Predicting final ischemic stroke lesions from initial diffusion-weighted images using a deep neural network

Sanaz Nazari-Farsani [a], Yannan Yu [a,b], Rui Duarte Armindo [a,c], Maarten Lansberg [d], David S. Liebeskind [e], Gregory Albers [d], Soren Christensen [d], Craig S. Levin [a], Greg Zaharchuk [a,*]

[a] Department of Radiology, Stanford University, CA, USA
[b] Internal Medicine Department, University of Massachusetts Memorial Medical Center, University of Massachusetts, Boston, USA
[c] Department of Neuroradiology, Hospital Beatriz Ângelo, Loures, Lisbon, Portugal
[d] Department of Neurology, Stanford University, CA, USA
[e] Department of Neurology, University of California Los Angeles, Los Angeles, CA, USA

A R T I C L E   I N F O

A B S T R A C T

Background: For prognosis of stroke, measurement of the diffusion-perfusion mismatch is a common practice for estimating tissue at risk of infarction in the absence of timely reperfusion. However, perfusion-weighted imaging (PWI) adds time and expense to the acute stroke imaging workup. We explored whether a deep convolutional neural network (DCNN) model trained with diffusion-weighted imaging obtained at admission could predict final infarct volume and location in acute stroke patients.

Methods: In 445 patients, we trained and validated an attention-gated (AG) DCNN to predict final infarcts as delineated on follow-up studies obtained 3 to 7 days after stroke. The input channels consisted of MR diffusion-weighted imaging (DWI), apparent diffusion coefficients (ADC) maps, and thresholded ADC maps with values less than 620×10^{-6} mm^2/s, while the output was a voxel-by-voxel probability map of tissue infarction. We evaluated performance of the model using the area under the receiver-operator characteristic curve (AUC), the Dice similarity coefficient (DSC), absolute lesion volume error, and the concordance correlation coefficient (ρ_c) of the predicted and true infarct volumes.

Results: The model obtained a median AUC of 0.91 (IQR: 0.84–0.96). After thresholding at an infarction probability of 0.5, the median sensitivity and specificity were 0.60 (IQR: 0.16–0.84) and 0.97 (IQR: 0.93–0.99), respectively, while the median DSC and absolute volume error were 0.50 (IQR: 0.17–0.66) and 27 ml (IQR: 7–60 ml), respectively. The model's predicted lesion volumes showed high correlation with ground truth volumes (ρ_c = 0.73, p < 0.01).

Conclusion: An AG-DCNN using diffusion information alone upon admission was able to predict infarct volumes at 3–7 days after stroke onset with comparable accuracy to models that consider both DWI and PWI. This may enable treatment decisions to be made with shorter stroke imaging protocols.

1. Introduction

Since time is a key factor affecting outcomes after treatment in acute ischemic stroke (AIS), patients should be imaged and treated as quickly as possible. Currently, the diffusion-perfusion mismatch paradigm is a commonly used method for triaging patients to endovascular treatment (Nighoghossian et al., 2003), with perfusion-weighted imaging (PWI) giving information about tissue at risk of infarction in the absence of reperfusion. Unlike diffusion weighted imaging (DWI), the patient preparation, scanning, and post-processing procedures for PWI are relatively time-consuming. In addition, the injection of the contrast agent can cause complications for patients with advanced kidney dysfunction (High et al., 2007; Kuo et al., 2007). Accordingly, prediction of final infarct volume using only DWI acquired at stroke onset has the potential to save time and minimize the cost and complications related to contrast agent injection.

Recently, there has been an increase in research activity surrounding lesion prediction using machine learning (ML) and deep learning (DL)

---

* Corresponding author at: The Richard M. Lucas Center for Imaging, Department of Radiology, Stanford University, 1215 Welch Rd, Mailcode 5488, Modular D, Stanford, CA 94305, USA.
E-mail address: gregz@stanford.edu (G. Zaharchuk).

(Gillmann et al., 2021; Hakim et al., 2021; Kijowski et al., 2020; Pinto et al., 2018; Yu et al., 2020). Deep convolutional neural networks (DCNNs) have outperformed other algorithms in a variety of medical image analyses, including stroke lesion segmentation and prediction with multi-modal magnetic resonance imaging (MRI) (Bernal et al., 2019; Cheng et al., 2017; Choi et al., 2016; Karthik et al., 2019; Pinto et al., 2018; Xue et al., 2020; Yu et al., 2021, 2020). More specifically, previous studies have predicted final infarct volume from either baseline CT perfusion (CTP) or DWI and PWI. Lucas et al. applied a U-Net CNN to the baseline CTP and reported a Dice similarity coefficient (DSC) of 0.46 (Lucas et al., 2018). Robben et al. deployed a CNN with four input channels including 3D CTP, down-sampled CTP, arterial input function (AIF, time), and metadata. They reported a DSC of 0.48 (Robben et al., 2020). Choi and colleagues deployed an ensemble of 3D multiscale residual U-Nets and a fully convolutional network which took DWI and PWI as input channels (Choi et al., 2016). Their model resulted in a DSC of 0.31. Pinto et al. developed a combined model of MR images (DWI and PWI) and clinical information for stroke lesion outcome prediction (Pinto et al., 2018), reporting a DSC of 0.35. Yu et al. also tried to predict the final ischemic stroke lesions from initial MRI (DWI and PWI) using an attention-gated U-Net that led to a DSC of 0.53 (Yu et al., 2020).

