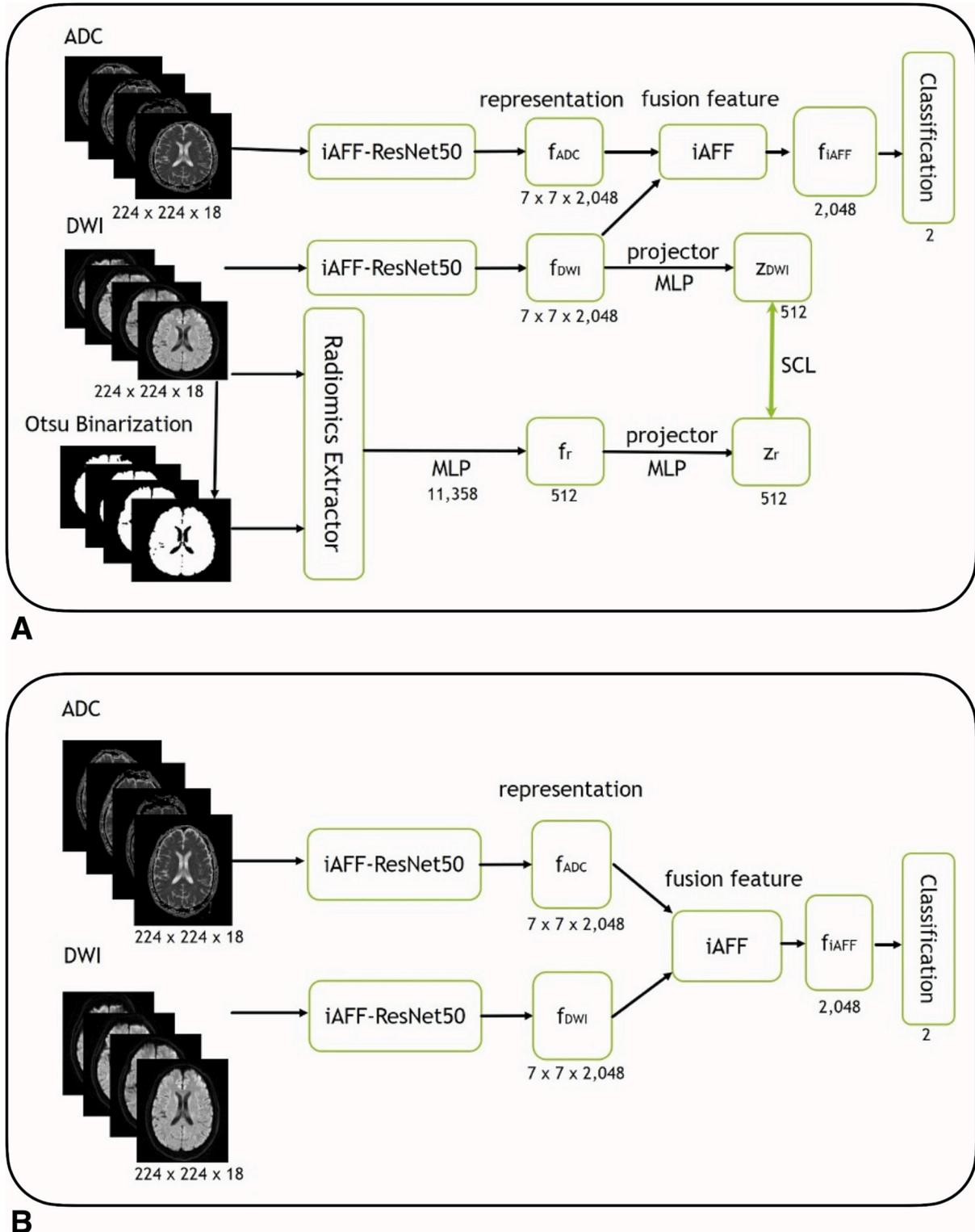


Fig. 2. The end-to-end network architectures for training (A) and prediction (B). ADC: apparent diffusion coefficient; DWI: diffusion-weighted imaging; iAFF: iterative attentional feature fusion; MLP: multilayer perceptron; ResNet50: 50-layer residual network; SCL: supervised contrastive learning.

### 2.3. Clinical risk scores

We utilized five clinical risk scores (Supplementary Table 1) that rely on clinical variables available at admission as the original prognostic models for comparison. The first model was the NIHSS score. The second model was the Stroke Prognostication using Age and NIHSS (SPAN) index, calculated by adding the patient's age and NIHSS score. It is a simple method for predicting functional outcomes in stroke patients treated with intravenous thrombolysis [3] and those undergoing mechanical thrombectomy [22]. The third model was the preadmission comorbidities, level of consciousness, age, and neurologic deficit (PLAN) score [23], which was developed to predict 30-day and 1-year mortality, as well as an mRS score of 5–6 at discharge. The fourth model, the Dutch Stroke Score (DSS) [5], was developed to predict a poor functional outcome (mRS > 2) at three months post-stroke. The fifth model was derived from a cohort of patients from the Acute Stroke Registry and Analysis of Lausanne (ASTRAL) [24]. It was designed to predict a 3-month poor functional outcome (mRS > 2).

### 2.4. Outcome variable

The outcome variable was a poor functional outcome as assessed using the mRS score at three months post-stroke. The mRS score was dichotomized into a poor functional outcome (mRS score of 3–6) versus a good functional outcome (mRS score of 0–2).

### 2.5. Deep learning imaging biomarker

For additional details, please consult the Supplementary Methods. The subsequent paragraphs offer a concise overview of the modeling process.

We developed an end-to-end DL model to predict outcomes at three months post-stroke. The network utilized DWI ( $b = 1000 \text{ s/mm}^2$ ) and ADC image sequences as inputs. During the training phase, we additionally used PyRadiomics to extract 11,358 radiomic features [25], including first-order statistical, grayscale, and shape-based features, from each DWI image sequence. The whole brain parenchyma was used for extracting radiomic features.

