

# UCLA

## UCLA Previously Published Works

### Title

Predicting Hypoperfusion Lesion and Target Mismatch in Stroke from Diffusion-weighted MRI Using Deep Learning.

### Permalink

<https://escholarship.org/uc/item/31d3g6wf>

### Journal

Radiology, 307(1)

### Authors

Yu, Yannan

Christensen, Soren

Ouyang, Jiahong

et al.

### Publication Date

2023-04-01

### DOI

10.1148/radiol.220882

Peer reviewed

# Predicting Hypoperfusion Lesion and Target Mismatch in Stroke from Diffusion-weighted MRI Using Deep Learning

Yannan Yu, MD • Soren Christensen, PhD • Jiabong Ouyang, MS • Fabien Scalzo, PhD • David S. Liebeskind, MD • Maarten G. Lansberg, MD • Gregory W. Albers, MD • Greg Zaharchuk, MD, PhD

From the Departments of Radiology (Y.Y., G.Z.), Neurology (S.C., M.G.L., G.W.A.), and Electrical Engineering (J.O.), Stanford University, 1201 Welch Rd, PS-04, Mailcode 5488, Stanford, CA 94305-5488; and Department of Neurology, University of California, Los Angeles, Los Angeles, Calif (F.S., D.S.L.). Received April 14, 2022; revision requested July 5; revision received September 8; accepted October 13. **Address correspondence to G.Z.** (email: [gregz@stanford.edu](mailto:gregz@stanford.edu)).

Supported by the National Institutes of Health (grant R01-NS066506) and a Spectrum award from Stanford University.

Conflicts of interest are listed at the end of this article.

See also the editorial by Kallmes and Rabinstein in this issue.

Radiology 2023; 307(1):e220882 • <https://doi.org/10.1148/radiol.220882> • Content codes: **NR** **MR** **AI**

**Background:** Perfusion imaging is important to identify a target mismatch in stroke but requires contrast agents and postprocessing software.

**Purpose:** To use a deep learning model to predict the hypoperfusion lesion in stroke and identify patients with a target mismatch profile from diffusion-weighted imaging (DWI) and clinical information alone, using perfusion MRI as the reference standard.

**Materials and Methods:** Imaging data sets of patients with acute ischemic stroke with baseline perfusion MRI and DWI were retrospectively reviewed from multicenter data available from 2008 to 2019 (Imaging Collaterals in Acute Stroke, Diffusion and Perfusion Imaging Evaluation for Understanding Stroke Evolution 2, and University of California, Los Angeles stroke registry). For perfusion MRI, rapid processing of perfusion and diffusion software automatically segmented the hypoperfusion lesion (time to maximum,  $\geq 6$  seconds) and ischemic core (apparent diffusion coefficient [ADC],  $\leq 620 \times 10^{-6}$  mm<sup>2</sup>/sec). A three-dimensional U-Net deep learning model was trained using baseline DWI, ADC, National Institutes of Health Stroke Scale score, and stroke symptom sidedness as inputs, with the union of hypoperfusion and ischemic core segmentation serving as the ground truth. Model performance was evaluated using the Dice score coefficient (DSC). Target mismatch classification based on the model was compared with that of the clinical-DWI mismatch approach defined by the DAWN trial by using the McNemar test.

**Results:** Overall, 413 patients (mean age, 67 years  $\pm$  15 [SD]; 207 men) were included for model development and primary analysis using fivefold cross-validation (247, 83, and 83 patients in the training, validation, and test sets, respectively, for each fold). The model predicted the hypoperfusion lesion with a median DSC of 0.61 (IQR, 0.45–0.71). The model identified patients with target mismatch with a sensitivity of 90% (254 of 283; 95% CI: 86, 93) and specificity of 77% (100 of 130; 95% CI: 69, 83) compared with the clinical-DWI mismatch sensitivity of 50% (140 of 281; 95% CI: 44, 56) and specificity of 89% (116 of 130; 95% CI: 83, 94) ( $P < .001$  for all).

**Conclusion:** A three-dimensional U-Net deep learning model predicted the hypoperfusion lesion from diffusion-weighted imaging (DWI) and clinical information and identified patients with a target mismatch profile with higher sensitivity than the clinical-DWI mismatch approach.

ClinicalTrials.gov registration nos. NCT02225730, NCT01349946, NCT02586415

© RSNA, 2022

Supplemental material is available for this article.

Acute ischemic stroke is a medical emergency that imposes a substantial burden on society. Reperfusion therapies, such as intravenous thrombolysis and endovascular thrombectomy, are the only validated treatments and have maximum benefit when administered early to properly selected patients (1–4). MRI- or CT-based perfusion-diffusion mismatch has shown efficacy for selecting patients up to 9 hours after onset of clinical symptoms for thrombolysis and up to 24 hours after symptom onset for thrombectomy (4–7). However, MRI-based perfusion-diffusion mismatch identification requires relatively long scanning and postprocessing times, with at least 10 minutes for preparation, sequence acquisition, and postprocessing. CT-based hypoperfusion core mismatch exposes patients to ionizing radiation and is still

limited for accurately identifying irreversibly dead tissue (8,9) or differentiating stroke mimics (10,11). In addition, perfusion imaging requires intravenous injection of contrast agents, which may be contraindicated in patients with renal failure and allergy. Mismatch between clinical presentation, based on the National Institutes of Health Stroke Scale (NIHSS) score, and the volume of the lesion as determined with diffusion-weighted imaging (DWI) has been used in the Clinical Mismatch in the Triage of Wake Up and Late Presenting Strokes Undergoing Neurointervention With Trevo (DAWN) trial to avoid these negative aspects, demonstrating high specificity compared with perfusion-diffusion mismatch (12). However, clinical-DWI mismatch sensitivity is low (previously reported to be 53%–62% [12,13]), meaning that patients

## Abbreviations

ADC = apparent diffusion coefficient, AUC = area under the receiver operating characteristic curve, DAWN = Clinical Mismatch in the Triage of Wake Up and Late Presenting Strokes Undergoing Neurointervention With Trevo, DEFUSE = Diffusion and Perfusion Imaging Evaluation for Understanding Stroke Evolution, DSC = Dice score coefficient, DWI = diffusion-weighted imaging, LVO = large vessel occlusion, NIHSS = National Institutes of Health Stroke Scale

## Summary

A three-dimensional deep learning model trained from a multicenter data set predicted the hypoperfusion lesion and target mismatch from diffusion-weighted MRI and clinical information in acute ischemic stroke.