Given the relative challenge of acquiring and processing PWI or CTP, we endeavored to determine whether it would be possible to estimate final lesion size from admission DWI only. If true, this could streamline the acute stroke imaging workup and yield insights that might be valuable for patient triage. To our knowledge, no prior studies have attempted to predict final infarct segmentations from only baseline DWI, which is the fundamental novelty of this paper.

2. Materials and methods

2.1. Study population

In this study, a total of 520 patients were reviewed from a single center registry (158 cases from the University of California, Los Angeles [UCLA]), and three clinical trials (117 cases from imaging Collaterals in Acute Stroke [iCAS] (Thamm et al., 2019; Zaharchuk et al., 2015), 60 cases from Diffusion and Perfusion Imaging Evaluation for Understanding Stroke Evolution [DEFUSE] (Ogata et al., 2013), and 110 cases from DEFUSE-2 (Lansberg et al., 2012)). All trial patients signed written informed consent, while the UCLA Institutional Review Board approved the stroke registry for retrospective data analysis. More detailed inclusion and exclusion criteria of the trials can be found in Lansberg et al. (2012), Thamm et al. (2019), and Zaharchuk et al. (2015). After excluding patients without confirmed anterior circulation stroke or lack of follow-up imaging in the 3 to 7 days period, we were left with an analysis cohort of 445 patients (see Fig. 1 for a flow chart of patient inclusion). For DEFUSE and DEFUSE-2, the patients' baseline scans were obtained within 12 h from stroke onset, while for iCAS, imaging was obtained within 24 h after stroke onset. In the UCLA dataset, the
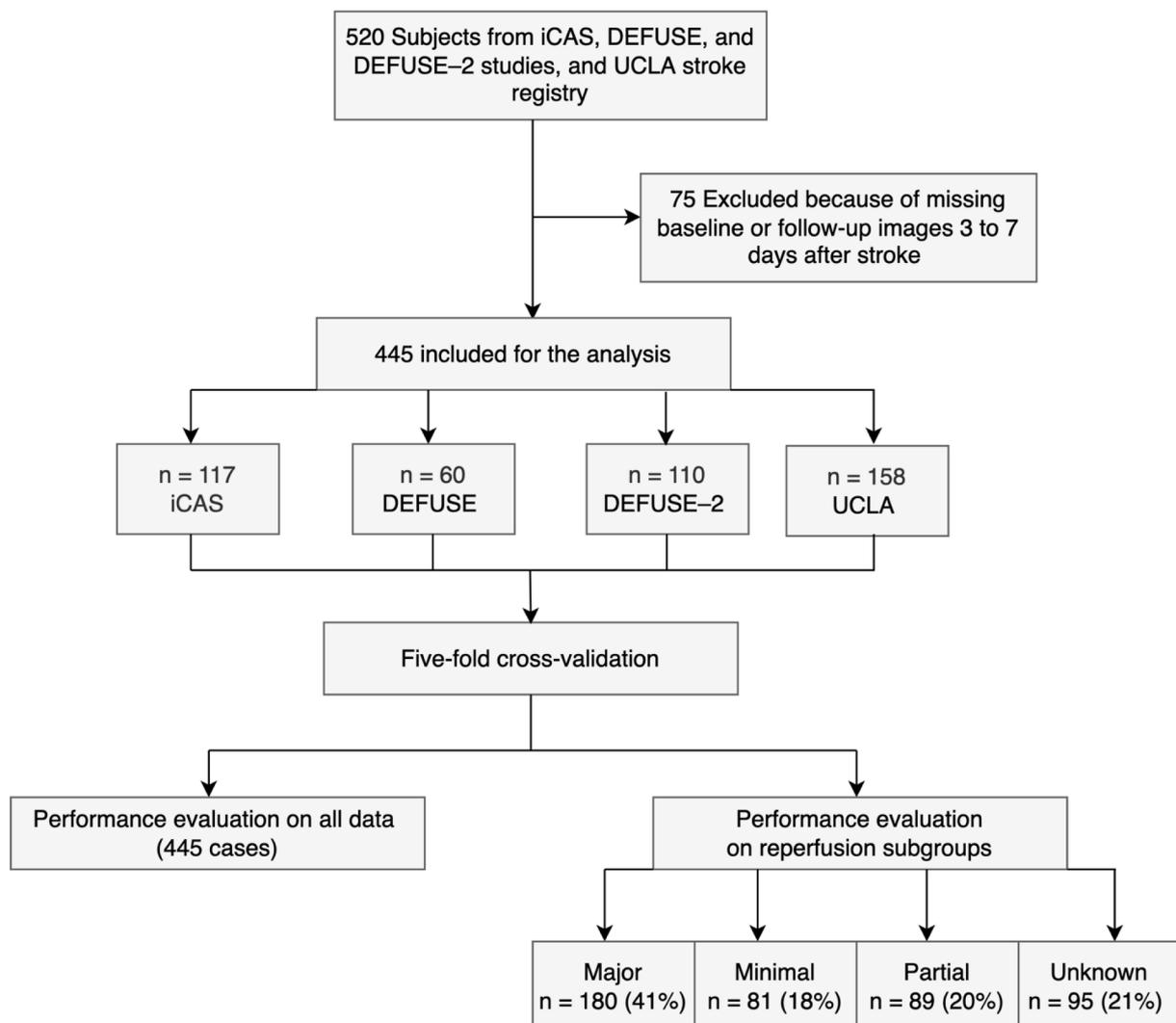
Fig. 1. Flow diagram of the study.

