



# Article **Prognostic Value of Combined Radiomic Features from Follow-Up DWI and T2-FLAIR in Acute Ischemic Stroke**

Alessia Gerbasi <sup>1,\*</sup>, Praneeta Konduri <sup>2,3</sup>, Manon Tolhuisen <sup>2,3</sup>, Fabiano Cavalcante <sup>3</sup>, Leon Rinkel <sup>4</sup>, Manon Kappelhof <sup>3</sup>, Lennard Wolff <sup>5</sup>, Jonathan M. Coutinho <sup>4</sup>, Bart J. Emmer <sup>3</sup>, Vincent Costalat <sup>6</sup>, Caroline Arquizan <sup>7</sup>, Jeannette Hofmeijer <sup>8</sup>, Maarten Uyttenboogaart <sup>9</sup>, Wim van Zwam <sup>10</sup>, Yvo Roos <sup>4</sup>, Silvana Quaglini <sup>1</sup>, Riccardo Bellazzi <sup>1</sup>, Charles Majoie <sup>3</sup> and Henk Marquering <sup>2,3</sup> on behalf of the MR CLEAN-NO IV Trial Investigators (Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in The Netherlands)

- <sup>1</sup> Department of Electrical, Computer and Biomedical Engineering, University of Pavia, 27100 PV Pavia, Italy
- <sup>2</sup> Department of Biomedical Engineering and Physics, Amsterdam UMC Location University of Amsterdam, 1105 AZ Amsterdam, The Netherlands
- <sup>3</sup> Department of Radiology and Nuclear Medicine, Amsterdam UMC Location University of Amsterdam, 1105 AZ Amsterdam, The Netherlands
- <sup>4</sup> Department of Neurology, Amsterdam UMC Location University of Amsterdam, 1105 AZ Amsterdam, The Netherlands
- <sup>5</sup> Department of Radiology and Nuclear Medicine, Erasmus MC University Medical Center, 3015 Rotterdam, The Netherlands
- <sup>6</sup> Department of Neuroradiology, Centre Hospitalier Universitaire de Montpellier, 34400 Montpellier, France
- <sup>7</sup> Department of Neurology, Centre Hospitalier Universitaire de Montpellier, 34400 Montpellier, France
- <sup>8</sup> Department of Neurology, Rijnstate Hospital, 6836 BH Arnhem, The Netherlands
- Department of Neurology and Department of Medical Imaging Center, University Medical Center Groningen, 9713 GZ Groningen, The Netherlands
- <sup>10</sup> Department of Radiology and Nuclear Medicine, Maastricht University Medical Center, 6229 HX Maastricht, The Netherlands
- Correspondence: alessia.gerbasi01@universitadipavia.it

Abstract: The biological pathways involved in lesion formation after an acute ischemic stroke (AIS) are poorly understood. Despite successful reperfusion treatment, up to two thirds of patients with large vessel occlusion remain functionally dependent. Imaging characteristics extracted from DWI and T2-FLAIR follow-up MR sequences could aid in providing a better understanding of the lesion constituents. We built a fully automated pipeline based on a tree ensemble machine learning model to predict poor long-term functional outcome in patients from the MR CLEAN-NO IV trial. Several feature sets were compared, considering only imaging, only clinical, or both types of features. Nested cross-validation with grid search and a feature selection procedure based on SHapley Additive exPlanations (SHAP) was used to train and validate the models. Considering features from both imaging modalities in combination with clinical characteristics led to the best prognostic model (AUC = 0.85, 95%CI [0.81, 0.89]). Moreover, SHAP values showed that imaging features from both sequences have a relevant impact on the final classification, with texture heterogeneity being the most predictive imaging biomarker. This study suggests the prognostic value of both DWI and T2-FLAIR follow-up sequences for AIS patients. If combined with clinical characteristics, they could lead to better understanding of lesion pathophysiology and improved long-term functional outcome prediction.

**Keywords:** acute ischemic stroke; functional independence; imaging bio-markers; machine learning; ensemble tree classifier; explainable AI; SHAP



Citation: Gerbasi, A.; Konduri, P.; Tolhuisen, M.; Cavalcante, F.; Rinkel, L.; Kappelhof, M.; Wolff, L.; Coutinho, J.M.; Emmer, B.J.; Costalat, V.; et al. Prognostic Value of Combined Radiomic Features from Follow-Up DWI and T2-FLAIR in Acute Ischemic Stroke. *J. Cardiovasc. Dev. Dis.* 2022, *9*, 468. https:// doi.org/10.3390/jcdd9120468 9

Academic Editor: Narayanaswamy Venketasubramanian

Received: 18 November 2022 Accepted: 15 December 2022 Published: 19 December 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

# 1. Introduction

Stroke is the second leading cause of death worldwide, with ischemic stroke being the most frequent type [1]. Acute ischemic stroke (AIS) is caused by thrombotic or embolic occlusion of a cerebral artery leading to reduced blood supply to part of the brain and consequent cell death. Treatments aim for the timely restoration of the blood supply to the occluded artery. The standard of care involves administering thrombolytic drugs (intravenous thrombolysis) followed by mechanical retrieval of the clot (endovascular treatment). Currently, the choice of the best therapy is the subject of ongoing research. Moreover, ischemic lesion evolution after stroke is a complex biological process. Notably, a considerable proportion of patients do not experience clinical improvement despite successful recanalization of the occluded artery and reperfusion of the ischemic area [2].

Although several well-known risk factors, such as lifestyle, common comorbidities, and aging, have been already identified as related to stroke onset and severity, several challenges remain open in this field, including the prediction of long-term prognosis and the identification of biomarkers for common treatment complications, to name only a few.

Neuroimaging has significantly improved AIS management, with magnetic resonance imaging (MRI) and computed tomography (CT) being the preferred imaging modalities. The use of MRI is particularly associated with a significant decrease in the rates of complications in AIS patients at the expense of only marginally increased length and cost of hospitalization [3]. Moreover, MRI is the most sensitive noninvasive neuroimaging technique for assessing brain changes after stroke as well as in predicting recovery, especially in the subacute stage [4]. Accurate functional outcome prediction after treatment is needed in order to provide patients with a prognosis, plan rehabilitation programs, and provide research avenues targeting secondary treatments [5,6].