## Key Results

- In this retrospective multicenter study of 413 patients with acute ischemic stroke, a deep learning model predicted the hypoperfusion lesion, as defined by perfusion MRI, using only clinical information and diffusion-weighted MRI, with a median Dice score coefficient of 0.61.
- The model achieved a sensitivity of 90% and specificity of 77% for identifying target mismatch based on the DEFUSE 3 trial criteria.
- The model had higher sensitivity and accuracy for identifying target mismatch compared with clinical-DWI mismatch using the DAWN trial criteria.

who could benefit from reperfusion may be underdiagnosed (12,13).

Previous studies have explored whether subtle signal changes (14–16) or lesion patterns (17) at initial DWI could predict infarct growth and identify stroke etiologies. Although those initial DWI features demonstrated some association with hypoperfused tissue when using traditional statistical analysis, prediction accuracy was far from ideal (15,18), thus limiting the clinical application. Convolutional neural networks, a machine learning technique, automatically extract features from images by using multiple convolutional layers to make predictions. Deep convolutional neural networks, such as U-Nets, have shown advantages in stroke lesion prediction compared with traditional threshold-based methods (19–21). In this study, we aimed to use a deep learning model to predict the hypoperfusion lesion in stroke and identify patients with a target mismatch profile from DWI and clinical information alone, using perfusion MRI as the reference standard.

## Materials and Methods

### Patients

Patients with acute ischemic stroke were reviewed from two prospective multicenter trials and one single-center registry; these are the imaging Collaterals in Acute Stroke (iCAS) trial (April 2014 to June 2019;  $n = 188$ ), Diffusion and Perfusion Imaging Evaluation for Understanding Stroke Evolution (DEFUSE) 2 trial (July 2008 to October 2012;  $n = 140$ ), and University of California, Los Angeles (UCLA) stroke registry (2012–2016;  $n = 196$ ). Patients from the multicenter randomized controlled trial DEFUSE 3 (May 2016 to May 2017;  $n = 182$ ) were reviewed separately as an external generalization cohort. The iCAS (22,23) is a

multicenter observational study enrolling participants with acute ischemic stroke symptoms attributable to anterior circulation, with NIHSS scores greater than or equal to 5 and onset-to-imaging times less than or equal to 24 hours. The DEFUSE 2 protocol enrolled similar participants within a shorter onset-to-imaging time frame ( $\leq 12$  hours) and has been previously reported (3,24). The iCAS, DEFUSE 2, and DEFUSE 3 studies (ClinicalTrials.gov registration numbers NCT02225730, NCT01349946, and NCT02586415, respectively) and UCLA registry data were approved by the participating site's institutional review boards and written consent was obtained from each participant. The current study was approved for retrospective analysis by the institutional review boards. Data used in this study were compliant with the Health Insurance Portability and Accountability Act.

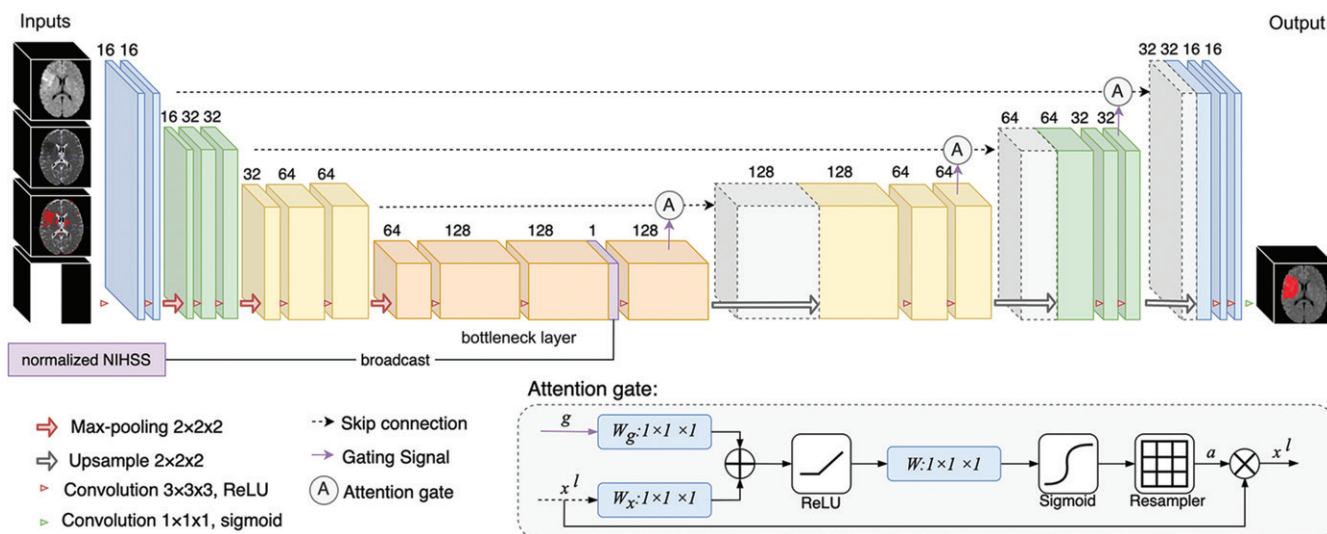
We excluded patients who did not have acute ischemic stroke located in the internal carotid artery or middle cerebral artery territories, those without adequate quality bolus perfusion MRI or DWI at arrival, those who had perfusion MRI with poor-quality reconstruction, or those with poor quality of time-to-maximum segmentation. We developed the model using subsets of patients from the iCAS and DEFUSE 2 studies and UCLA stroke registry using a fivefold split for the purposes of training, validating, and testing. The testing data set was the largest and most representative sample and was considered the primary analysis of model performance. We further evaluated the model using an external generalization cohort comprised of patients from the DEFUSE 3 study (25). Of note, all patients in the generalization cohort had a target mismatch profile because this was an inclusion criterion for the DEFUSE 3 trial. This had two consequences: (a) only sensitivity could be tested for classification and (b) all patients had large size differences between hypoperfusion and core lesions, making prediction more challenging.

### Imaging Protocol

All images were acquired at 1.5 T or 3 T. Patients underwent MRI according to each site's standard protocol, including DWI ( $b = 1000 \text{ sec/mm}^2$ ) and dynamic susceptibility contrast-enhanced perfusion MRI using gadolinium-based contrast agents. Rapid processing of perfusion and diffusion software (Rapid; RapidAI [<https://www.rapidai.com/>]) was used for postprocessing to reconstruct perfusion parameter maps and generate the ischemic core lesion (apparent diffusion coefficient [ADC] threshold  $\leq 620 \times 10^{-6} \text{ mm}^2/\text{sec}$ ) and hypoperfusion lesion (time-to-maximum threshold  $\geq 6$  seconds). The union of the hypoperfusion lesion and ischemic core was used as ground truth, as it is the relevant region for defining the perfusion-diffusion mismatch. For readability we use the term "hypoperfusion lesion," while recognizing that it could potentially include tissue that has subsequently reperfused but already suffered infarction.