Table 1
Clinical data in all patinets and reperfusion groups.

Demographic	Reperfusion Status				
	All (N = 445)	Major (N = 180)	Partial (N = 89)	Minimal (N = 81)	Unknown (N = 95)
Male, number (%)	222 (50)	84 (46)	40 (45)	42 (51)	56 (58)
Age, mean (SD)	67 (15)	69 (15)	73 (14)	66 (16)	66 (13)
Hypertension, number (%)	305 (68)	121 (67)	62 (69)	57 (70)	65 (68)
Diabetes, number (%)	115 (26)	48 (26)	22 (25)	21 (26)	24 (25)
Dyslipidemia, number (%)	170 (38)	65 (36)	39(43)	34 (42)	32(24)
Atrial fibrillation, number (%)	137 (31)	64 (35)	35 (39)	18 (22)	20 (21)
Treatment methods, number (%)	169	55	35	38	41
IV tPA only	(38)124	(31)58	(39)26	(47)14	(43)20
Direct thrombectomy	(28)123	(32)63	(29)24	(17)16	(21)26
Bridging therpy	(28)29	(35)4	(27)4	(20)13	(28)8
No treatment	(6)	(2)	(5)	(16)	(8)
Onset to treatment time, hr, median (IQR)	6.2 (4.7–8.7)	5.9 (4.6–9.4)	5.5 (3.8–7.5)	6.2 (4.4–8.3)	6.8 (5.4–9.4)
Baseline lesion core, mL, median (IQR)	15 (3–39)	12 (3–28)	24 (10–62)	22 (5–66)	8 (0–35)
Baseline NIHSS, median (IQR)	13 (8–19)	14 (8–19)	16 (11–19)	13(9–18)	10 (5–15)
Symptomatic hemorrhage	158 (35)	64 (35)	46 (51)	22 (27)	26 (27)
Reperfusion rate, median (%)	81 (26–100)	100 (93–100)	54 (37–69)	−16 (-44–6)	
Final infarct volume, mL, median (IQR)	50 (15–123)	30 (11–73)	107 (47–186)	80 (31–225)	38 (4–101)
90-day mRS	3 (1–4)	2(1–3)	4(2–5)	3(1–4)	3(1–4)

majority of patients had their baseline images within 6 h, but there were a few with missing data points. More detail on the cohorts can be found in Table 1.

2.2. Imaging protocol

All the subjects of this study underwent baseline DWI according to the specific site's protocol. These were obtained on both 1.5 T and 3 T scanners, including all major vendors, with echo-planar imaging using a range of standard clinical parameters (TR 4000–10000 ms, TE 70–107 ms, slice thickness 3–5 mm, FoV 20–24 cm, b = 1000 s/mm^2). Three to seven days after stroke onset, DWI or T2-weighted fluid-attenuated inversion recovery (FLAIR) images were obtained. The ground truth (GT) final infarct lesions were segmented on the follow-up DWI or FLAIR by a neuroradiologist who was blinded to all clinical information.