The network has different architectures for training and prediction. Fig. 2A illustrates the architecture for training the network. The lower branch of the training architecture utilizes the 50-layer residual network (ResNet50) [26] with the iterative attentional feature fusion (iAFF) module [27] as an encoder. Additionally, radiomic features undergo dimension reduction by a multilayer perceptron (MLP) and are projected to  $Z_r$  using a second MLP. Supervised contrastive learning is employed to align the feature space of  $Z_{\text{DWI}}$  and  $Z_r$ . This alignment aims to enhance the feature extraction capability of the iAFF-ResNet50 encoder, enabling it to simultaneously learn low-level features extracted directly from images and high-level radiomic features. The upper branch of the architecture employs another iAFF-ResNet50 encoder to extract features from the input ADC sequences. The features extracted from ADC and DWI ( $f_{\text{ADC}}$  and  $f_{\text{DWI}}$ ) are fused and subsequently passed through a fully connected layer to generate classification results.

Fig. 2B illustrates the architecture of the prediction network. This architecture uses only the features extracted from ADC and DWI sequences by the iAFF-ResNet50 encoders as inputs. Instead of directly extracting radiomic features, the encoder for the DWI sequences is utilized because it has already learned how to extract both low-level (image-level) and high-level (radiomic) features during the training stage. The prediction network generated the DL score for each patient in the test set, namely the DL imaging biomarker.

In addition to the proposed model, we trained several baseline models to compare their performance. The first two models used DWI or ADC images as the only input and were trained using the ResNet50 [26]. The third model used radiomic features extracted from DWI images as the input and was trained using a four-layer MLP network. Finally, we

created a model that performed score fusion on the scores generated by the first three baseline models. The performance of each model was evaluated on the validation and test sets, with the primary evaluation metric being the area under the receiver operating characteristic curve (AUC).

### 2.6. Statistical analysis

Categorical variables are presented as counts and percentages, while continuous variables are reported as means (standard deviations) or medians (interquartile ranges). Differences between groups were tested using Chi-square tests for categorical variables and *t*-tests or Mann-Whitney U tests for continuous variables, as appropriate.

Logistic regression was performed by including one of the clinical risk scores and the DL imaging biomarker to determine the independent effect of the biomarker on the outcome. The model discrimination of the original prognostic models and those enhanced by the DL imaging biomarker was evaluated on the test set using the AUC [19]. AUCs were calculated and compared using the DeLong method. The model calibration was assessed using the Hosmer-Lemeshow goodness-of-fit test and a visual inspection of the calibration plot, which shows the observed risk versus the predicted risk. Additionally, the continuous net reclassification improvement (NRI) and integrated discrimination improvement (IDI) indices [28] were estimated. Higher values of NRI and IDI indices indicate superior discrimination.

All statistical analyses were performed using Stata 15.1 (StataCorp, College Station, Texas) and R version 4.2.1 (R Foundation for Statistical Computing, Vienna, Austria). Two-tailed *p*-values of 0.05 were considered significant.

## 3. Results

The study population (Table 1) included 3338 patients (2018 males, 1320 females), with an average age of 68.9 years (standard deviation 12.7 years). Among these patients, 1514 (45.4 %) experienced a poor functional outcome three months after having a stroke. Patients with poor functional outcomes were older, more likely to be female, and more likely to have hypertension, diabetes, atrial fibrillation, congestive heart failure, cancer, and pre-stroke dependence. However, they were less likely to have hyperlipidemia than those with good functional outcomes. Additionally, patients with poor functional outcomes had significantly higher levels of glucose, NIHSS, SPAN, PLAN, DSS, and ASTRAL scores.

### 3.1. Model performance

The training set consisted of 2018 patients, while the validation and test set each included 660 patients (Fig. 1). Supplementary Table 2 presents the results of four baseline models and our model on the validation set. Our proposed model achieved the highest AUC (0.804) on the validation set, which was significantly higher than models using DWI images (0.747,  $p < 0.001$ ), ADC images (0.758,  $p = 0.006$ ), or radiomic features (0.778,  $p = 0.032$ ) as the sole input. However, our proposed model showed comparable performance to the model using score fusion (0.789,  $p = 0.237$ ).

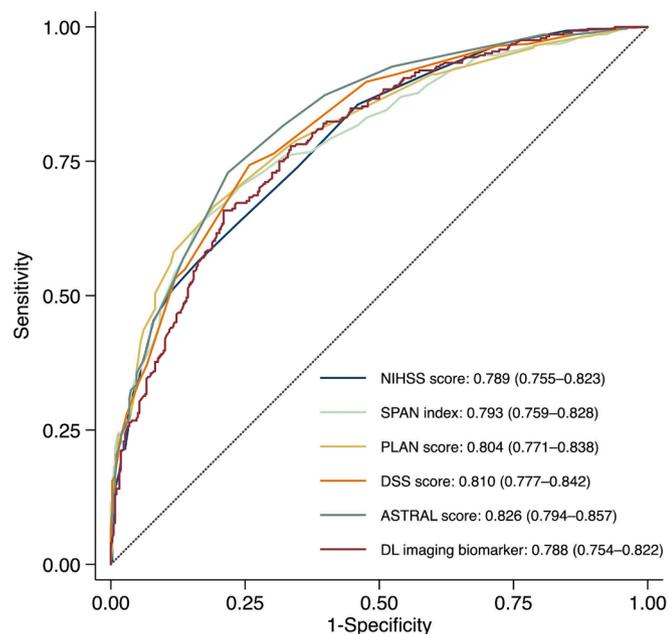
On the test set (Supplementary Table 3), our proposed model achieved the highest AUC (0.788), which was significantly higher than models using DWI images (0.733,  $p = 0.001$ ) or ADC images (0.730,  $p = 0.003$ ) as the only input. It performed similarly to models using radiomic features (0.769,  $p = 0.144$ ) or score fusion (0.780,  $p = 0.566$ ). The DL score generated by our proposed model was used as the DL imaging biomarker in the subsequent analysis.