### Data Preprocessing

An experienced neuroradiology researcher (Y.Y., with 8 years of experience) reviewed the ground truth segmentation from Rapid software. For segmentations with suboptimal quality, the same researcher manually removed the artifacts in the segmentation with ITK-SNAP ([www.itksnap.org](http://www.itksnap.org/)) (26) (Fig S1).



**Figure 1:** Block diagram shows the attention-gated three-dimensional U-Net model with clinical data fusion and a schematic of the attention gate. Input images include four three-dimensional image slabs sized at  $128 \times 128 \times 16$  pixels: diffusion-weighted imaging ( $b = 1000 \text{ sec/mm}^2$ ), apparent diffusion coefficient (ADC), ADC mask thresholded at  $620 \times 10^{-6} \text{ mm}^2/\text{sec}$ , and a mask indicating the side of stroke. Normalized National Institutes of Health Stroke Scale (NIHSS) scores are broadcast to the shape of the bottleneck layer and linked with the image features. The number of channels is denoted above each box and each block represents a four-dimensional vector. In an attention gate, the output of the previous layer ( $g$ ) and the symmetric encoding layer ( $x$ ) undergo convolution (with a  $1 \times 1$ -pixel kernel), summation, and rectified linear unit (ReLU) activation. Then another convolution with sigmoid activation is applied to extract the attention coefficient ( $a$ ), which is then multiplied with the skip connection.

All images were coregistered and normalized to the Montreal Neurological Institute template with SPM12 ([www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)) software implemented in Matlab version 2016b (MathWorks). Each brain volume was formatted to a size of  $128 \times 128 \times 60$  pixels. Of note, the spatial coverage of perfusion imaging was usually smaller than diffusion imaging; only voxels with both diffusion and perfusion information were included in the model and analysis.

DWI ( $b = 1000 \text{ sec/mm}^2$  images) and ADC data were normalized by the mean of their parenchymal tissue value. To preserve important information about absolute ADC values, a mask was created at an ADC less than  $620 \times 10^{-6} \text{ mm}^2/\text{sec}$  by using simple thresholding. An image volume was created to indicate the side of stroke. If the stroke was unilateral, half of the image was labeled as 1 and the other half as 0; if the stroke was bilateral, all pixels were labeled as 1. Baseline NIHSS scores were normalized into a range from 0 to 2 by dividing the NIHSS score by 21.

### Model Structure, Training, and Testing

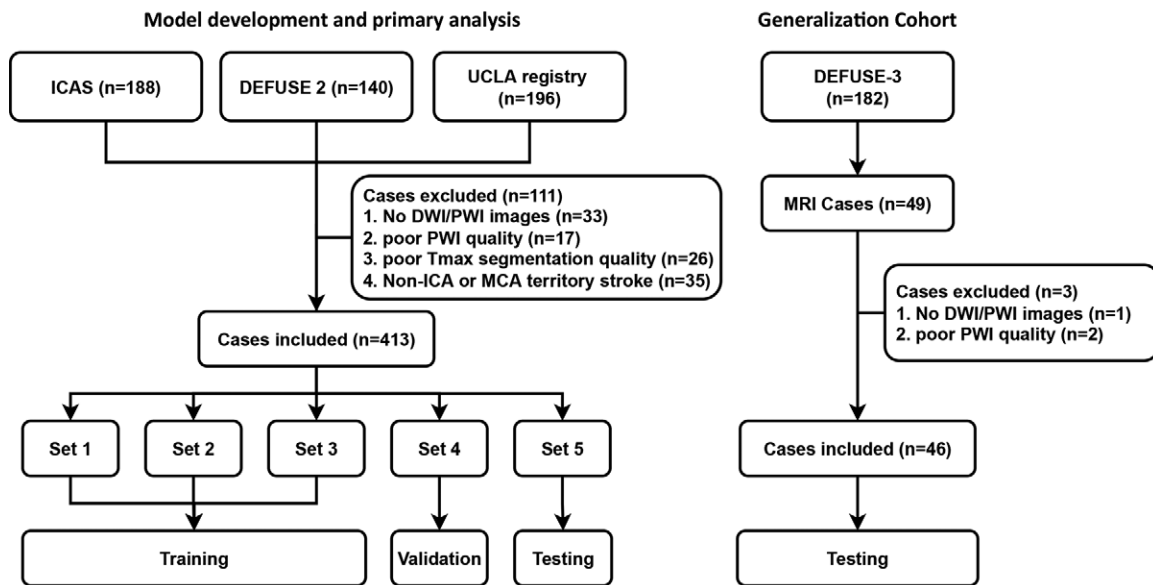
We used an attention-gated three-dimensional U-Net model with imaging and clinical data fusion. A three-dimensional U-Net (27) is identical to a two-dimensional U-Net (28,29) except that the sample and ground truth size and kernel size for convolution and max pooling layers have a third dimension. The normalized clinical information was linked with image features at the bottleneck layer before the decoders (Fig 1) (30,31). The U-Net takes a slab (16 sections with a dimension of  $128 \times 128 \times 16$  pixels) of DWI, the ADC, and the thresholded ADC mask, with the mask indicating the side of the stroke, as imaging input and the baseline NIHSS score as numeric input. A slab of hypoperfusion lesion served as the ground truth. The model outputs a probability map with voxel values ranging from 0 to 1, whereby

a value closer to 1 indicates that the voxel is more likely to be in the hypoperfusion lesion. During testing, image slabs were extracted using the sliding window method. The slabs from model outputs were then combined and averaged to generate the final probability map.

The loss function used was a combination of binary cross entropy, volume error, and Dice score coefficient (DSC) loss. Other hyperparameters, based on previous experience of similar tasks, included a learning rate of 0.0005, total epochs of 80, batch size of 32, and use of the Adam optimizer algorithm with exponential decay. Fivefold cross-validation was performed. We randomly divided the patients into five data sets; three sets were used for training, one set for validation, and one set for testing for each fold. The best model was selected based on the lowest validation loss function among all training epochs. The generalization cohort was tested using each of the five models trained during fivefold cross-validation of the primary analysis cohort. The predictions from these five models were averaged to calculate the overall performance. The code used for data processing, model training and testing, and data analysis can be found at [https://github.com/yannanyu/dwi\\_hypoperfusion\\_paper](https://github.com/yannanyu/dwi_hypoperfusion_paper).