Previous studies exploring the prognostic value of MRI features have identified the size of the area with diffusion restriction and the apparent diffusion coefficient (ADC) as being associated with clinical outcome. Furthermore, infarct volume measurements (in the subacute phase between 14 h and 48 h) are predictors of long-term functional outcome, measured in terms of 3-month modified Rankin Scale (mRS) [7–9]. A texture analysis based on ADC maps and T2-FLAIR images highlighted an association between image texture features and functional outcome [10].

It is not always easy to fully explain the reasons for poor functional outcome and to timely predict its occurrence in AIS patients. In this context, the predictive power of machine learning (ML) and deep learning (DL) models and their ability to extract hidden patterns from neuroimages have been explored in several studies [11,12]. Most of the published works focused on pre-treatment imaging features extracted from a single imaging modality. However, a tool able to support the interpretation and prognostic value assessment of follow-up neuroimaging sequences in the subacute stage would be useful in clinical practice. There could be early detectable signs for unfavorable long-term clinical outcome among follow-up imaging characteristics that could help to gain new insights into the lesion evolution mechanism and its pathophysiology. MRI follow-up protocols usually include DWI and T2-FLAIR sequences to assess ischemic lesion status after treatment. DWI is considered the gold standard modality for evaluating the status of the lesion in the subacute stage thanks to its ability to quantify the movement of water molecules [13]. During ischemia, intracellular water accumulates in the injured cells as a consequence of changes in the osmotic gradients caused by the malfunctioning of the sodium-potassium pump. This results in a bright signal in MR diffusion imaging, allowing the detection of infarction and cytotoxic edema. On the other hand, T2-FLAIR is a structural imaging sequence in which T2 prolongation is commonly observed hours after stroke. T2 prolongation is mainly related to increased water content in the ischemic tissue, and usually represents vasogenic edema [14]. Although several prognostic models have been proposed in the literature [15,16], lack of reproducibility and interpretability of the prediction results is often the most critical aspect. Moreover, most of the related works are single-center or single-country studies, and thus present a significant risk of bias. Finally, none of the previous studies have explored the prognostic value of combined imaging features extracted from these two follow-up MR sequences; yet, these may actually capture different aspects of stroke lesion status after treatment, and multimodal predictive modeling approaches could lead to more accurate and reliable prognosis prediction from a clinical perspective.

The aim of this study is to evaluate the prognostic value of the combination of two follow-up MR sequences. We hypothesize that the prediction of stroke functional outcome benefits from combined follow-up DWI and T2-FLAIR imaging features in terms of 90-day mRS. We aim to build a complete prognostic model considering both follow-up imaging features extracted from the two sequences and routinely available clinical characteristics, thereby exploiting all patient data available up to 24 h. Finally, we are interested in exploring the contribution of each feature to the prediction of the prognostic model and studying the clinical interpretation of the identified imaging predictors.

#### 2. Materials and Methods

The complete study workflow is schematized in Figure 1. We propose an image analysis pipeline, starting with the preprocessing of the MR scans and the infarct lesion segmentation in DWI and T2-FLAIR sequences. Radiomic features are subsequently extracted from the segmented lesions and combined with clinical characteristics in order to train and validate an ensemble tree-based ML classifier for outcome prediction. Finally, the models' interpretability is evaluated using SHapley Additive exPlanations (SHAP) [17].



**Figure 1.** Schematic representation of the study workflow. Follow-up DWI and T2-FLAIR scans are preprocessed and radiomic features are extracted from the segmented lesions. Five different sets of features are then used to train an ensemble tree-based ML classifier to predict binary 90-day mRS (0–2 vs. 3–6). Each set respectively includes: (1) only DWI radiomic features (*DWI set*), (2) only T2-FLAIR radiomic features (*FLAIR set*), (3) both DWI and T2-FLAIR radiomic features (*imaging set*), (4) only clinical characteristics (*clinical set*), and (5) a combination of all the radiomic features and the clinical characteristics available up to 24 h (*combination set*). Finally, the models' interpretability is evaluated using SHapley Additive exPlanations (SHAP).

### 2.1. Data Collection

The patients included in this study were enrolled in the MR CLEAN-NO IV [18] trial, which is a multicenter randomized clinical trial including 539 patients  $\geq$  18 years of age with a pre-stroke mRS < 3 with an intracranial large vessel occlusion (LVO) confirmed on CT/MR angiography within 4.5 h of symptom onset eligible for both intravenous thrombolysis (IVT) and endovascular treatment (EVT). The patients were randomized to receive IVT followed by EVT or EVT only. The follow-up imaging protocol included NCCT or MRI at 24 h for all patients. A detailed description of the inclusion and exclusion criteria is provided in the study protocol [19]. The trial was carried out in 20 hospitals in

the Netherlands, Belgium, and France. Therefore, the slice thickness of the scans varied (from 3 to 6 mm), and the field strength used for the acquisition was 1.5 or 3 Tesla.

Baseline clinical features, 24-h NIH Stroke Scale score, and 90-day mRS were available for all the patients included in the study.

Figure 2 shows a flow-chart of the patient inclusion process. We included patients with high-quality follow-up DWI and T2-FLAIR scans, resulting in 164 patients from MR CLEAN-NO IV trial. Scans presenting severe artifacts were excluded from this study.



Figure 2. Flow chart of the patient inclusion process.

#### 2.2. Image Preprocessing

The main image preprocessing steps are summarized in Figure 3. All the scans were skull-stripped and registered to the standard MNI-space via rigid transformations using the open source SPM12 toolbox [20], obtaining an isotropic voxel dimension of 1 mm. Intensities were normalized using the white stripe normalization method [21]. Finally, the lesion was segmented in both imaging sequences.