2.3. Image analysis and preprocessing

All images of this study were co-registered and spatially normalized to Montreal Neurological Institute (MNI) template space using SPM12 software (Statistical Parametric Mapping, The Wellcome Trust Centre for Neuroimaging). Spatial normalization accelerates the model training since the model does not require learning the brain structure and orientation of the individuals. For the intensity normalization of the DWI and apparent diffusion coefficient (ADC) images, first, all the background (non-brain) pixels were set to zero. Then, using these voxels,

irrelevant features of the background (Oktay et al., 2018). AGs proven to be specifically useful for the prediction of small abnormalities such as tumors and necrosis (Mathews and Mohamed, 2022). The model had 14 convolutional layers (3x3) with ReLU activation functions followed by a final convolutional layer with a sigmoid activation function to obtain the probability values. Literature on the ISLES 2018 challenge (Hakim et al., 2021) revealed that application of a hybrid loss function including Dice loss and weighted binary cross-entropy results in better performance and stability of the model. According to our previous experience adding volume loss to the hybrid loss function further improves the volume prediction in stroke lesions (Yu et al., 2020).

Thus, we used a combination of four loss functions in the model: weighted binary cross-entropy, mean absolute error (L1 loss), Dice similarity coefficient (DSC), and volume loss. Weighted binary cross-entropy balances the positive and negative voxels in the brain, since stroke lesions are only present in a relatively small number of the overall brain image voxels. The weights for positive and negative voxels were determined based on the positive and negative ratio of the voxels across each training batch. Where N_- and N_+ are the number of negative and positive voxels per batch respectively.

$$R_0 = \frac{1}{\frac{N_-}{N_+} + 1}$$

$$R_1 = 1 - R_0$$

$$Weighted\ binary\ cross\ entropy = -\frac{1}{N}\sum_{i=0}^{N} R_1\, y_i \log(p_i) + R_0\,(1 - y_i)\log(1 - p_i)$$

mean normalization was performed. To preserve important quantitative information on ADC images, a binary mask was created of voxels with ADC values less than 620×10^{-6} mm^2/s using simple threshold filtering.

2.4. Neural network

The architecture of the DCNN was a 3D attention gated (AG) U-net with ReLU activation function and ADAM optimizer (Kingma and Ba, 2014). AGs highlight the relevant features of the target and suppress the

p_i is the predicted probability and y_i is the ground truth value for the ith voxel (0 = no infarct, 1 = infarct). N is the total number of voxels.

$$L1\ loss = \sum_{i=0}^{N} \left| y_i - p_i \right|$$

$$Dice\ loss = 1 - \frac{2N_{TP}}{2N_{TP} + N_{FP} + N_{FN}}$$

N_{TP}, N_{FP}, and N_{FN} represent the number of true positives, false positive, and false negative respectively. The volume loss was calculated as absolute sum of the difference between the predicted and true values, divided by the total number of voxels.

$$Volume\ loss = \frac{\left|\sum_{i=0}^{N}p_i - \sum_{i=0}^{N}y_i\right|}{N_+}$$

When combining the loss functions, we assigned a weight of 0.5 to DSC loss and volume loss since these two loss functions penalize the weights of the DCNN model for the same parameter. By assigning an index of 0.5 to these two loss functions, they were adjusted to the same scale as the weighted cross-entropy loss and the L1 loss.

$$Loss = Weighted\ binary\ cross\ entropy + L1\ loss + 0.5 \times DSC\ loss + 0.5 \times Volume\ loss$$

The network received 3 input channels; DWI, ADC, and thresholded ADC. We also examined the performance of the network when only DWI or ADC (but not both) were used as inputs. The model was trained over 80 epochs with a batch size of 32 and learning rate of 0.0005. Dropout ratio of 0.25 was used in our network to reduce overfitting and to speed up training (Srivastava et al., 2014). For data augmentation, images were mirrored around the midline. The goal was to predict the binary masks of the final infarct lesions delineated on the 3–7 day follow-up images. The model's output was a probability map of voxel values ranging from 0 to 1 with values close to 1 indicating a higher probability that the voxel is part of the final infarct. We applied a threshold of 0.5 to classify the voxels into either lesion or non-lesion tissue. More detail about the network architecture can be found in Fig. 2.

2.5. Performance evaluation

To utilize all available data and to test the generalizability of the model performance across all subjects, five-fold cross-validation (CV) was performed so that in each iteration the model was trained on 4 folds and tested on the fifth fold (see Supplementary Materials for a breakdown of the different trials into the 5 different folds). No patients were simultaneously in the training and test sets. To measure the ability of the model to distinguish infarct from non-infarcted regions, area under the curve (AUC) was calculated. AUC has a range between 0 and 1 with higher values showing better performance of the network in distinguishing the classes, in this case, infarcted vs non-infarcted tissue. Using a probability threshold of 0.5, we calculated sensitivity, specificity, DSC, volume error, absolute volume error, and Youden index. DSC was calculated as:

$$Dice\ similarity\ coefficient = \frac{2N_{TP}}{2N_{TP} + N_{FP} + N_{FN}}$$

DSC has a range of 0 to 1 with higher DSC values corresponding to a better overlap of the two segmentations. Lesion volume error and absolute lesion volume error were calculated as the difference and absolute difference between the volumes of the predicted lesion and the ground truth lesion. Youden index (sensitivity + specificity −1) is a statistic that reports on the performance of a dichotomous test, ranging from 0 (no value) to 1 (perfect prediction).