### Performance Evaluation

The area under the receiver operating characteristic curve (AUC), DSC, and volume difference were calculated for each patient to evaluate hypoperfusion segmentation performance. AUC values were calculated within the ipsilateral stroke hemisphere, except when there were bilateral strokes. The DSC reflects the overlap between the prediction and the ground truth. It ranges from 0 to 1 with higher numbers representing more overlap, so that a DSC greater than 0.5 was considered good agreement. To calculate the DSC and lesion volume



**Figure 2:** Flowcharts of patient inclusion and exclusion. For each of the five folds in the primary analysis cohort, there were 247 patients in the training set, 83 patients in the validation set, and 83 patients in the test set. There was no overlap in patients between training, validation, or test sets. DEFUSE = Diffusion and Perfusion Imaging Evaluation for Understanding Stroke Evolution, DWI = diffusion-weighted imaging, ICA = internal carotid artery, ICAS = Imaging Collaterals in Acute Stroke, MCA = middle cerebral artery, PWI = perfusion-weighted imaging, UCLA = University of California, Los Angeles.

difference between prediction and ground truth, a threshold probability of 0.5 was chosen. If the patient has no hypoperfusion lesion (meaning that ground truth is 0), the DSC will always be 0 and the AUC cannot be calculated. Therefore, patients with no hypoperfusion lesion were excluded when reporting DSC and AUC values but included when reporting volume differences.

To evaluate triage, we tested whether each patient met the target mismatch criteria by using the DEFUSE 3 trial criteria (5) as follows: (a) ratio of hypoperfusion and core volume greater than or equal to 1.8, (b) hypoperfusion minus core volume greater than or equal to 15 mL, and (c) core volume less than or equal to 70 mL. To compare the model with existing clinical-DWI mismatch criteria, DAWN trial (ClinicalTrials.gov no. NCT02142283) criteria (4) were used as follows: (a) age greater than or equal to 80 years, NIHSS score greater than or equal to 10, core volume less than 21 mL; (b) age less than 80 years, NIHSS score greater than or equal to 10, core volume less than 31 mL; and (c) age less than 80 years, NIHSS score greater than or equal to 20, core volume 31–51 mL. Sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were calculated for target mismatch for the model prediction and clinical-DWI mismatch defined using DAWN trial criteria, with the DEFUSE 3 target mismatch criteria as reference.

Two subgroups were separately analyzed to determine different application scenarios for the model. These were (a) patients with and without a baseline core lesion as defined by Rapid software and (b) patients with large vessel occlusion (LVO), defined as occlusion of the internal carotid artery or the M1 segment of the middle cerebral artery, and those without. A scatterplot of baseline core volumes against DSCs was plotted to explore the model behavior.

The proportion of good hypoperfusion prediction (defined as a DSC >0.5) was compared between different levels of baseline core volume.

Additional comparisons were performed between images scanned at 1.5 T and 3 T (see Appendix S1).

### Statistical Analysis

We maximized our sample size for statistical analysis by incorporating all available data sets and implementing fivefold cross-validation. Statistical analysis was performed using Stata version 15.0 (StataCorp). Values were expressed as means ± SDs if the variable was normally distributed and medians with IQRs if the variable was not normally distributed. The exact McNemar test was performed to compare the accuracy of classifying target mismatch from model prediction with that of the clinical-DWI mismatch approach. The Mann-Whitney *U* test was used to compare DSC values of model performance between subgroups. The one-sample *t* test was used to compare the proportions of good hypoperfusion prediction (DSC >0.5) at different baseline core volume levels. The concordance correlation coefficient ( $\rho_c$ ) was used to analyze lesion volume predictions. Correlation was considered either excellent ( $\rho_c >0.70$ ), moderate ( $\rho_c 0.50\text{--}0.70$ ), or low ( $\rho_c <0.50$ ) (32). All tests were two sided and  $P \leq .007$  was considered indicative of a statistically significant difference based on Bonferroni correction.

## Results

### Patient Characteristics

Of 524 patients available from three studies, 413 patients met the inclusion and exclusion criteria and were included in the

**Table 1: Baseline Characteristics of the Primary Analysis and Generalization Cohorts**

Characteristic	Primary Analysis Cohort (n = 413)	Generalization Cohort (n = 46)	P Value
Age (y)*	67 ± 15	69 ± 15	.40
Sex			.17
M	207 (50)	18 (39)	
F	206 (50)	28 (61)	
Occluded territory			<.001
ICA or M1 segment	299 (72)	45 (98)	
M2 or M3 segments	90 (22)	1 (2)	
M4 segment or no occlusion	24 (6)	...	
Side of stroke			.99
Left	217 (53)	24 (52)	
Right	186 (45)	21 (46)	
Bilateral	10 (2)	1 (2)	
NIHSS score <sup>†</sup>	14 (9–19)	12 (6–18)	.21
Core volume (mL) <sup>†</sup>	19 (6–49)	17 (6–27)	.02
Hypoperfusion volume (mL) <sup>†</sup>	82 (43–132)	94 (70–134)	.52
Mismatch ratio <sup>‡</sup>	3.9 (2.2–14.3)	6.4 (4.1–16.7)	.002
Mismatched cases	283 (69)	46 (100)	<.001
Modified Rankin Scale score at 3 months <sup>†</sup>	3 (1–4)	3 (2–5)	.60

Note.—Except where indicated, data are numbers of patients, with percentages in parentheses. M1–M4 are segments of the middle cerebral artery. ICA = internal carotid artery, NIHSS = National Institutes of Health Stroke Scale.

\* Data are means ± SDs.

† Data are medians, with IQRs in parentheses.

‡ Maximum was set as 20 in cases with a baseline core volume of 0 or mismatch ratio greater than 20.

analysis (Fig 2). Hypoperfusion lesion segmentation quality was good in 328 patients and suboptimal in 85 patients, whereby suboptimal segmentations were then manually edited (Fig S1). Therefore, for each of the five folds, there were 247 patients in the training set, 83 patients in the validation set, and 83 patients in the test set. The generalization cohort comprised 46 of the 49 DEFUSE 3 study participants with MRI (two were excluded due to poor perfusion MRI quality and one due to absent perfusion MRI). The generalization cohort had more LVOs, smaller baseline core volumes, and higher mismatch ratios than the primary analysis cohort (Table 1).