**Figure 3.** Image preprocessing pipeline. The same steps are repeated for each of the input MR sequences, obtaining the infarct lesion segmentation as the final output from both imaging sequences.

Infarcts in DWI scans were automatically segmented using a deep learning software developed in-house and then manually checked and corrected by two expert neuroradiologists (with more than twenty and five years of experience, respectively) using ITK-SNAP software [22]. Infarcts in T2-FLAIR sequences were automatically segmented using DeepNeuroSeg [23], a deep fully convolutional network trained to detect white matter hyperintensities (WMH) from FLAIR that ranked first in WMH Segmentation Challenge at MICCAI 2017. All resulting segmentations were manually checked and corrected using ITK-SNAP software. Figure 4 shows an example of T2-FLAIR and DWI lesion segmentation from the same patient. For each modality, a few slices sampled from the preprocessed scan are reported in the first line. Lesions appear hyperintense on both imaging modalities, and the segmented area after the preprocessing steps (red) is displayed as overlapping the original slices in the second row.



(a) DWI



(b) T2-FLAIR

**Figure 4.** Example of infarct lesion segmentation on a selected slices sampled from (**a**) DWI and (**b**) T2-FLAIR preprocessed volumes. For each modality, original slices are displayed in the first row and lesion segmentation (red) is shown as overlapping the correspondent slices in the second row.

## 2.3. Features

Radiomic features were extracted from the segmented lesion, which is defined as the region of interest (ROI) in the following using the open-source PyRadiomics [24] Python package. Although deep neural networks have been proven to be powerful feature extractors in several computer vision applications, and are able to capture hidden patterns in images, we decided to avoid this strategy in our study due to the limited sample size and consequent high risk of overfitting. The extracted radiomic features (n = 107) extensively describe the lesion, capturing different aspects, and can be subdivided in the following classes: first-order statistics, size/shape based, and high-order textural characteristics. The first-order statistics describe the distribution of voxel intensities within the ROI (e.g., energy, entropy); the size- and shape-based features include both 2D and 3D measures, among which is the follow-up lesion volume, which is often used in literature as a prognostic marker; finally, the textural features include measures derived from correlation matrices describing spatial patterns of voxel intensities. There are five main textural matrices: (1) the Gray Level Co-occurrence Matrix (GLCM), which describes the second-order joint probability function of an image region constrained by the mask; (2) the Gray Level Size Zone Matrix (GLSZM), which quantifies gray level zones in an image, defined as a the number of connected voxels that share the same gray level intensity; (3) the Gray Level Run

Length Matrix (GLRLM), which quantifies gray level runs, defined as the length in number of pixels, of consecutive pixels that have the same gray level value; (4) the Neighbouring Gray Tone Difference Matrix (NGTDM), which quantifies the difference between a gray value and the average gray value of its neighbors within a defined distance; and (5) the Gray Level Dependence Matrix (GLDM), which quantifies gray level dependencies in an image, defined as a the number of connected voxels within a defined distance that are dependent on the center voxel.

We considered five different sets of features to predict binary 90-day mRS (0–2 vs. 3–6) as the model outcome, a clinical definition for favorable long-term functional independence. Each set respectively includes: (1) only DWI radiomic features (named the *DWI set*, n = 107); (2) only T2-FLAIR radiomic features (named the *FLAIR set*, n = 107); (3) both DWI and T2-FLAIR radiomic features (named the *imaging set*, n = 214); (4) only clinical characteristics (named the *clinical set*, n = 23); and (5) a combination of all the radiomic features and the clinical characteristics available up to 24 h (named the *combination set*, n = 237).

Clinical characteristics were included in the final models, as it is well known that a number of these are strongly associated with functional outcome. Therefore, we decided to further evaluate the prognostic value of imaging features on top of routinely acquired clinical characteristics in order to simulate a complete model exploiting all the patient data available up to 24 h. Clinical characteristics included: clinical history data (e.g., age, gender), baseline clinical measurements and imaging derived scores routinely acquired in the acute phase (e.g., glucose, ASPECTS), treatment, and post-treatment characteristics (e.g., type of treatment, 24-h NIH Stroke Scale). A complete list of the clinical characteristics considered in this study is provided in the Supplementary Materials.

#### 2.4. Image Analysis Pipeline and Experimental Setup

We adopted the Extreme Gradient Boosting (XGB) model as a classifier, implemented in the XGBoost Python library [25]. XGB is a high-performing ensemble tree-based algorithm designed to deal with high dimensional and mixed input data types. One of the strengths of this algorithm is its capability to automatically handle collinear features and missing data; therefore, we preferred to avoid any imputing strategy for the missing clinical values.

We implemented a feature selection strategy based on SHAP values [17]. Lundberg et al. [26] showed that several common methods for estimating feature importance are inconsistent, as they can lower a feature's assigned importance when the true impact of that feature actually increases. Therefore, they developed TreeSHAP, a consistent individualized feature attribution algorithm for tree ensembles based on SHAP values. It has been shown that this method results in reduced computational complexity depending on the maximum depth of the trees instead of the number of possible feature combinations, and it has demonstrated obtained results in better agreement with human intuition.

Because the analyzed sample size is relatively small, nested stratified *k*-fold cross validation (CV) was used to train and test the model in order to avoid biases and overly-optimistic results. The schema of the experimental setup is shown in Figure 5, where the outer and inner loops are schematically represented on the left and right, respectively. In our pipeline, *k* was set to 5, meaning that that 20% of the dataset is used as a test set at each split. In the inner CV, a grid-search strategy was implemented to tune the model's parameters (*learning\_rate, max\_depth, min\_child\_weight, colsample\_bytree, colsample\_bynode, subsample, alpha*) in order to maximise validation accuracy. A detailed description of the parameters is available in the official documentation [27]. For each feature, mean absolute SHAP values were computed on each validation set of the inner loop. The average SHAP values at the end of inner CV were used to obtain the final feature importance ranking. All the features with a mean absolute SHAP value equal to 0 where automatically discarded. For each round of the outer CV, the final model with the selected parameters and features was retrained on the whole outer training set and tested on the unseen test set to assess the model's generalization capabilities.