Fig. 3 displays the receiver operating characteristic curves and AUC values of the original prognostic models and the DL imaging biomarker in the test set. The DL imaging biomarker achieved an AUC of 0.788, which was comparable to the AUC values of the NIHSS (0.789,  $p = 0.947$ ), SPAN (0.793,  $p = 0.778$ ), PLAN (0.804,  $p = 0.376$ ), and DSS

**Table 1**  
Characteristics of the study population.

Characteristic	All N = 3338	Good functional outcome (n = 1824)	Poor functional outcome (n = 1514)	P
Age, mean (SD)	68.9 (12.7)	65.2 (12.5)	73.5 (11.4)	<0.001
Female	1320 (39.5)	617 (33.8)	703 (46.4)	<0.001
Hypertension	2633 (78.9)	1393 (76.4)	1240 (81.9)	<0.001
Diabetes	1404 (42.1)	700 (38.4)	704 (46.5)	<0.001
Hyperlipidemia	1913 (57.3)	1096 (60.1)	817 (54.0)	<0.001
Atrial fibrillation	536 (16.1)	192 (10.5)	344 (22.7)	<0.001
Congestive heart failure	131 (3.9)	44 (2.4)	87 (5.7)	<0.001
Cancer	238 (7.1)	100 (5.5)	138 (9.1)	<0.001
Pre-stroke dependence	290 (8.7)	23 (1.3)	267 (17.6)	<0.001
Onset-to-arrival > 3 h	2449 (73.4)	1332 (73.0)	1117 (73.8)	0.625
Glucose, mean (SD), mmol/L	8.93 (4.44)	8.63 (4.10)	9.30 (4.80)	<0.001
NIHSS, median (IQR)	5 (3–9)	4 (2–6)	8 (5–17)	<0.001
SPAN, median (IQR)	77 (66–86)	71 (61–79)	85 (75–95)	<0.001
PLAN, median (IQR)	8 (6–11)	7 (6–8)	11 (8–16)	<0.001
DSS, median (IQR)	8 (5–11)	6 (4–9)	11 (8–16)	<0.001
ASTRAL, median (IQR)	21 (18–26)	19 (16–21)	26 (21–36)	<0.001

Data are expressed in number (percentage) unless specified otherwise. ASTRAL: Acute Stroke Registry and Analysis of Lausanne; DSS: Dutch Stroke Score; IQR: interquartile range; NIHSS: National Institutes of Health Stroke Scale; PLAN: preadmission comorbidities, level of consciousness, age, and neurological deficit; SD: standard deviation; SPAN: Stroke Prognostication using Age and NIHSS.



**Fig. 3.** Receiver operating characteristic curves for predicting a poor functional outcome in the test set. ASTRAL: Acute Stroke Registry and Analysis of Lausanne; DL: deep learning; DSS: Dutch Stroke Score; NIHSS: National Institutes of Health Stroke Scale; PLAN: preadmission comorbidities, level of consciousness, age, and neurological deficit; SPAN: Stroke Prognostication using Age and NIHSS.

score (0.810,  $p = 0.237$ ), but was inferior to the AUC value of the ASTRAL score (0.826,  $p = 0.038$ ).

Supplementary Table 4 shows the odds ratio and the corresponding  $p$ -value of the DL imaging biomarker in each enhanced model. After controlling for the clinical risk score, the DL imaging biomarker was significantly associated with poor functional outcomes. Table 2 compares the performance of the original and enhanced prognostic models in predicting poor functional outcomes at three months post-stroke. The original prognostic models had AUCs ranging from 0.789 to 0.826. When the DL imaging biomarker was added to the original models, their AUCs increased to a range of 0.825–0.839. The increase in AUC was statistically significant for NIHSS, SPAN, PLAN, and DSS scores. The NRI and IDI indices (Table 2) also showed a statistically significant improvement in predictive performance for all original prognostic models with the addition of the DL imaging biomarker.

Fig. 4 displays the calibration plots of the original and enhanced models. It demonstrates that the DL imaging biomarker-enhanced models were generally better calibrated than the original models. This is evidenced by the closer distribution of all points around the 45-degree line and the lower Hosmer-Lemeshow statistics.

### 3.2. Exemplary cases

We present two case examples to underscore the clinical relevance of these predictions in patients with AIS. The first case involves a 72-year-old woman who exhibited progressive right-sided weakness and slurred speech. She has a medical history of diabetes and hypertension. Her MR imaging revealed a hyperintense area on the DWI (Fig. 5A) and hypointensity on the ADC maps (Fig. 5B) on the left side of the pons. Her scores were as follows: NIHSS 9, SPAN 81, PLAN 11, DSS 11, and ASTRAL 26. Her three-month mRS score was 2, indicating a good functional outcome. Based on these clinical risk scores, she was expected to have a more than 50 % chance of a poor functional outcome. However, when the DL imaging biomarker was added, all enhanced prognostic models predicted a low risk of a poor functional outcome. The predicted probabilities of a poor functional outcome were 28 %, 28 %, 36 %, 36 %, and 33 %, respectively.

In the second case, an 82-year-old man presented with acute dizziness, vomiting, and slurred speech. He has a medical history of hypertension and atrial fibrillation. His MR imaging showed multiple patch-like hyperintense lesions on the DWI (Fig. 5C) with low signals on the ADC maps (Fig. 5D) in bilateral cerebellar hemispheres and right-sided medulla. His scores were as follows: NIHSS 4, SPAN 86, PLAN 15, DSS 8, and ASTRAL 23. His mRS score was 4 at three months post-stroke, meaning a poor functional outcome. According to the NIHSS, DSS, and ASTRAL scores, he had a less than 50 % chance of having a poor functional outcome. However, after incorporating the DL imaging biomarker, the predicted probability of a poor functional outcome increased to 54 %, 56 %, and 60 %, respectively.