### Primary Analysis Performance

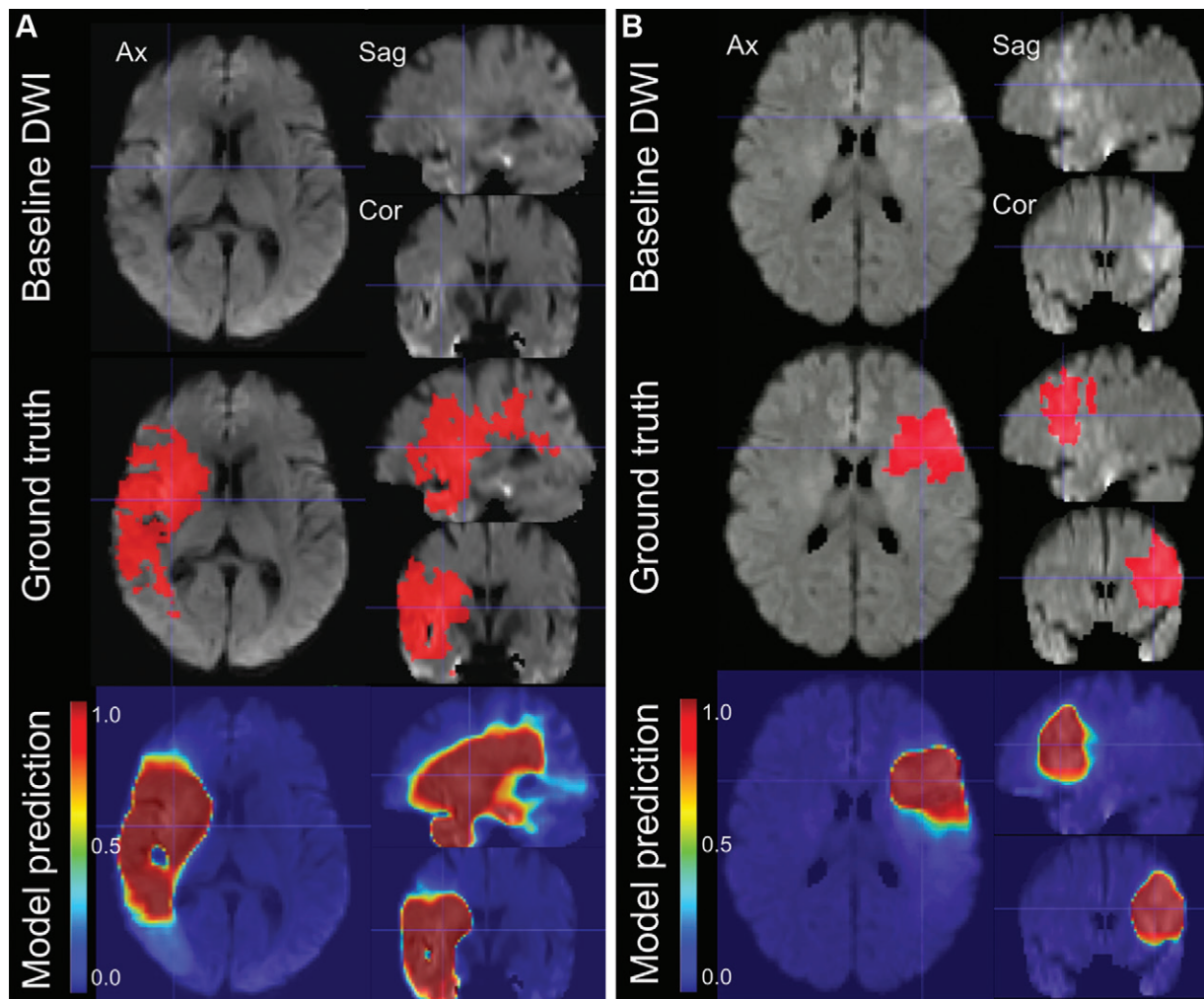
The model achieved a median AUC of 0.91 (IQR, 0.89–0.94), median DSC of 0.61 (IQR, 0.45–0.71), median volume difference of 4 mL (IQR, –37 to 41 mL), and median absolute volume difference of 40 mL (IQR, 17–70 mL) (see Table S1 for 95% CIs of the model segmentation performance). For each patient in the test set, it took approximately 8 seconds for the model to generate the test result and 24 seconds in total for the

algorithm to load and convert model predictions to a three-dimensional volume. In patients with no hypoperfusion lesion, although there was never intersection between prediction and truth, the median absolute volume difference was small (1 mL [IQR, 0–14 mL]). Predicted hypoperfusion volume had excellent correlation with true volume ( $\rho_c = 0.78$ , 95% CI: 0.75, 0.82; Fig S2). Example test cases are shown in Figure 3. The model also achieved more accurate classification for target mismatch (Table 2) at an accuracy of 86% (354 of 413 patients; 95% CI: 82, 89), which was higher than the clinical-DWI mismatch approach accuracy of 62% (95% CI: 57, 67) ( $P < .001$ ). Sensitivity and specificity of the model were 90% (254 of 283 patients; 95% CI: 86, 93) and 77% (100 of 130 patients; 95% CI: 69, 83), compared with 50% (140 of 281 patients; 95% CI: 44, 56) and 89% (116 of 130 patients; 95% CI: 83, 94) for the clinical-DWI mismatch approach. If the two methods were combined and a case was considered positive if either the proposed model or the clinical-DWI method yielded a positive classification, the accuracy would be 88% (361 of 411 patients; 95% CI: 84, 90), sensitivity 94% (264 of 281 patients; 95% CI: 90, 96), and specificity 75% (97 of 130 patients; 95% CI: 66, 81). In 157 patients in which the model prediction of target mismatch was not consistent with the clinical-DWI criteria, 135 (86%) had a target mismatch. The model showed comparable or more accurate classification of target mismatch in all subgroups (Table 2).

In patients with a baseline core lesion greater than 0 mL as determined with Rapid software, the model better predicted the hypoperfusion lesion and target mismatch than in patients with a baseline Rapid software–determined core lesion equal to 0 mL (median DSC, 0.65 vs 0.25 [ $P < .001$ ]; accuracy, 90% vs 68%) (Table 3). In patients with Rapid software–determined core lesions greater than 0, 81% (95% CI: 76, 85) of test cases had a DSC greater than 0.5, which further improved to 89% (95% CI: 82, 91) and 95% (95% CI: 91, 97;  $P < .001$ ) when the baseline core volume was greater than or equal to 10 mL and greater than or equal to 20 mL, respectively (Fig S3). Figure S4 shows examples of test cases with baseline core lesions equal to 0 as determined with Rapid software. In patients with LVO, the model was more robust than in patients without LVO (DSC, 0.63 vs 0.51 [ $P < .001$ ]; accuracy, 93% vs 66%). Figure S5 shows examples of test cases without LVO.