After five-fold cross-validating the model across the whole dataset, we analyzed its performance once across all subjects and once across patient subgroups based on their documented reperfusion status at 24 h following stroke, which was assessed using the reperfusion rate (Bivard et al., 2013; Yu et al., 2020), defined as:

$$Reperfusion\ Rate = 100 \times 1 - \frac{Tmax_{24hr} > 6\ seconds\ lesion}{Tmax_{baseline} > 6\ seconds\ lesion}$$

where Tmax represents the time to the peak of the residue function as measured by PWI (RAPID, Ischemaview, Redwood City, CA, USA). A reperfusion rate of > 80 % was considered as major reperfusion, > 20 % and < 80 % as partial reperfusion, and < 20 % as minimal (Bivard et al., 2013). Patients with missing follow-up reperfusion status were categorized as "unknown".

We further analyzed the model performance in subgroups based on the time interval between the stroke onset to treatment with tissue plasminogen activator (t-PA) and stroke onset to imaging. In this analysis, the median of the time intervals was considered as the cut-off value to create two groups of short-time period and long-time periods for each paradigm. Finally, while the primary analysis focused on a probability threshold of 0.5, performance was also evaluated with other thresholds ranging between 0.1 and 0.9.

2.6. Statistical analysis

The statistical analyses of this study were performed in Python 3.7.0, using Scipy package (1.5.1) (Virtanen et al., 2020). For the comparison of DSC and absolute volume error between different reperfusion groups, the Kruskal-Wallis equality of populations rank test was performed. Paired-sample Wilcoxon tests were performed to compare the performance of the deep learning model and a simple ADC thresholding model, where final infarct was considered to be tissue with ADC $< 620 \times 10^{-6}$ s/mm^2 at baseline. In this comparison, we examined whether the performance of the proposed model was better than a simple model that only considered baseline abnormal DWI as the final infarct.
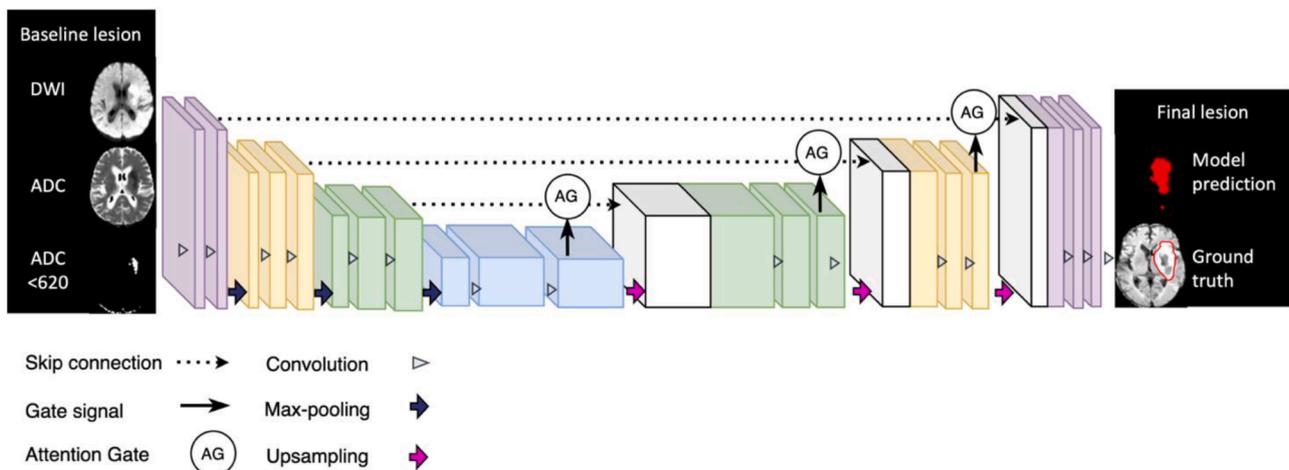
Fig. 2. The block diagram of the attention-gated U-net, as well as the network's input and output.

Table 2
Model performance in all patients and reperfusion groups.