The pipeline was executed for each of the five feature sets previously defined. For each classifier, final average performance on the test sets was evaluated in terms of classification accuracy, precision, sensitivity, specificity, and area under the ROC curve (AUC). We compared the models pairwise based on their AUC using DeLong's test, with the highest AUC as a reference.



**Figure 5.** Schema of experimental setup, showing a schematic representation of the nested cross validation in the outer loop on the left and the inner loop on the right. Each row on the left represents the whole dataset, which is divided in five different folds; one fold (blue) is used as an independent test set, while the remaining four folds form the training set. Repeating this split five times ensures that each single fold is considered as a test set once. In the inner loop on the right, the training set is subsequently split in five folds again in order to obtain a validation set (green) used to estimate the model's performance during the grid search and feature selection procedure. The best model with the chosen parameters and features is then retrained on the outer training set (red) and tested on the outer test set (blue). SHAP values for the final model interpretability are computed on the test sets.

#### 2.5. Explainability

The explainability of AI models is of the utmost importance in the medical field, especially when dealing with challenging clinical tasks such as prognosis prediction where a black-box approach would be useless in clinical practice. The final aim in this case is not only to build a reliable model, it is to better understand the impact of the identified predictors that could possibly help to gain new insights into lesion pathophysiology and help clinicians to interpret and trust the model's outcome.

For the best model, SHAP values were computed on the test sets to better understand the role of each feature in predicting the desired outcome. We used SHAP summary plots to show the most important features with the largest impact on outcome prediction for the best experiment. This is an easy to read beeswarm plot colored by feature values, with the features ordered based on their impact on the model's output. Positive SHAP values are associated with unfavorable outcomes, while negative values are associated with good clinical outcomes. In order to visually inspect the meaning of the extracted radiomic characteristics, we ordered our cohort of patients based on the relevant features shown by the resulting SHAP summary plot and analyzed four different case studies.

## 3. Results

A complete list of the clinical characteristics, including clinical history and measurements, baseline imaging scores, and treatment and post-treatment characteristics of the patients included in this study and the original MR CLEAN-NO IV population, is provided in the Supplementary Materials.

The prediction accuracy measures computed on the test sets for each collection of features are reported in Table 1. Table 2 provides the results of DeLong's test. The model including all imaging features (*imaging set*) shows slightly higher performance compared to the models using features from a single modality. The average AUC achieved with the *imaging set* is 71% (+2% and +3% compared to the average AUC of *DWI* and *FLAIR set*, respectively). The model trained on the *clinical set* shows a higher predictive power

compared to models using imaging features alone (+12% compared to the average AUC of the *imaging set*). Finally, the *combination set* led to the best prognostic model, outperforming all the other models in terms of the considered metrics, with an average AUC of 85% (Figure 6).

**Table 1.** Average metrics (95% CI) computed on the test sets for all the tested feature sets (*DWI set*: only DWI radiomic features, *FLAIR set*: only T2-FLAIR radiomic features, *imaging set*: both DWI and T2-FLAIR radiomic features, *clinical set*: only clinical characteristics, *combination set*: combination of all the radiomic features and the clinical characteristics). For each feature set, Accuracy (Acc), Sensitivity (Sens), Specificity (Spec), Precision (Prec), and Area under the ROC curve (AUC) are reported.

Feature Set	Acc	Sens	Spec	Prec	AUC
DWI	0.66 [0.56, 0.76]	0.58 [0.38, 0.78]	0.72 [0.56, 0.88]	0.62 [0.48, 0.76]	0.69 [0.55, 0.82]
FLAIR	0.64 [0.58, 0.70]	0.56 [0.46, 0.66]	0.71 [0.62, 0.80]	0.58 [0.50, 0.66]	0.68 [0.59, 0.77]
Imaging	0.67 [0.60, 0.74]	0.56 [0.47, 0.65]	0.76 [0.66, 0.86]	0.65 [0.55, 0.75]	0.71 [0.63, 0.79]
Clinical	0.75 [0.70, 0.80]	0.71 [0.58, 0.84]	0.79 [0.74, 0.84]	0.72 [0.67, 0.77]	0.83 [0.78, 0.88]
Combination	0.79 [0.72, 0.86]	0.71 [0.62, 0.80]	0.84 [0.74, 0.94]	0.78 [0.68, 0.88]	0.85 [0.81, 0.89]

**Table 2.** DeLong's test *p*-values for pairwise AUC comparisons ( $\alpha = 0.05$ ), with the highest AUC as a reference. Statistically significant results are provided in bold.

Feature Set	AUC	<i>p</i> -Value
Combination—DWI	0.85–0.69	<0.001
Combination—FLAIR	0.85-0.68	<0.001
Combination—Imaging	0.85-0.71	<0.001
Combination—Clinical	0.85–0.83	0.33



**Figure 6.** Average ROC curve (blue) obtained from the outer cross-validation of the best-performing model (*combination set*).

Figure 7 shows the SHAP summary plot of the most important features with the largest impact on outcome prediction for the best experiment (*combination set*). The features with the largest impact include both clinical characteristics and radiomic features. It can be seen that both DWI and T2-FLAIR radiomic features are ranked among the most relevant predictors, with texture features the most frequent ones. The only exceptions are *Skewness* measured on DWI, which is a first-order feature and an indicator of asymmetry in gray

level distribution, and *Sphericity* measured on T2-FLAIR, which is a shape-related feature measuring the roundness of the shape of the lesion region. All the other image predictors are texture features, and despite the different textural matrices being able to capture slightly different aspects of gray values distributions, they are essentially describing the heterogeneity of the examined region (a detailed description of each feature is reported in the official documentation [28]). The beeswarm plot shows that more heterogeneous lesions (↑ Dependence non uniformity, GLDM–DWI; ↓ Strength, NGTDM–FLAIR; ↑ Short run low level emphasis, GLRLM–FLAIR; ↑ Max correlation coefficient, GLCM–FLAIR; ↑ Gray level non uniformity, GLSZM–FLAIR) are more associated with unfavorable clinical outcomes.