### Generalization Cohort Performance

In the external generalization cohort, the model achieved similar performance as that in the primary analysis data set, with a median AUC of 0.93 (IQR, 0.90–0.94;  $P = .52$ ), median DSC of 0.62 (IQR, 0.53–0.72;  $P = .26$ ), median volume difference of 7 mL (IQR, –24 to 32 mL;  $P = .99$ ), and median absolute volume difference of 30 mL (IQR, 15–65 mL;  $P = .79$ ). In patients with a baseline core volume greater than or equal to 1 mL, 80% (95% CI: 64, 90) of test cases had a DSC greater than 0.5, which further improved to 90% (95% CI: 71, 97) and 95% (95% CI: 66, 99;  $P < .001$ ) when the baseline core volume was greater than or equal to 10 mL and greater than or equal to 20 mL, respectively (Figs S2, S3). Sensitivity for the model to classify target mismatch was 96%



**Figure 3:** (A) Images in a 60-year-old woman with a National Institutes of Health Stroke Scale (NIHSS) score of 2 and right M1 segment occlusion exemplify a large vessel occlusion case. Rapid software identified a hypoperfusion lesion of 146 mL and a core of 32 mL. The model predicted 201 mL for the hypoperfusion lesion (as shown in the bottom row), with accurate spatial location and a Dice score coefficient (DSC) of 0.71. (B) Images in a 40-year-old man with an NIHSS score of 7 and left M2 segment occlusion exemplify a case without large vessel occlusion. Rapid software identified a hypoperfusion lesion of 62 mL and a core of 30 mL. The model predicted 103 mL for the hypoperfusion lesion, with accurate spatial location and a DSC of 0.64. Ax = axial, Cor = coronal, DWI = diffusion-weighted imaging, Sag = sagittal.

(44 of 46 patients; 95% CI: 83, 99) compared with 41% (19 of 46 patients; 95% CI: 28, 56) for the clinical-DWI mismatch approach.

## Discussion

Perfusion-diffusion imaging mismatch is crucial for acute ischemic stroke triage; however, it requires contrast agent injection and costly postprocessing software. Previous literature reported the potential value of using subtle diffusion-weighted imaging (DWI) signal changes to predict infarct growth in stroke patients with limited accuracy. In this study, we trained and evaluated a convolutional neural network to predict hypoperfusion and target mismatch using DWI and clinical information alone as input data, compared with clinical-DWI mismatch using DAWN trial criteria. The results demonstrated that the proposed model prediction was more sensitive (90% vs 50%) and accurate (86% vs 62%) for identification of patients with a target mismatch profile than the clinical-DWI

mismatch approach ( $P < .001$  for both). In subgroup analyses, we showed that the best use case for the proposed model is in large vessel occlusion (LVO) with baseline DWI signal changes. In cases with no LVO and/or no baseline DWI lesion, the model accuracy decreased but was overall equivalent to the clinical-DWI mismatch approach (66% and 68% vs 60% and 61%, respectively) and the model was more sensitive than the clinical-DWI approach (69% and 66% vs 35% and 46%, respectively) (Table 2). Performance in the generalization cohort achieved 96% sensitivity in identifying patients selected with the target mismatch profile, despite greater technical challenges due to smaller core lesions and higher mismatch ratios.

As reperfusion therapy is the most powerful treatment in stroke, triaging more patients who can benefit from the treatment is crucial. Leslie-Mazwi et al (13) showed that patients who did not meet DAWN trial clinical-DWI criteria but met DEFUSE 3 criteria still benefitted from thrombectomy treatment, indicating that clinical-DWI criteria are very specific but

**Table 2: Patient-wise Performance of Deep Learning Model and Clinical-DWI Criteria to Identify Target Mismatch**

Variable	Sensitivity	Specificity	PPV	NPV	Accuracy	<i>P</i> Value
Primary analysis ( <i>n</i> = 413)						
U-Net model	254/283 (90)	100/130 (77)	254/284 (89)	100/129 (78)	354/413 (86)	<.001
Clinical-DWI method	140/281 (50)	116/130 (89)	140/154 (91)	116/257 (45)	256/411 (62)	
Baseline core >0 mL ( <i>n</i> = 326)						
U-Net model	217/227 (96)	78/99 (79)	217/238 (91)	78/88 (89)	295/326 (90)	<.001
Clinical-DWI method	114/225 (51)	89/99 (90)	114/124 (92)	89/200 (45)	203/324 (63)	
Baseline core equal to 0 mL ( <i>n</i> = 87)						
U-Net model	37/56 (66)	22/31 (71)	37/46 (80)	22/41 (54)	59/87 (68)	.38
Clinical-DWI method	26/56 (46)	27/31 (87)	26/30 (87)	27/57 (47)	53/87 (61)	
Large vessel occlusion ( <i>n</i> = 299)						
U-Net model	210/222 (95)	62/70 (89)	210/218 (96)	62/74 (84)	272/292 (93)	<.001
Clinical-DWI method	117/222 (53)	66/70 (94)	117/121 (97)	66/171 (39)	183/292 (63)	
No large vessel occlusion ( <i>n</i> = 114)						
U-Net model	37/54 (69)	38/59 (64)	37/58 (64)	38/55 (69)	75/113 (66)	.29
Clinical-DWI method	18/52 (35)	49/59 (83)	18/28 (64)	49/83 (59)	67/111 (60)	

Note.—Except where indicated, data are numbers of patients, with percentages in parentheses. U-Net is the three-dimensional convolutional neural network. The clinical-DWI method is based on DAWN trial criteria. Two patients had no age data; therefore, the total number of patients for the clinical-DWI method is 411. *P* values were calculated using the exact McNemar test between the accuracy of the U-Net model and clinical-DWI mismatch criteria in identifying patients with a target mismatch profile. DAWN = Clinical Mismatch in the Triage of Wake Up and Late Presenting Strokes Undergoing Neurointervention With Trevo trial, DWI = diffusion-weighted imaging, NPV = negative predictive value, PPV = positive predictive value.

insufficiently sensitive to identify all patients who might benefit from reperfusion. Therefore, although the model prediction had lower specificity than the clinical-DWI criteria, it may be advantageous to avoid underdiagnosis for this application.

A possible clinical application for the proposed model is to accelerate MRI-based stroke protocols. Although CT is more commonly used in suspected acute ischemic stroke due to availability and center preferences, MRI better enables detection of infarcted tissue and possible stroke mimics (33) and plays a substantial role in acute stroke triage in many comprehensive stroke centers worldwide. Higher cost and longer scan times are two concerns with MRI. Shorter scan times can be achieved by adjusting MRI protocols (34) and by using synthetic MRI (35,36) and machine learning (37). If DWI shows a nonzero core lesion and MR angiography demonstrates LVO, the perfusion MRI sequence may be omitted based on the model's high performance to predict target mismatch in these subgroups. The performance of the model drops significantly if there is no core DWI lesion; this performance is not surprising as one could argue there is no way for the algorithm to identify where the hypoperfusion, if present, is located. Therefore, if DWI is negative and/or there is no evidence of LVO at MR angiography, perfusion MRI could still be performed.