Metrics	Reperfusion Status					
Median (IQR)	All patients (n = 445)	Major (n = 180)	Partial (n = 89)	Minimal (n = 81)	Unknown (n = 95)	P-value
AUC	0.91 (0.84–0.96)	0.92 (0.83–0.95)	0.91 (0.86–0.95)	0.92 (0.84–0.96)	0.89 (0.76–0.96)	0.45
Sensitivity	0.60 (0.16–0.84)	0.61 (0.15–0.84)	0.65 (0.36–0.83)	0.60 (0.37–0.70)	0.41 (0.01–0.80)	0.04
Specificity	0.97 (0.93–0.99)	0.97 (0.94–0.99)	0.95 (0.90–0.98)	0.96 (0.93–0.99)	0.98 (0.94–0.99)	< 0.007
Youden Index	0.55 (0.13–0.76)	0.58 (0.12–0.78)	0.62 (0.33–0.74)	0.58 (0.25–0.76)	0.35 (0.01–0.73)	0.05
DSC	0.50 (0.17–0.66)	0.46 (0.12–0.62)	0.58 (0.37–0.70)	0.56 (0.33–0.70)	0.36 (0.01–0.60)	< 0.007
Volume error, ml	0 (−22–30)	4 (-10–30)	2 (-35–48)	−7 (-54–20)	0 (−22–31)	0.03
Absolute volume error, ml	27 (7–60)	20 (6–41)	39 (17–77)	39 (11–76)	20 (4–68)	< 0.007

P-value listed represents the significance of differences between any of the groups (minimal, major, partial, and unknown).

Concordance correlation coefficient (ρ_c) and Bland-Altman plots were used to analyze the relationship of the lesion volumes predicted by the neural network and the ground truth. To analyze the relationship between the model performance in terms of DSC and the baseline lesion volume size, the Spearman correlation coefficient was calculated. We also evaluated model performance based on reperfusion status and on ground truth lesion size. All tests were two-sided, and after Bonferroni correction for multiple comparisons (n = 7) in reperfusion groups, a p-value of less than or equal to 0.007 was considered as statistically significant.

3. Results

The evaluation metrics of the DCNN model including AUC, sensitivity, specificity, Youden Index, DSC, volume error, and absolute volume error are summarized in Table 2 for all the subjects and in subgroups. The DCNN model showed a median AUC of 0.91 (IQR: 0.84–0.96). Using a probability threshold of 0.5, median sensitivity, specificity, and Youden Index were 0.60 (IQR: 0.16–0.84), 0.97 (IQR: 0.93–0.99), and 0.50 (IQR: 0.21–0.70) respectively. Comparing the predicted lesions to the ground truth resulted in a median DSC of 0.50 (IQR: 0.17–0.66), volume error of 0 ml (IQR: −22 to 30 ml), and an absolute volume error of 27 ml (IQR: 7–60 ml) (Table 2). Subgroup analysis of the results based on the time interval between the stroke onset and t-PA, and the time between stroke onset and imaging did not reveal any significant differences (see Supplementary Materials). The model performance using different probability thresholds ranging from 0.1 to 0.9 are summarized in Table 3.

The DCNN model predicted the final stroke lesions significantly better than the simple ADC thresholding method, with a median DSC of 0.50 compared to 0.18 (p < 0.01). Similar findings were seen for

absolute volume error, with a median absolute volume error of 27 ml compared to 64 ml, p < 0.01) (Table 4). Examples of final lesion prediction using the DCNN model and simple thresholding model are illustrated in Fig. 3.

The volumes predicted by the model showed a high correlation of 0.73 with the ground truth lesion volumes (p < 0.001) (Fig. 4). Direct comparison of the model outputs and ground truth was performed using Bland-Altman plot with 95 % limits of agreement (Fig. 5), with bias close to zero except for the minimal reperfusion group, where a bias of approximately 20 ml was seen. Finally, as shown in Fig. 6, there was a strong correlation between the model accuracy and lesion size ($\rho = 0.64$, p < 0.01).

In 43 cases, the model predicted a lesion volume of zero. These subjects had significantly smaller baseline lesions (median baseline lesion size of 0 ml vs 18 ml, p < 0.01), final lesions (median final lesion size of 3.2 ml vs 59.5 ml, p < 0.01), and milder strokes (median baseline NIHSS of 6 vs 14, p < 0.01) compared to the rest of the subjects. In addition, there were 94 cases in our dataset without any admission DWI lesion as determined by thresholding criteria. The performance of the model as measured by DSC was significantly lower in this group compared with cases with non-zero admission DWI lesions (median DSC of 0.10 vs 0.55, P < 0.01).

Finally, we examined performance of this model (that used both DWI and ADC inputs) with models that used one or the other (but not both). In general, we found reduced performance, particularly for DSC when using these models. Further information can be found in the Supplementary Materials.