SHAP value (impact on model output)

**Figure 7.** SHAP value beeswarm plot of the most relevant predictors on the test sets. Features are sorted according to their impact on model output. Positive and negative SHAP values are associated with unfavorable and favorable clinical outcomes, respectively. The colors represent the original feature values, with higher values in pink and lower values in blue. For each clinical feature (CLIN), the time of measurement is reported (BL = baseline, FU = follow-up at 24 h). For each radiomic feature, the imaging modality (FLAIR or DWI) and the type of feature are provided in brackets (FO = first order, SHAPE = 2D or 3D shape, NGTDM/GLRLM/GLCM/GLSZM = texture).

Figure 8 shows four different study cases (L1-4) demonstrating the meaning of texture heterogeneity for visual inspection. For each lesion, the axial view from both DWI and T2-FLAIR modality is reported. The lesion appears hyperintense on both imaging sequences. L1 and L2 are two examples of lesions from patients experiencing poor long-term functional outcome (90-day mRS of 6 and 5 respectively). L1 is a very severe case; the extension of the lesion is quite large, involving both GM and WM and presenting PH1 hemorrhage. L2 is less extensive compared to L1, although it too involves both GM and WM. Patients with worse outcomes present more heterogeneous lesions, often involving both GM and WM. In the presence of hemorrhage, heterogeneity in terms of radiomic features is even higher (e.g., texture heterogeneity measured in terms of Dependence non uniformity normalized—GLDM on DWI was 0.23 and 0.13 in L1 and L2, respectively). On the other hand, in patients that achieved functional independence (e.g., L3 and L4 presented a 90-day mRS of 2 and 0, respectively), the lesions were less heterogeneous (texture heterogeneity in terms of Dependence non uniformity normalized—GLDM on DWI of 0.08 and 0.04, respectively), and mostly involved GM areas.



(**g**) L4—DWI

(h) L4—FLAIR

**Figure 8.** Axial views of DWI and T2-FLAIR scans from four different lesions (L1–L4). The lesions look hyperintense on both imaging sequences. L1 and L2 are two examples of patients with poor long-term functional outcome (90-day mRS of 6 and 5 respectively). L1 is quite extensive and presents PH1 hemorrage. Both lesions include WM and GM areas of the brain (texture heterogeneity measured in terms of Dependence non uniformity normalized—GLDM on DWI of 0.23 and 0.13, respectively). L3 and L4 instead are two examples of patients with good long-term functional outcome (90-day mRS of 2 and 0, respectively). Both lesions involve GM only (texture heterogeneity measured in terms of Dependence non uniformity normalized – GLDM on DWI of 0.08 and 0.04, respectively).

# 4. Discussion

In this study, we showed the predictive value of radiomic features extracted from routinely acquired follow-up MR sequences (DWI and T2-FLAIR) for long-term functional outcome prediction in AIS patients. In the comparison of prediction models trained on five different feature sets, respectively including only radiomic features extracted by one follow-up MR sequence (DWI or T2-FLAIR), all the available imaging sequences (DWI and T2-FLAIR), only clinical characteristics, and finally a combination of all the radiomic features and the available clinical characteristics, we showed that the highest accuracy is found with the model trained on this last feature set. A completely data-driven ranking of the best model's predictors based on SHAP values showed the relevance of radiomic features from both imaging modalities and confirmed, in line with previous clinical studies, the high predictive value of 24-h NIH Stroke Scale, and age [29,30].

Among the most relevant imaging characteristics is the heterogeneity of the lesion measured in terms of different texture features on both MR sequences, which has been shown to be more related to worse long-term functional outcome. This finding is interesting from a clinical perspective because heterogeneous lesions could indicate edema formation in the subacute stage or hemorrhage. In fact, focal cerebral ischemia and post-ischemic reperfusion can cause cerebral capillary dysfunction, resulting in edema formation and hemorrhagic conversion [31]. Because edema strongly influences patient prognosis, predicting its occurrence in the early stages is vital to providing timely intervention. However, there is a lack of early bio-markers, and the underlying physiological mechanisms of cerebral edema and hemorrhagic transformations are still being searched [32]. The work of Jiang et al. [33] supports our observation, showing how radiomic features from infarction and CSF (measured on DWI and T2-FLAIR respectively) may offer effective imaging biomarkers for predicting edema. In addition, Zhai et al. [34] showed DWI-based texture analysis with good predictive validity for hemorrhagic transformation in patients with acute massive cerebral infarction.

Another interesting aspect that can be captured by texture features is the involvement of white matter (WM) and gray matter (GM) in the ischemic lesion. Both GM and WM can be affected by ischemic stroke, as they need a constant supply of oxygen and glucose. However, the collateral circulation in WM is reduced compared to GM and the reduced blood supply makes WM highly vulnerable to ischemia, leading to more rapid and severe injuries [35]. Damage in WM integrity has been shown to be related to cognitive deficits in stroke patients and worse post-stroke functional outcomes [36]. More heterogeneous lesions are more likely to involve both GM and WM, indicating more severe injuries with fewer expectations of full functional recovery. Nonetheless, the lesions in patients with a favorable functional outcome were often smaller than those that did not achieve functional independence. Very large lesions are more likely to involve both GM and WM, resulting in a more heterogeneous areas, while very small lesions are often localized in GM or WM only. Therefore, heterogeneity indirectly takes into account follow-up lesion volume, which is known to be a useful prognostic factor [37]. Despite this, when the size of the lesion is comparable (e.g., L2, L3 in Figure 8), the involvement of both GM and WM still results in higher values of heterogeneity, indicating a higher risk of poor long-term functional outcome.