The proposed model does not provide interpretable reasoning as to why a voxel is classified as part of the hypoperfused lesion. The prediction may result from the subtle signal changes not severe enough to be classified as ischemic core or from using the probability of common hypoperfusion lesions for each type of DWI lesion pattern. Previous literature suggests that a higher ADC threshold, such as  $740\text{--}780 \times 10^{-6} \text{ mm}^2/\text{sec}$ , correlates with infarct growth and likely represents at-risk tissue (15,16). Therefore, hypoperfused tissue may present with subtle DWI

or ADC abnormalities (14–16,38). It could be argued that the model just learns the shape of the middle cerebral artery distribution, which is possible if the data set only contains proximal middle cerebral artery occlusion and internal carotid artery occlusion. However, we found relatively accurate predictions even in non-LVO subgroups with DSC values greater than 0.5 (Table 3), indicating there were features in baseline DWI that implied the size of occluded arteries.

Our study had several limitations. First, as the proposed model was trained only using internal carotid artery and middle cerebral artery territory strokes, it cannot be applied to detect hypoperfusion in other stroke territories. Second, we did not test the model in patients without stroke with DWI hyperintensities; therefore, the proposed model cannot be used to diagnose stroke or assess the perfusion status of nonacute pathologic findings. Third, other sequences routinely performed in addition to DWI, such as gradient echo and MR angiography, were not included in the model, although they may help predict hypoperfusion lesion. Fourth, as sequence selection may vary at different sites, we did not include every possible sequence. Fifth, there is potential overfitting of the model to the primary analysis data set because we used fivefold cross-validation; however, this data set is already relatively diverse due to varying scanners from different institutions and varying field strengths. Additionally, the similar performance in the separate, held-out DEFUSE 3 test data set is further evidence of the model's robustness. Finally, we have not exhaustively experimented with all recent deep learning model structures or all hyperparameter combinations, although we recognize that this could improve performance.

In conclusion, by using a three-dimensional U-Net convolutional neural network, we have shown the feasibility of predicting hypoperfusion lesions and triaging patients with



**Table 3: Voxel-wise Model Performance Metrics in Different Subgroups**

Group	AUC	DSC	Volume Difference (mL)	Absolute Volume Difference (mL)
All patients (n = 413)	0.91 (0.89–0.94)	0.61 (0.45–0.71)	4 (–37 to 41)	40 (17–70)
Baseline core lesion >0 mL (n = 326)	0.92 (0.90–0.94)	0.65 (0.53–0.73)	15 (–30 to 47)	41 (20–70)
Baseline core lesion equal to 0 mL (n = 87)	0.88 (0.85–0.91)	0.25 (0.13–0.42)	–17 (–62 to 1)	28 (8–69)
Large vessel occlusion (n = 299)	0.91 (0.88–0.93)	0.63 (0.48–0.72)	2 (–44 to 41)	43 (22–73)
No large vessel occlusion (n = 114)	0.92 (0.89–0.95)	0.51 (0.20–0.66)	5 (–14 to 38)	27 (10–52)

Note.—Data are medians, with IQRs in parentheses. Baseline core lesion size was determined with rapid processing of perfusion and diffusion software. AUC = area under the receiver operating characteristic curve, DSC = Dice score coefficient.

stroke using only diffusion-weighted imaging (DWI), the National Institutes of Health Stroke Scale score, and knowledge of stroke sidedness. The proposed model is proof of concept that information from sequences that require contrast agents (eg, hypoperfusion lesion) may be predicted from noncontrast-enhanced images. Compared with the clinical presentation and DWI mismatch approach, the model prediction was more accurate and much more sensitive for identifying patients with a target mismatch profile. Such a tool may be useful to reduce MRI scan times and costs in acute stroke protocols. Further finetuning and careful design of a data set to balance different stroke subtypes is needed for stroke cases in less common vascular territories and in patients without large vessel occlusion (LVO). Finetuning the model with more balanced categorization of LVO and non-LVO may help the model to avoid potential bias.

**Author contributions:** Guarantors of integrity of entire study, **Y.Y., G.Z.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **Y.Y., G.Z.**; clinical studies, **Y.Y., D.S.L., G.Z.**; experimental studies, **Y.Y., J.O., G.Z.**; statistical analysis, **Y.Y., F.S., G.Z.**; and manuscript editing, **Y.Y., S.C., J.O., D.S.L., M.G.L., G.W.A., G.Z.**

**Data sharing:** All data generated or analyzed during the study are included in the published paper.

**Disclosures of conflicts of interest:** **Y.Y.** Recipient of a Stanford spectrum SPADA pilot grant. **S.C.** No relevant relationships. **J.O.** No relevant relationships. **F.S.** No relevant relationships. **D.S.L.** Consulting fees from Cerenovus, Genentech, Medtronic, Stryker, and Rapid Medical. **M.G.L.** Grants from National Institute of Neurological Disorders and Stroke; consulting fees from Biogen, Nektar Therapeutics, and NuvOx Pharma; payments for expert testimony. **G.W.A.** Institutional grant from National Institutes of Health; consulting fees from Genentech and iSchemaView; patents planned, issued, or pending; advisory board, Genentech; stockholder, iSchemaView. **G.Z.** Editorial board, *Radiology*.

**References**

- Jahan R, Saver JL, Schwamm LH, et al. Association Between Time to Treatment With Endovascular Reperfusion Therapy and Outcomes in Patients With Acute Ischemic Stroke Treated in Clinical Practice. *JAMA* 2019;322(3):252–263.
- Goyal M, Menon BK, van Zwam WH, et al; HERMES collaborators. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet* 2016;387(10029):1723–1731.
- Lansberg MG, Straka M, Kemp S, et al; DEFUSE 2 study investigators. MRI profile and response to endovascular reperfusion after stroke (DEFUSE 2): a prospective cohort study. *Lancet Neurol* 2012;11(10):860–867.