4. Discussion

Using only baseline DWI acquired at stroke onset, we demonstrate

Table 3
Summary of the model performance using different probability thresholds ranging from 0.1 to 0.9.

Threshold	Sensitivity	Specificity	Youden Index	DSC	Volume error (ml)	Absolute volume error (mL)
0.1	0.71	0.95	0.63	0.45	17	39
0.2	0.66	0.96	0.61	0.46	9	33
0.3	0.63	0.97	0.60	0.48	3	30
0.4	0.61	0.97	0.56	0.49	2	28
0.5	**0.59**	**0.98**	**0.55**	**0.50**	**0**	**27**
0.6	0.57	0.98	0.52	0.49	0	25
0.7	0.54	0.98	0.50	0.48	−1	24
0.8	0.51	0.98	0.47	0.47	−3	22
0.9	0.45	0.99	0.42	0.46	−7	21

*Values are presented as median.

Table 4

Comparison of the deep learning model and ADC-thresholding model performance in all patients using a probability threshold of 0.5.

Metrics' median (IQR)	AUC	Sensitivity	Specificity	DSC	Volume error, (ml)	Absolute volume error, (ml)
DCNN	0.91	0.60	0.97	0.50	0	27
	(0.84–0.96)	(0.16–0.84)	(0.93–0.99)	(0.17–0.66)	(−22–30)	(7–60)
ADC-thresholding	0.62	0.26	0.98	0.18	7	64
	(0.56–0.69)	(0.12–0.40)	(0.97–0.99)	(0.10–0.35)	(−54–66)	(25–96)
P-value	< 0.001	< 0.001	< 0.001	< 0.001	0.50	< 0.001
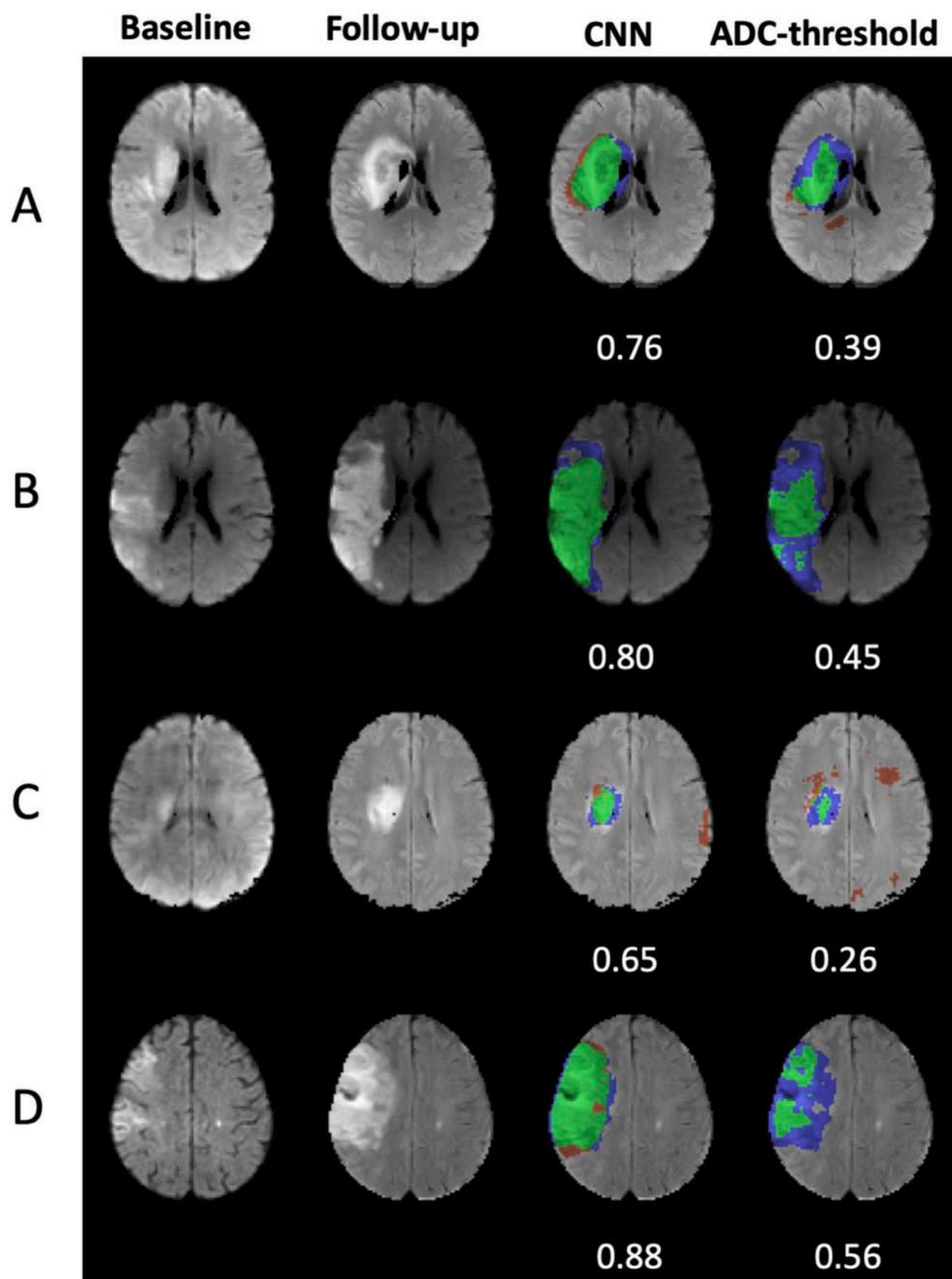
Fig. 3. Examples of CNN model prediction compared with ADC thresholding in different reperfusion groups including major (A), minimal (B), partial (C), and unknown (D). The Dice similarity coefficients (DSC) shown below the images were calculated compared to the ground truth in all slices. Green area shows true positives, blue area false negatives, and red area false positives. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that an attention-gated DCNN can accurately predict the final infarct volume 3–7 days after stroke onset in a large dataset of 445 AIS patients. The main highlight of the study is that neither the PW images nor the reperfusion status of patients were considered when training the DCNN model. The model, however, shows comparable accuracy in terms of DSC when compared to prior studies that used both DWI/PWI or CT perfusion images as input (DSC of 0.24–0.53 as compared to 0.50 in our study) (Lucas et al., 2018; Pinto et al., 2018; Robben et al., 2020; Yu et al., 2020). As PWI can extend the scan time and cost of the initial stroke imaging workup and may be contraindicated for some patients, this suggests an alternative method using only diffusion information might be useful for prediction of the final infarct size (High et al., 2007; Kuo et al., 2007).