At present, most of the available radiomic evidence regarding stroke is derived from single-center studies, often leading to unstable and unreproducible association of radiomic features with clinical events due to selection bias [16]. This multi-center study suggests that functional outcome prediction using a combination of imaging and clinical characteristics leads to accurate results taking into account all the parameters routinely available up to 24 h. Moreover, we show that lesion heterogeneity measured from DWI and T2-FLAIR in terms of radiomic features can serve as a biomarker for poor clinical outcome. An accurate prognosis in the subacute stage is highly relevant for patients who ultimately have a poor outcome, as it enables early planning of care adapted to their needs [38]. Radiomics potential in stroke research has not been fully exploited yet, mostly due to the difficulty of collecting

multi-center and large-scale imaging data. Nonetheless, the identification of stable imaging biomarkers would help to better understand the complex pathophysiological cascade involved in lesion formation and evolution after an AIS, optimize secondary prevention strategies, and facilitate the development of personalized precision medicine in post-stroke patients [16].

This study has a few limitations. Our results are not directly generalizable to patients not compliant to the MR CLEAN-NO IV inclusion criteria. Therefore, future studies assessing the impact of the identified predictors for stroke patients with minor vessel occlusions or posterior circulation occlusions are needed. Moreover, we showed that 24-h radiomic features, if used on top of clinical characteristics, have the potential to improve long-term prognosis prediction. However, in our study the differences are small. This is most likely due to the relatively small size of the analyzed sample. Future works including a larger population could probably lead to increased statistical significance and model performance. Furthermore, with a larger sample size it would be interesting to try endto-end DL approaches comparing the predictive power and interpretability of radiomic and automatically extracted features. Research in this field is in its early stages; studies on larger populations are needed to further assess the scalability and robustness of the proposed method and the identified biomarkers, which is essential before considering its use in the clinical practice. Finally, we only considered mRS score as long-term functional outcome, as it is the most used indicator in clinical practice; however different scores better reflecting the psychological and emotional status of the patient could be used as endpoints, possibly leading to a better understanding of the disease mechanisms.

## 5. Conclusions

In this study, we showed that radiomic imaging features describing infarct lesion on follow-up MR sequences (DWI and T2-FLAIR) are predictive of poor long-term functional outcome in AIS patients, with features describing texture heterogeneity being the most predictive imaging biomarkers. In addition, we proved that using imaging features on top of routinely available clinical characteristics, and thus exploiting all patient data available up to 24 h, leads to the most accurate prognostic model. Further research in this direction may help in better targeting secondary treatments to mitigate lesion evolution in the subacute phase and thereby improving long-term functional outcome, hopefully reducing the burden of disease.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jcdd9120468/s1, Table S1: Comparison of baseline clinical/imaging characteristics, treatment and post-treatment clinical characteristics between patients with favorable (90-day mRS  $\leq$  2) and unfavorable (90-day mRS > 2) long-term clinical outcome; Table S2: Comparison of baseline clinical/imaging characteristics, treatment and post-treatment and post-treatment and post-treatment and post-treatment clinical outcome; Table S2: Comparison of baseline clinical/imaging characteristics, treatment and post-treatment clinical characteristics between included and excluded patients from MR CLEAN-NO IV trial.

Author Contributions: Conceptualization, A.G., P.K. and H.M.; methodology, A.G.; software, A.G.; validation, A.G., P.K. and H.M.; formal analysis, A.G.; investigation, A.G.; resources, M.T., F.C., L.R., M.K., L.W., J.M.C., B.J.E., V.C., C.A., J.H., M.U., W.v.Z., Y.R., C.M. and H.M.; data curation, A.G., M.T. and C.M.; writing—original draft preparation, A.G.; writing—review and editing, P.K., M.T., F.C., L.R., M.K., L.W., J.M.C., B.J.E., V.C., C.A., J.H., M.U., W.v.Z., Y.R., S.Q., R.B., C.M. and H.M.; visualization, A.G.; supervision, H.M., S.Q. and R.B.; project administration, A.G.; funding acquisition, B.J.E., W.v.Z. and C.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** MR CLEAN-NO IV was funded through the CONTRAST consortium, which acknowledges support from the Netherlands Cardiovascular Research Initiative, an initiative of the Dutch Heart Foundation (CVON2015-01: CONTRAST), and from the Brain Foundation Netherlands (HA2015.01.06). The collaboration project is additionally financed by the Ministry of Economic Affairs by means of the PPP Allowance made available by the Top Sector Life Sciences and Health to stimulate public-private partnerships (LSHM17016). This work was funded in part through unrestricted funding by Stryker, Medtronic, and Cerenovus. The funding sources were not involved in study design, monitoring, data collection, statistical analyses, interpretation of results, or manuscript writing. This sub-study was partially funded by INSIST (www.insist-h2020.eu, accessed on 12 July 2022), a European Union Horizon 2020 research and innovation program (Grant Agreement Number: 777072).

**Institutional Review Board Statement:** This study was conducted in accordance with the Declaration of Helsinki. MR CLEAN NO-IV received ethical approval from the Central Medical Ethics Committee and the Research Board of the Erasmus MC University Medical Center, Rotterdam, the Netherlands, approval code MEC-2017-368 on 19 October 2017.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Trial data can be made available on reasonable request via mrclean@ erasmusmc.nl.

**Acknowledgments:** We thank J.W. Hoving and M.S. Koopman (Department of Radiology and Nuclear Medicine, Amsterdam UMC—location AMC) for their contributions in verifying the 24-h DWI lesions.