- Nogueira RG, Jadhav AP, Haussen DC, et al. Thrombectomy 6 to 24 Hours after Stroke with a Mismatch between Deficit and Infarct. *N Engl J Med* 2018;378(1):11–21.
- Albers GW, Marks MP, Kemp S, et al; DEFUSE 3 Investigators. Thrombectomy for Stroke at 6 to 16 Hours with Selection by Perfusion Imaging. *N Engl J Med* 2018;378(8):708–718.
- Ma H, Campbell BCV, Parsons MW, et al; EXTEND Investigators. Thrombolysis Guided by Perfusion Imaging up to 9 Hours after Onset of Stroke. *N Engl J Med* 2019;380(19):1795–1803.
- Campbell BC, Mitchell PJ, Kleinig TJ, et al; EXTEND-IA Investigators. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *N Engl J Med* 2015;372(11):1009–1018.
- Boned S, Padroni M, Rubiera M, et al. Admission CT perfusion may overestimate initial infarct core: the ghost infarct core concept. *J Neurointerv Surg* 2017;9(1):66–69.
- Martins N, Aires A, Mendez B, et al. Ghost Infarct Core and Admission Computed Tomography Perfusion: Redefining the Role of Neuroimaging in Acute Ischemic Stroke. *Intervent Neurol* 2018;7(6):513–521.
- Prodi E, Danielli L, Manno C, et al. Stroke Mimics in the Acute Setting: Role of Multimodal CT Protocol. *AJNR Am J Neuroradiol* 2022;43(2):216–222.
- Rowley H, Vagal A. Stroke and Stroke Mimics: Diagnosis and Treatment. In: Hodler J, Kubik-Huch RA, von Schulthess GK, eds. *Diseases of the Brain, Head and Neck, Spine 2020–2023*. IDKD Springer Series. Cham, Switzerland: Springer, 2020;25–36.
- Prosser J, Butcher K, Allport L, et al. Clinical-diffusion mismatch predicts the putative penumbra with high specificity. *Stroke* 2005;36(8):1700–1704.
- Leslie-Mazwi TM, Hamilton S, Mlynash M, et al. DEFUSE 3 Non-DAWN Patients. *Stroke* 2019;50(3):618–625.
- Montiel NH, Rosso C, Chupin N, et al. Automatic prediction of infarct growth in acute ischemic stroke from MR apparent diffusion coefficient maps. *Acad Radiol* 2008;15(1):77–83.
- Rosso C, Hevia-Montiel N, Deltour S, et al. Prediction of infarct growth based on apparent diffusion coefficients: penumbral assessment without intravenous contrast material. *Radiology* 2009;250(1):184–192.
- Oppenheim C, Grandin C, Samson Y, et al. Is there an apparent diffusion coefficient threshold in predicting tissue viability in hyperacute stroke? *Stroke* 2001;32(11):2486–2491.
- Cheng B, Knaack C, Forkert ND, Schnabel R, Gerloff C, Thomalla G. Stroke subtype classification by geometrical descriptors of lesion shape. *PLoS One* 2017;12(12):e0185063.
- Na DG, Thijs VN, Albers GW, Moseley ME, Marks MP. Diffusion-weighted MR imaging in acute ischemia: value of apparent diffusion coefficient and signal intensity thresholds in predicting tissue at risk and final infarct size. *AJNR Am J Neuroradiol* 2004;25(8):1331–1336.
- Yu Y, Xie Y, Thamm T, et al. Tissue at Risk and Ischemic Core Estimation Using Deep Learning in Acute Stroke. *AJNR Am J Neuroradiol* 2021;42(6):1030–1037.
- Yu Y, Xie Y, Thamm T, et al. Use of Deep Learning to Predict Final Ischemic Stroke Lesions From Initial Magnetic Resonance Imaging. *JAMA Netw Open* 2020;3(3):e200772.
- Wang K, Shou Q, Ma SJ, et al. Deep Learning Detection of Penumbral Tissue on Arterial Spin Labeling in Stroke. *Stroke* 2020;51(2):489–497.
- Zaharchuk G, Marks MP, Do HM, et al. Introducing the Imaging the Collaterals in Acute Stroke (iCAS) Multicenter MRI Trial. *Stroke* 2015;46(suppl\_1)AWMP16.

23. Thamm T, Guo J, Rosenberg J, et al. Contralateral Hemispheric Cerebral Blood Flow Measured With Arterial Spin Labeling Can Predict Outcome in Acute Stroke. *Stroke* 2019;50(12):3408–3415.
24. Albers GW, Thijs VN, Wechsler L, et al; DEFUSE Investigators. Magnetic resonance imaging profiles predict clinical response to early reperfusion: the diffusion and perfusion imaging evaluation for understanding stroke evolution (DEFUSE) study. *Ann Neurol* 2006;60(5):508–517.
25. Albers GW, Lansberg MG, Kemp S, et al. A multicenter randomized controlled trial of endovascular therapy following imaging evaluation for ischemic stroke (DEFUSE 3). *Int J Stroke* 2017;12(8):896–905.
26. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 2006;31(3):1116–1128.
27. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. arXiv 1606.06650 [preprint] <https://arxiv.org/abs/1606.06650>. Posted June 21, 2016. Accessed May 1, 2019.
28. Oktay O, Schlemper J, Le Folgoc L, et al. Attention U-Net: Learning Where to Look for the Pancreas. arXiv 1804.03999 [preprint] <https://arxiv.org/abs/1804.03999>. Posted April 11, 2018. Accessed November 20, 2018.
29. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv 1505.04597 [preprint] <https://arxiv.org/abs/1505.04597>. Posted May 18, 2015. Accessed July 1, 2018.
30. Liu Y, Jain A, Eng C, et al. A deep learning system for differential diagnosis of skin diseases. *Nat Med* 2020;26(6):900–908.
31. Robben D, Boers AMM, Marquering HA, et al. Prediction of final infarct volume from native CT perfusion and treatment parameters using deep learning. *Med Image Anal* 2020;59:101589.
32. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J* 2012;24(3):69–71.
33. Brunser AM, Hoppe A, Illanes S, et al. Accuracy of diffusion-weighted imaging in the diagnosis of stroke in patients with suspected cerebral infarct. *Stroke* 2013;44(4):1169–1171.
34. Nael K, Khan R, Choudhary G, et al. Six-minute magnetic resonance imaging protocol for evaluation of acute ischemic stroke: pushing the boundaries. *Stroke* 2014;45(7):1985–1991.
35. Tanenbaum LN, Tsiouris AJ, Johnson AN, et al. Synthetic MRI for Clinical Neuroimaging: Results of the Magnetic Resonance Image Compilation (MAGiC) Prospective, Multicenter, Multireader Trial. *AJNR Am J Neuroradiol* 2017;38(6):1103–1110.
36. Ryu KH, Baek HJ, Skare S, et al. Clinical Experience of 1-Minute Brain MRI Using a Multicontrast EPI Sequence in a Different Scan Environment. *AJNR Am J Neuroradiol* 2020;41(3):424–429.
37. Bash S, Wang L, Airriess C, et al. Deep Learning Enables 60% Accelerated Volumetric Brain MRI While Preserving Quantitative Performance: A Prospective, Multicenter, Multireader Trial. *AJNR Am J Neuroradiol* 2021;42(12):2130–2137.
38. Desmond PM, Lovell AC, Rawlinson AA, et al. The value of apparent diffusion coefficient maps in early cerebral ischemia. *AJNR Am J Neuroradiol* 2001;22(7):1260–1267.