## 4. Discussion

This study found that the DL imaging biomarker obtained from MR images can help predict the functional outcome after AIS. This imaging biomarker significantly predicted poor functional outcomes at three months post-stroke, independent of the clinical risk score. Additionally, it enhanced the predictive capability of the clinical risk scores regarding model discrimination and calibration.

### 4.1. Rationale of DL imaging biomarker

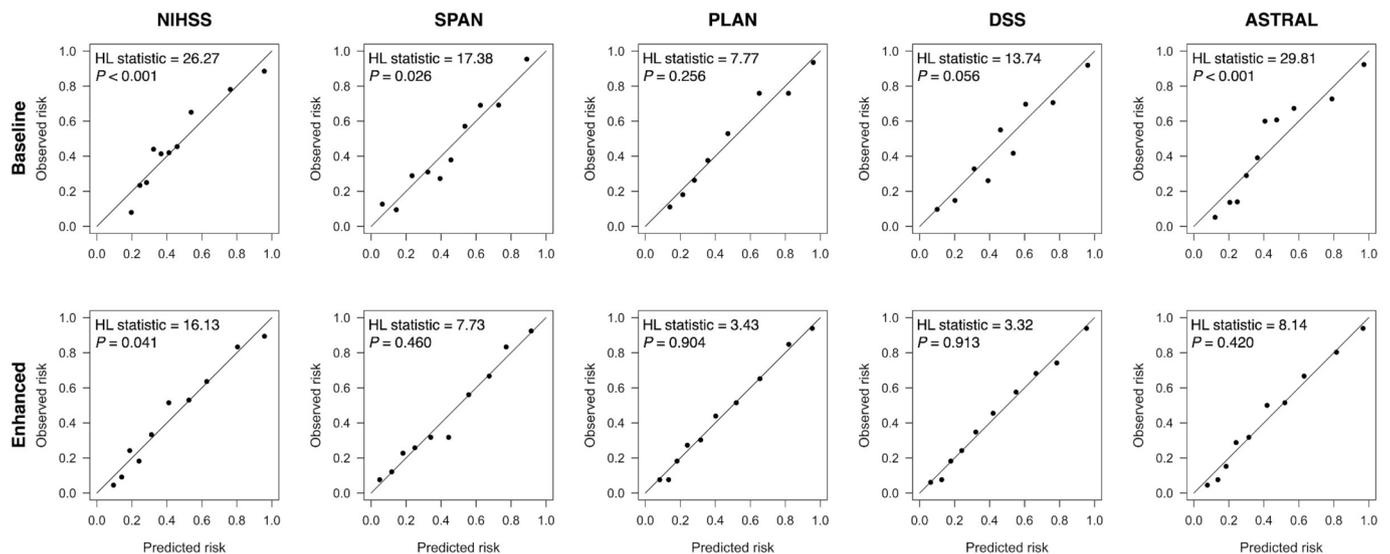
After vessel occlusion, DWI can demonstrate the ischemic brain tissue within minutes, with a reduction of the ADC. DWI is more sensitive than CT in detecting early infarction [29]. Moreover, the infarct volume measured as early as 24 h on DWI strongly correlates with the final infarct volume measured on fluid-attenuated inversion recovery

**Table 2**

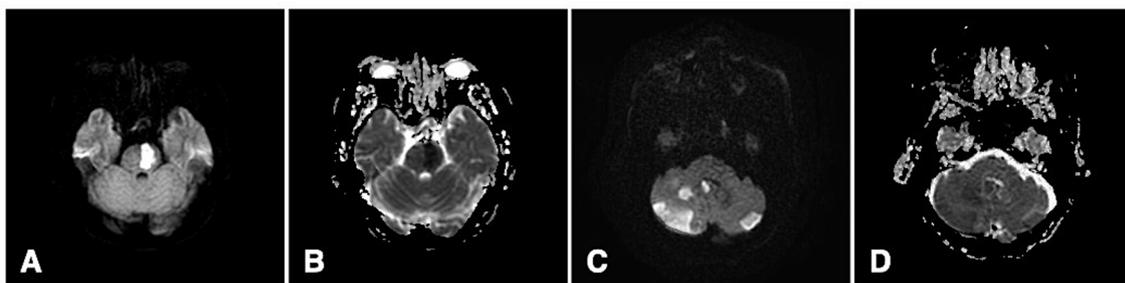
Performance comparison in predicting a poor functional outcome at three months between original and enhanced prognostic models on the test set.

	AUC (95 % CI) (baseline)	AUC (95 % CI) (baseline + imaging DL biomarker)	$\Delta$ AUC	<i>P</i>	NRI, (95 % CI)	<i>P</i>	IDI, (95 % CI)	<i>P</i>
NIHSS	0.789 (0.755–0.823)	0.831 (0.800–0.862)	0.042	0.003	0.638 (0.492–0.784)	<0.001	0.090 (0.067–0.112)	<0.001
SPAN	0.793 (0.759–0.828)	0.825 (0.793–0.856)	0.032	0.001	0.587 (0.439–0.734)	<0.001	0.060 (0.042–0.078)	<0.001
PLAN	0.804 (0.771–0.838)	0.834 (0.803–0.864)	0.030	0.003	0.566 (0.418–0.714)	<0.001	0.055 (0.037–0.073)	<0.001
DSS	0.810 (0.777–0.842)	0.836 (0.805–0.866)	0.026	0.003	0.523 (0.375–0.672)	<0.001	0.052 (0.035–0.069)	<0.001
ASTRAL	0.826 (0.794–0.857)	0.839 (0.809–0.869)	0.013	0.177	0.411 (0.260–0.561)	<0.001	0.054 (0.036–0.071)	<0.001

ASTRAL: Acute Stroke Registry and Analysis of Lausanne; AUC: area under the receiver operating characteristic curve; CI: confidence interval; DL: deep learning; DSS: Dutch Stroke Score; IDI: integrated discrimination improvement; NIHSS: National Institutes of Health Stroke Scale; NRI: net reclassification improvement; PLAN: preadmission comorbidities, level of consciousness, age, and neurological deficit; SPAN: Stroke Prognostication using Age and NIHSS.