**Conflicts of Interest:** The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. Konduri was funded by INSIST (www.insist-h2020.eu, accessed on 12 July 2022): a European Union's Horizon 2020 research and innovation programme (Grant Agreement Number: 777072) and the RadPath AI project (25158). W.v.Zwam reports that Maastricht University Medical Center received compensation from Stryker® and Cerenovus® for consultations by W.v.Z.; Y.Roos is a shareholder of Nico.lab International Ltd.; J.Coutinho reports that Amsterdam UMC received research support from Medtronic; R.Bellazzi is co-founder and shareholder of two spin-offs of the University of Pavia (Engenome s.r.l. and Biomeris s.r.l.) that operate in the field of data management, artificial intelligence and bioinformatics; C.Majoie reports grants from CVON/the Dutch Heart Foundation, the TWIN Foundation, the European Commission, the Dutch Health Evaluation program, and Stryker outside the submitted work (paid to the institution) and is a shareholder of Nico.lab International Ltd.; H.Marquering is a co-founder and shareholder of Nico.lab International Ltd.; H.Marquering interests.

## References

- 1. Wittenauer, R.; Smith, L. Background paper 6.6 ischaemic and haemorrhagic stroke. In *Priority Medicines for Europe and the World. A Public Health Approach to Innovation*; World Health Organization: Geneva, Switzerland, 2012.
- Rabinstein, A.A.; Albers, G.W.; Brinjikji, W.; Koch, S. Factors that may contribute to poor outcome despite good reperfusion after acute endovascular stroke therapy. *Int. J. Stroke* 2019, 14, 23–31. [CrossRef] [PubMed]
- 3. Lee, H.; Yang, Y.; Liu, B.; Castro, S.A.; Shi, T. Patients With Acute Ischemic Stroke Who Receive Brain Magnetic Resonance Imaging Demonstrate Favorable In-Hospital Outcomes. *J. Am. Heart Assoc.* **2020**, *9*, e016987. [CrossRef] [PubMed]
- DiBella, E.; Sharma, A.; Richards, L.; Prabhakaran, V.; Majersik, J.; Hashemizadeh Kolowri, S. Beyond diffusion tensor MRI methods for improved characterization of the brain after ischemic stroke: A Review. *Am. J. Neuroradiol.* 2022, 43, 661–669. [CrossRef] [PubMed]
- 5. Kwah, L.K.; Herbert, R.D. Prediction of walking and arm recovery after stroke: a critical review. *Brain Sci.* **2016**, *6*, 53. [CrossRef] [PubMed]
- 6. MEDICA, E.M. Factors enhancing Activities of Daily Living after stroke in specialized rehabilitation. An observational multicenter study within the Sunnaas International Network. *Eur. J. Phys. Rehabil. Med.* **2017**, *53*, 725–734.
- 7. Thijs, V.N.; Lansberg, M.G.; Beaulieu, C.; Marks, M.P.; Moseley, M.E.; Albers, G.W. Is early ischemic lesion volume on diffusionweighted imaging an independent predictor of stroke outcome? A multivariable analysis. *Stroke* 2000, *31*, 2597–2602. [CrossRef]
- Zaidi, S.F.; Aghaebrahim, A.; Urra, X.; Jumaa, M.A.; Jankowitz, B.; Hammer, M.; Nogueira, R.; Horowitz, M.; Reddy, V.; Jovin, T.G. Final infarct volume is a stronger predictor of outcome than recanalization in patients with proximal middle cerebral artery occlusion treated with endovascular therapy. *Stroke* 2012, 43, 3238–3244. [CrossRef]
- 9. Yoo, A.J.; Chaudhry, Z.A.; Nogueira, R.G.; Lev, M.H.; Schaefer, P.W.; Schwamm, L.H.; Hirsch, J.A.; González, R.G. Infarct volume is a pivotal biomarker after intra-arterial stroke therapy. *Stroke* 2012, *43*, 1323–1330. [CrossRef]
- 10. Wang, H.; Lin, J.; Zheng, L.; Zhao, J.; Song, B.; Dai, Y. Texture analysis based on ADC maps and T2-FLAIR images for the assessment of the severity and prognosis of ischaemic stroke. *Clin. Imaging* **2020**, *67*, 152–159. [CrossRef]
- 11. Hilbert, A.; Ramos, L.A.; van Os, H.J.; Olabarriaga, S.D.; Tolhuisen, M.L.; Wermer, M.J.; Barros, R.S.; van der Schaaf, I.; Dippel, D.; Roos, Y.; et al. Data-efficient deep learning of radiological image data for outcome prediction after endovascular treatment of patients with acute ischemic stroke. *Comput. Biol. Med.* **2019**, *115*, 103516. [CrossRef]