**Fig. 4.** Calibration plots of the baseline and DL imaging biomarker-enhanced models. ASTRAL: Acute Stroke Registry and Analysis of Lausanne; DSS: Dutch Stroke Score; HL: Hosmer-Lemeshow; NIHSS: National Institutes of Health Stroke Scale; PLAN: preadmission comorbidities, level of consciousness, age, and neurological deficit; SPAN: Stroke Prognostication using Age and NIHSS.



**Fig. 5.** Magnetic resonance images from two exemplary cases. Case 1 presents with left pontine infarction, as demonstrated by hyperintensity on the DWI (A) and hypointensity on the ADC maps (B). Case 2 exhibits multiple infarcts in the bilateral cerebellar hemispheres and right-sided medulla, indicated by several patch-like hyperintense lesions on the DWI (C) and low signals on the ADC maps (D).

sequences 90 days post-stroke [30]. Previous studies have successfully used DL techniques on MR images to create prognostic models for AIS [14–16,31]. Therefore, we chose DWI and ADC images acquired between 1 and 7 days after stroke onset as the source images to construct the DL imaging biomarker.

On the other hand, radiomics enables the extraction of quantitative features from medical images, providing an objective assessment of visible stroke lesions and subtle changes in surrounding tissues. Several prior studies have utilized radiomic features from CT [10] or MR [13,32–34] images to predict functional outcomes after AIS. Most of these studies required manually segmenting infarct lesions before extracting radiomic features. In addition, they generally used logistic

regression or conventional machine learning algorithms for predictive modeling.

This study differed from previous ones in several ways. Firstly, our model used both MR images and radiomic features as input, allowing for simultaneous learning of high-level radiomic features and low-level imaging features. Secondly, the extracted radiomic features were not directly used for outcome prediction. Instead, they were used to enhance the feature extraction capability of the DL encoder through supervised contrastive learning. As a result, our model does not need to compute radiomic features during the prediction stage. Nevertheless, our model achieved comparable performance to that using score fusion from separate models based on DWI, ADC, or radiomic features,

demonstrating the effectiveness of our DL model training strategy. Thirdly, radiomic features were extracted from the entire brain parenchyma instead of just infarct lesions, eliminating the need for lesion segmentation.

#### 4.2. Imaging biomarkers for stroke prognosis

Imaging biomarkers related to stroke outcomes include the extent and severity of ischemic injury, the degree of hemodynamic compromise, the patency or occlusion of large vessels, the size of salvageable brain tissue, and the presence of hemorrhagic transformation [35]. In specific clinical scenarios, such as predicting outcomes for AIS patients undergoing mechanical thrombectomy, the volume of salvageable brain tissue and the status of blood vessels are critical factors. Imaging techniques like MR perfusion, CT perfusion, and CT angiography are particularly useful in visualizing salvageable brain tissue and vessel status and, thus, are valuable in predicting functional outcomes in these scenarios [11,12,15,36,37].

On the other hand, this study focused on predicting outcomes in a broad range of AIS patients, regardless of whether they received reperfusion therapy. MR images obtained 24 h after stroke onset were used for outcome prediction. These images can offer information about the size and location of infarct lesions, along with the effect of treatments administered during the hyperacute stage of stroke, especially reperfusion therapy.

Many MR biomarkers help predict stroke outcomes. In addition to the size and location of infarct lesions, the characteristics of the brain surrounding the ischemic lesion and the “underlying” brain, representing the pre-stroke cerebral status, are closely related to stroke outcomes [38]. One strong predictor of post-stroke functional outcomes is the pre-stroke medical history, which may be indicated by chronic brain lesions on brain imaging. This idea is supported by a study that used radiomic features from MR images of the entire brain parenchyma to determine the relative brain age, significantly predicting functional outcomes of AIS [33]. Interestingly, the relative brain age was also significantly associated with hypertension, diabetes, atrial fibrillation, coronary artery disease, a history of smoking, and prior stroke. Since the current study utilized images of the whole brain parenchyma, our model could learn from characteristics beyond the infarct, such as the underlying brain.

#### 4.3. Clinical significance and application

Previous studies have shown that information extracted from brain imaging can help predict outcomes in AIS patients. Models combining information from both imaging and clinical data tend to have better predictive performance than models based on imaging or clinical data alone [10,11,15,16,37,39]. This study also found improved predictive performance by combining the DL imaging biomarker with existing clinical risk scores. Instead of developing a clinical-imaging fusion model from scratch, we chose not to do so for the following reasons. Clinical risk scores are typically calculated using readily available clinical variables. Creating a complex DL model using extensive clinical data may cause issues with unstandardized data formats or missing data.

Although the addition of the DL imaging biomarker to these risk scores significantly improved model discrimination, the increases in the AUC were all below 0.1. Similar findings were reported in previous studies examining the added value of biomarkers to existing clinical risk scores, where the AUC increases rarely exceeded 0.05 [18,19]. Given the marginal increase in predictive value by adding the DL imaging biomarker, its clinical relevance in current risk scores is debatable. However, the DL imaging biomarker achieved an AUC comparable to most of the clinical risk scores, making it a viable alternative for predicting the functional outcomes of AIS. This opens the possibility of automating AIS prognosis through a prediction module in the PACS system [40]. Future studies might be needed to assess whether using the

DL imaging biomarker for prediction impacts clinical decision-making.

#### 4.4. Limitations

This study has the following limitations. Firstly, it is a single-center study, and the generalizability of its findings has yet to be fully established. While we evaluated the DL imaging biomarker using a hold-out test set, an independent external dataset would have been more suitable for assessing generalizability. Secondly, we did not delineate infarct lesions using manual or automated segmentation prior to extracting radiomic features. This step could have helped capture specific tissue and lesion properties from the infarcted area. However, since characteristics of the underlying brain may contain relevant information that correlates with functional outcomes of AIS [33], we chose to extract radiomic features from the entire brain parenchyma. Based on these radiomic features, the model showed adequate predictive ability, with AUC values of 0.778 in the validation set and 0.769 in the test set (Supplementary Tables 2 and 3). Additionally, it eliminates the need for lesion segmentation during the prediction stage.