- Nishi, H.; Oishi, N.; Ishii, A.; Ono, I.; Ogura, T.; Sunohara, T.; Chihara, H.; Fukumitsu, R.; Okawa, M.; Yamana, N.; et al. Deep Learning–Derived High-Level Neuroimaging Features Predict Clinical Outcomes for Large Vessel Occlusion. *Stroke* 2020, 51, 1484–1492. [CrossRef] [PubMed]
- 13. Birenbaum, D.; Bancroft, L.W.; Felsberg, G.J. Imaging in acute stroke. West. J. Emerg. Med. 2011, 12, 67.
- 14. Ayata, C.; Ropper, A.H. Ischaemic brain oedema. J. Clin. Neurosci. 2002, 9, 113–124. [CrossRef] [PubMed]
- 15. Shlobin, N.A.; Baig, A.A.; Waqas, M.; Patel, T.R.; Dossani, R.H.; Wilson, M.; Cappuzzo, J.M.; Siddiqui, A.H.; Tutino, V.M.; Levy, E.I. Artificial Intelligence for Large Vessel Occlusion Stroke: A Systematic Review. *World Neurosurg.* **2021**, *159*, 207–220.e1. [CrossRef]
- 16. Chen, Q.; Xia, T.; Zhang, M.; Xia, N.; Liu, J.; Yang, Y. Radiomics in stroke neuroimaging: Techniques, applications, and challenges. *Aging Dis.* **2021**, *12*, 143. [CrossRef] [PubMed]
- 17. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 2017, 30, 4768–4777.
- LeCouffe, N.E.; Kappelhof, M.; Treurniet, K.M.; Rinkel, L.A.; Bruggeman, A.E.; Berkhemer, O.A.; Wolff, L.; van Voorst, H.; Tolhuisen, M.L.; Dippel, D.W.; et al. A randomized trial of intravenous alteplase before endovascular treatment for stroke. *N. Engl. J. Med.* 2021, 385, 1833–1844. [CrossRef]
- Treurniet, K.M.; LeCouffe, N.E.; Kappelhof, M.; Emmer, B.J.; van Es, A.C.; Boiten, J.; Lycklama, G.J.; Keizer, K.; Lonneke, S.; Lingsma, H.F.; et al. MR CLEAN-NO IV: intravenous treatment followed by endovascular treatment versus direct endovascular treatment for acute ischemic stroke caused by a proximal intracranial occlusion—Study protocol for a randomized clinical trial. *Trials* 2021, 22, 141. [CrossRef]
- 20. Penny, W.D.; Friston, K.J.; Ashburner, J.T.; Kiebel, S.J.; Nichols, T.E. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*; Elsevier: Amsterdam, The Netherlands, 2011.
- Shinohara, R.T.; Sweeney, E.M.; Goldsmith, J.; Shiee, N.; Mateen, F.J.; Calabresi, P.A.; Jarso, S.; Pham, D.L.; Reich, D.S.; Crainiceanu, C.M.; et al. Statistical normalization techniques for magnetic resonance imaging. *Neuroimage Clin.* 2014, 6, 9–19. [CrossRef]
- 22. Hazlett, Y.P.P.J.C. H Gimpel Smith R Ho S Gee JC Gerig G User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* **2006**, *31*, 1116.
- 23. Li, H.; Jiang, G.; Zhang, J.; Wang, R.; Wang, Z.; Zheng, W.S.; Menze, B. Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images. *NeuroImage* 2018, *183*, 650–665. [CrossRef]
- Van Griethuysen, J.J.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.; Fillion-Robin, J.C.; Pieper, S.; Aerts, H.J. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 2017, 77, e104–e107. [CrossRef]
- Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; ACM: New York, NY, USA, 2016; pp. 785–794. [CrossRef]
- 26. Lundberg, S.M.; Erion, G.G.; Lee, S.I. Consistent Individualized Feature Attribution for Tree Ensembles. *arXiv* 2018, arXiv:1802.03888.
- 27. XGBoost Documentation. Available online: https://xgboost.readthedocs.io/en/stable/parameter.html (accessed on 12 July 2022).
- 28. Pyradiomics Documentation. Available online: https://pyradiomics.readthedocs.io/en/latest/features.html (accessed on 12 July 2022).
- Hendrix, P.; Melamed, I.; Collins, M.; Lieberman, N.; Sharma, V.; Goren, O.; Zand, R.; Schirmer, C.M.; Griessenauer, C.J. NIHSS 24 h after mechanical thrombectomy predicts 90-day functional outcome. *Clin. Neuroradiol.* 2022, *32*, 401–406. [CrossRef] [PubMed]
- 30. Zhang, M.Y.; Mlynash, M.; Sainani, K.L.; Albers, G.W.; Lansberg, M.G. Ordinal Prediction Model of 90-Day Modified Rankin Scale in Ischemic Stroke. *Front. Neurol.* **2021**, *12*, 727171. [CrossRef]
- Stokum, J.A.; Gerzanich, V.; Simard, J.M. Molecular pathophysiology of cerebral edema. J. Cereb. Blood Flow Metab. 2016, 36, 513–538. [CrossRef] [PubMed]
- Whiteley, W.N.; Slot, K.B.; Fernandes, P.; Sandercock, P.; Wardlaw, J. Risk factors for intracranial hemorrhage in acute ischemic stroke patients treated with recombinant tissue plasminogen activator: a systematic review and meta-analysis of 55 studies. *Stroke* 2012, 43, 2904–2909. [CrossRef] [PubMed]
- Jiang, L.; Zhang, C.; Wang, S.; Ai, Z.; Shen, T.; Zhang, H.; Duan, S.; Yin, X.; Chen, Y.C. MRI Radiomics Features From Infarction and Cerebrospinal Fluid for Prediction of Cerebral Edema after Acute Ischemic Stroke. *Front. Aging Neurosci.* 2022, 14, 782036. [CrossRef]
- 34. Zhai, H.; Liu, Z.; Wu, S.; Cao, Z.; Xu, Y.; Lv, Y. Predictive value of magnetic resonance imaging-based texture analysis for hemorrhage transformation in large cerebral infarction. *Front. Neurosci.* **2022**, *16*, 923708. [CrossRef]
- 35. Wang, Y.; Liu, G.; Hong, D.; Chen, F.; Ji, X.; Cao, G. White matter injury in ischemic stroke. *Prog. Neurobiol.* **2016**, *141*, 45–60. . [CrossRef]
- Hong, S.; Giese, A.K.; Schirmer, M.D.; Bonkhoff, A.K.; Bretzner, M.; Rist, P.; Dalca, A.V.; Regenhardt, R.W.; Etherton, M.R.; Donahue, K.L.; et al. Excessive White Matter Hyperintensity Increases Susceptibility to Poor Functional Outcomes After Acute Ischemic Stroke. *Front. Neurol.* 2021, 12, 700616. [CrossRef] [PubMed]
- Boers, A.M.; Jansen, I.G.; Beenen, L.F.; Devlin, T.G.; San Roman, L.; Heo, J.H.; Ribó, M.; Brown, S.; Almekhlafi, M.A.; Liebeskind, D.S.; et al. Association of follow-up infarct volume with functional outcome in acute ischemic stroke: A pooled analysis of seven randomized trials. *J. Neurointerventional Surg.* 2018, *10*, 1137–1142. [CrossRef] [PubMed]
- Almenkerk, S.v.; Smalbrugge, M.; Depla, M.F.; Eefsting, J.A.; Hertogh, C.M. What predicts a poor outcome in older stroke survivors? A systematic review of the literature. *Disabil. Rehabil.* 2013, 35, 1774–1782. [CrossRef] [PubMed]