### 5. Conclusions

We developed a DL model that generates DL scores using DWI and ADC images in patients with AIS. The DL score can serve as an imaging biomarker, predicting the functional outcomes of AIS. This lays the foundation for automated outcome prediction in the PACS system. While the proposed DL model may seem complex, advancements in computational power have significantly reduced the training time. By creating a standalone application for making predictions and an interface for importing images from the PACS system, this prediction tool can be implemented in a clinical routine. Furthermore, this imaging biomarker has the potential to enhance the predictive ability of current clinical risk scores, assisting clinicians in improving their functional outcome prediction for AIS patients.

#### CRedit authorship contribution statement

**Tzu-Hsien Yang:** Writing – original draft, Investigation, Data curation, Conceptualization. **Ying-Ying Su:** Writing – original draft, Investigation, Data curation, Conceptualization. **Chia-Ling Tsai:** Writing – original draft, Methodology, Investigation. **Kai-Hsuan Lin:** Writing – original draft, Software, Investigation, Data curation. **Wei-Yang Lin:** Writing – review & editing, Supervision, Methodology, Investigation. **Sheng-Feng Sung:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization.

#### 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.

#### Acknowledgments

The authors thank the help from the Clinical Data Center, Ditmanson Medical Foundation Chia-Yi Christian Hospital, for providing administrative and technical support.

This research was supported by the Ditmanson Medical Foundation Chia-Yi Christian Hospital-National Chung Cheng University Joint Research Program [grant number CYCH-CCU-2022-14] and the Advanced Institute of Manufacturing with High-tech Innovations from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan. The research funder had no role in the design and conduct of the study, interpretation of the data, or decision to submit for publication.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ejrad.2024.111405>.

## References

- [1] GBD 2019 Stroke 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 (2021) 795–820, [https://doi.org/10.1016/s1474-4422\(21\)00252-0](https://doi.org/10.1016/s1474-4422(21)00252-0).
- [2] GBD 2016 Lifetime Risk of Stroke Collaborators, Global, regional, and country-specific lifetime risks of stroke, 1990 and 2016, *New Engl. J. Med.* 379 (2018) 2429–2437, <https://doi.org/10.1056/nejmoa1804492>.
- [3] G. Saposnik, A.K. Guzik, M. Reeves, B. Oviagele, S.C. Johnston, Stroke prognostication using age and NIH stroke scale: SPAN-100, *Neurology* 80 (2013) 21–28, <https://doi.org/10.1212/wnl.0b013e31827b1ace>.
- [4] B.A. Drozdowska, S. Singh, T.J. Quinn, Thinking about the future: a review of prognostic scales used in acute stroke, *Front. Neurol.* 10 (2019) 274, <https://doi.org/10.3389/fneur.2019.00274>.
- [5] I.R. de Ridder, S.A. Dijkland, M. Scheele, H.M. den Hertog, M. Dirks, W. F. Westendorp, P.J. Nederkoorn, D. van de Beek, G.M. Ribbers, E.W. Steyerberg, H. F. Lingsma, D.W. Dippel, Development and validation of the Dutch Stroke Score for predicting disability and functional outcome after ischemic stroke: a tool to support efficient discharge planning, *Eur. Stroke J.* 3 (2017) 165–173, <https://doi.org/10.1177/2396987318754591>.
- [6] K. Matsumoto, Y. Nohara, H. Soejima, T. Yonehara, N. Nakashima, M. Kamouchi, Stroke prognostic scores and data-driven prediction of clinical outcomes after acute ischemic stroke, *Stroke* 51 (2020) 1477–1483, <https://doi.org/10.1161/strokeaha.119.027300>.
- [7] J. Lee, K.M. Park, S. Park, Interpretable machine learning for prediction of clinical outcomes in acute ischemic stroke, *Front. Neurol.* 14 (2023) 1234046, <https://doi.org/10.3389/fneur.2023.1234046>.
- [8] B. Cheng, N.D. Forkert, M. Zavaglia, C.C. Hilgetag, A. Golsari, S. Siemonsen, J. Fiehler, S. Pedraza, J. Puig, T.-H. Cho, J. Alawneh, J.-C. Baron, L. Ostergaard, C. Gerloff, G. Thomalla, Influence of stroke infarct location on functional outcome measured by the modified Rankin scale, *Stroke* 45 (2014) 1695–1702, <https://doi.org/10.1161/strokeaha.114.005152>.
- [9] X. Meng, J. Ji, Infarct volume and outcome of cerebral ischaemia, a systematic review and meta-analysis, *Int. J. Clin. Pr.* 75 (2021) e14773, <https://doi.org/10.1111/ijcp.14773>.
- [10] L. Zhang, J. Wu, R. Yu, R. Xu, J. Yang, Q. Fan, D. Wang, W. Zhang, Non-contrast CT radiomics and machine learning for outcomes prediction of patients with acute ischemic stroke receiving conventional treatment, *Eur. J. Radiol.* 165 (2023) 110959, <https://doi.org/10.1016/j.ejrad.2023.110959>.
- [11] M.S. Jabal, O. Joly, D. Kallmes, G. Harston, A. Rabinstein, T. Huynh, W. Brinjikji, Interpretable machine learning modeling for ischemic stroke outcome prediction, *Front. Neurol.* 13 (2022) 884693, <https://doi.org/10.3389/fneur.2022.884693>.
- [12] Y. Chen, D.J. Tozer, W. Liu, E.J. Peake, H.S. Markus, Prediction of response to thrombolysis in acute stroke using neural network analysis of CT perfusion imaging, *Eur. Stroke J.* 8 (2023) 629–637, <https://doi.org/10.1177/23969873231183206>.
- [13] Y. Li, Y. Liu, Z. Hong, Y. Wang, X. Lu, Combining machine learning with radiomics features in predicting outcomes after mechanical thrombectomy in patients with acute ischemic stroke, *Comput. Methods Prog. Biomed.* 225 (2022) 107093, <https://doi.org/10.1016/j.cmpb.2022.107093>.
- [14] E. Moulton, R. Valabregue, M. Pliotin, G. Marnat, S. Saleme, B. Lapergue, S. Lehericy, F. Clarencon, C. Rosso, Interpretable deep learning for the prognosis of long-term functional outcome post-stroke using acute diffusion weighted imaging, *J. Cereb. Blood Flow Metab.* 43 (2022) 198–209, <https://doi.org/10.1177/0271678x221129230>.
- [15] L. Herzog, L. Kook, J. Hamann, C. Globas, M.R. Heldner, D. Seiffge, K. Antonenko, T. Dobrocky, L. Panos, J. Kaesmacher, U. Fischer, J. Gralla, M. Arnold, R. Wiest, A. R. Luft, B. Sick, S. Wegener, Deep learning versus neurologists: functional outcome prediction in LVO stroke patients undergoing mechanical thrombectomy, *Stroke* 54 (2023) 1761–1769, <https://doi.org/10.1161/strokeaha.123.042496>.
- [16] Y. Liu, Y. Yu, J. Ouyang, B. Jiang, G. Yang, S. Ostmeier, M. Wintermark, P. Michel, D.S. Liebeskind, M.G. Lansberg, G.W. Albers, G. Zaharchuk, Functional outcome prediction in acute ischemic stroke using a fused imaging and clinical deep learning model, *Stroke* 54 (2023) 2316–2327, <https://doi.org/10.1161/strokeaha.123.044072>.
- [17] L. Fast, U. Temuulen, K. Villringer, A. Kufner, H.F. Ali, E. Siebert, S. Huo, S. K. Piper, P.S. Sperber, T. Liman, M. Endres, K. Ritter, Machine learning-based prediction of clinical outcomes after first-ever ischemic stroke, *Front. Neurol.* 14 (2023) 1114360, <https://doi.org/10.3389/fneur.2023.1114360>.
- [18] B. Hotter, S. Hoffmann, L. Ulm, C. Meisel, A. Bustamante, J. Montaner, M. Katan, C. J. Smith, A. Meisel, External validation of five scores to predict stroke-associated pneumonia and the role of selected blood biomarkers, *Stroke* 52 (2021) 325–330, <https://doi.org/10.1161/strokeaha.120.031884>.
- [19] J.M. Rivera-Caravaca, F. Marín, J.A. Vilchez, J. Gálvez, M.A. Esteve-Pastor, V. Vicente, G.Y.H. Lip, V. Roldán, Refining stroke and bleeding prediction in atrial fibrillation by adding consecutive biomarkers to clinical risk scores, *Stroke* 50 (2019) 1372–1379, <https://doi.org/10.1161/strokeaha.118.024305>.
- [20] A. Echle, N.T. Rindtorff, T.J. Brinker, T. Luedde, A.T. Pearson, J.N. Kather, Deep learning in cancer pathology: a new generation of clinical biomarkers, *Brit. J. Cancer.* 124 (2021) 686–696, <https://doi.org/10.1038/s41416-020-01122-x>.
- [21] C. Jiang, C. Qian, Z. Jiang, Y. Teng, R. Lai, Y. Sun, X. Ni, C. Ding, Y. Xu, R. Tian, Robust deep learning-based PET prognostic imaging biomarker for DLBCL patients: a multicenter study, *Eur. J. Nucl. Med. Mol. Imaging* (2023) 1–12, <https://doi.org/10.1007/s00259-023-06405-y>.
- [22] J.M. Ospel, S. Brown, M. Kappelhof, W. van Zwam, T. Jovin, D. Roy, B.C. V. Campbell, P. Mitchell, Y. Roos, F. Guillemin, B. Buck, K. Muir, S. Bracad, P. White, R.M. de Rochemont, M. Goyal, for the HERMES Investigators, Comparing the prognostic impact of age and baseline National Institutes of Health stroke scale in acute stroke due to large vessel occlusion, *Stroke* 52 (2021) 2839–2845, <https://doi.org/10.1161/strokeaha.120.032364>.
- [23] M.J. O'Donnell, J. Fang, C. D'Uva, G. Saposnik, L. Gould, E. McGrath, M.K. Kapral, for the Investigators of the Registry of the Canadian Stroke Network, The PLAN score: a bedside prediction rule for death and severe disability following acute ischemic stroke, *Arch. Intern. Med.* 172 (2012) 1548–1556, <https://doi.org/10.1001/2013.jamainternmed.30>.
- [24] G. Ntaios, M. Faouzi, J. Ferrari, W. Lang, K. Vemmos, P. Michel, An integer-based score to predict functional outcome in acute ischemic stroke, *Neurology* 78 (2012) 1916–1922, <https://doi.org/10.1212/wnl.0b013e318259e221>.
- [25] M.E. Mayerhoefer, A. Materka, G. Langs, I. Häggström, P. Szczypinski, P. Gibbs, G. Cook, Introduction to radiomics, *J. Nucl. Med.* 61 (2020) 222893, <https://doi.org/10.2967/jnumed.118.222893>.
- [26] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778, doi: 10.1109/cvpr.2016.90.
- [27] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, K. Barnard, Attentional feature fusion, in: 2021 IEEE Winter Conf. Appl. Comput. Vis. (WACV), 2021, pp. 3559–3568, doi: 10.1109/wacv48630.2021.00360.
- [28] M.J. Pencina, R.B. D'Agostino, E.W. Steyerberg, Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers, *Stat. Med.* 30 (2011) 11–21, <https://doi.org/10.1002/sim.4085>.
- [29] S.P. Kloska, M. Wintermark, T. Engelhorn, J.B. Fiebach, Acute stroke magnetic resonance imaging: current status and future perspective, *Neuroradiology* 52 (2010) 189–201, <https://doi.org/10.1007/s00234-009-0637-1>.
- [30] B.C.V. Campbell, H.T.H. Tu, S. Christensen, P.M. Desmond, C.R. Levi, C.F. Bladin, N. Hjort, M. Ashkanian, C. Sølling, G.A. Donnan, S.M. Davis, L. Østergaard, M. W. Parsons, Assessing response to stroke thrombolysis: validation of 24-hour multimodal magnetic resonance imaging, *Arch. Neurol.* 69 (2012) 46–50, <https://doi.org/10.1001/archneurol.2011.232>.
- [31] H. Nishi, N. Oishi, A. Ishii, I. Ono, T. Ogura, T. Sunohara, H. Chihara, R. Fukumitsu, M. Okawa, N. Yamana, H. Imamura, N. Sadamas, T. Hatano, I. Nakahara, N. Sakai, S. Miyamoto, Deep learning-derived high-level neuroimaging features predict clinical outcomes for large vessel occlusion, *Stroke* 51 (2020) 1484–1492, <https://doi.org/10.1161/strokeaha.119.028101>.
- [32] L. Jiang, Z. Miao, H. Chen, W. Geng, W. Yong, Y.-C. Chen, H. Zhang, S. Duan, X. Yin, Z. Zhang, Radiomics analysis of diffusion-weighted imaging and long-term unfavorable outcomes risk for acute stroke, *Stroke* 54 (2023) 488–498, <https://doi.org/10.1161/strokeaha.122.040418>.
- [33] M. Bretzner, A.K. Bonkhoff, M.D. Schirmer, S. Hong, A. Dalca, K. Donahue, A.-K. Giese, M.R. Etherton, P.M. Rist, M. Nardin, R.W. Regenhardt, X. Leclerc, R. Lopes, M. Gautherot, C. Wang, O.R. Benavente, J.W. Cole, A. Donati, C. Griessenauer, L. Heitsch, L. Holmegaard, K. Jood, J. Jimenez-Conde, S.J. Kittner, R. Lemmens, C.R. Levi, P.F. McArdle, C.W. McDonough, J.F. Meschia, C.-L. Phuah, A. Rolfs, S. Ropele, J. Rosand, J. Roquer, T. Rundek, R.L. Sacco, R. Schmidt, P. Sharma, A. Slowik, A. Sousa, T.M. Stanne, D. Strbian, T. Tatlisumak, V. Thijs, A. Vagal, J. Wasselius, D. Woo, O. Wu, R. Zand, B.B. Worrall, J. Maguire, A. G. Lindgren, C. Jern, P. Golland, G. Kuchcinski, N.S. Rest, MRI-GENIE and GISCOME Investigators and the International Stroke Genetics Consortium, Radiomics-derived brain age predicts functional outcome after acute ischemic stroke, *Neurology* 100 (2023) e822–e833, <https://doi.org/10.1212/wnl.0000000000201596>.
- [34] G. Quan, R. Ban, J.-L. Ren, Y. Liu, W. Wang, S. Dai, T. Yuan, FLAIR and ADC image-based radiomics features as predictive biomarkers of unfavorable outcome in patients with acute ischemic stroke, *Front. Neurosci.* 15 (2021) 730879, <https://doi.org/10.3389/fnins.2021.730879>.
- [35] C.S. Kidwell, MRI biomarkers in acute ischemic stroke: a conceptual framework and historical analysis, *Stroke* 44 (2018) 570–578, <https://doi.org/10.1161/strokeaha.111.626093>.
- [36] A. Hilbert, L.A. Ramos, H.J.A. van Os, S.D. Olabarriaga, M.L. Tolhuisen, M.J. H. Wermer, R.S. Barros, I. van der Schaaf, D. Dippel, Y.B.W.E.M. Roos, W.H. van Zwam, A.J. Yoo, B.J. Emmer, G.J.L. Nijeholt, A.H. Zwinderman, G.J. Strijkers, C.B. L.M. Majoie, H.A. Marquering, Data-efficient deep learning of radiological image data for outcome prediction after endovascular treatment of patients with acute ischemic stroke, *Comput. Biol. Med.* 115 (2019) 103516, <https://doi.org/10.1016/j.cbiomed.2019.103516>.
- [37] L.A. Ramos, H. van Os, A. Hilbert, S.D. Olabarriaga, A. van der Lugt, Y.B.W.E.M. Roos, W.H. van Zwam, M.A.A. van Walderveen, M. Ernst, A.H. Zwinderman, G. J. Strijkers, C.B.L.M. Majoie, M.J.H. Wermer, H.A. Marquering, Combination of radiological and clinical baseline data for outcome prediction of patients with an acute ischemic stroke, *Front. Neurol.* 13 (2022) 809343, <https://doi.org/10.3389/fneur.2022.809343>.

- [38] S. Sagnier, I. Sibon, The new insights into human brain imaging after stroke, *J. Neurosci. Res.* 100 (2022) 1171–1181, <https://doi.org/10.1002/jnr.24525>.
- [39] S. Bacchi, T. Zerner, L. Oakden-Rayner, T. Kleinig, S. Patel, J. Jannes, Deep learning in the prediction of ischaemic stroke thrombolysis functional outcomes a pilot study, *Acad. Radiol.* 27 (2020) e19–e23, <https://doi.org/10.1016/j.acra.2019.03.015>.
- [40] K. Osorno-Castillo, R.D. Fonnegra, G.M. Díaz, Integration of machine learning models in PACS systems to support diagnostic in radiology services, *Commun. Comput. Inf. Sci.* (2020) 233–244, [https://doi.org/10.1007/978-3-030-61834-6\\_20](https://doi.org/10.1007/978-3-030-61834-6_20).