Comparing the performance of the model in different reperfusion
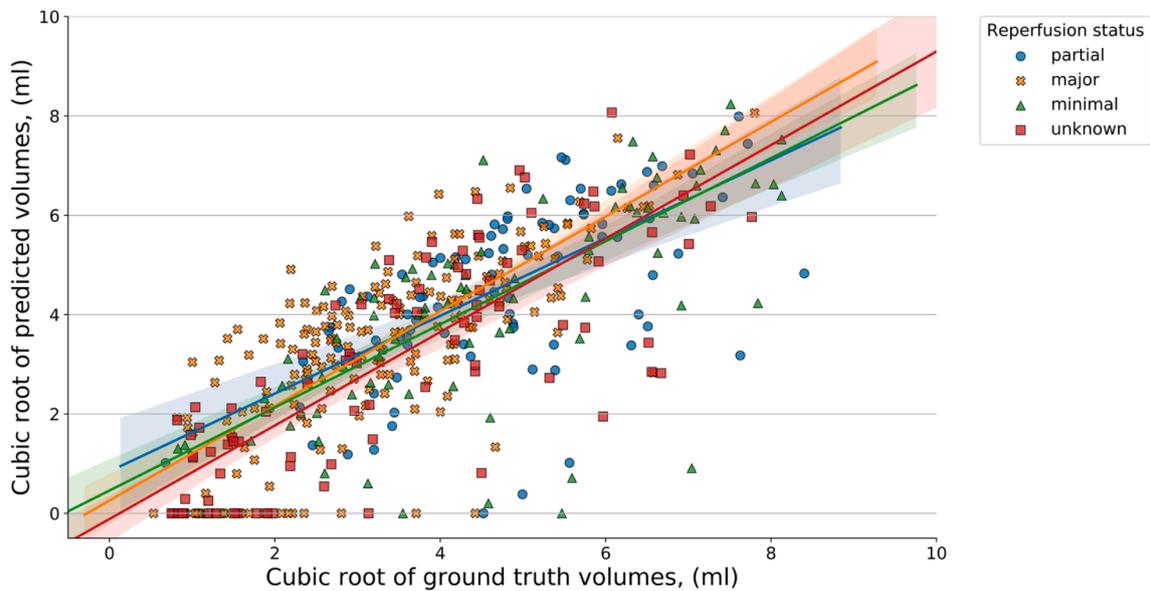
Fig. 4. The correlation of lesion volumes from the model prediction and ground truth manually delineated infarct size at 3–7 days after stroke, plotted using the cube root of the lesion sizes ($\rho_c = 0.73$, $p < 0.01$). In each subset, the solid lines represent the best linear fit function and the colored areas represent the 95 % confidence interval.
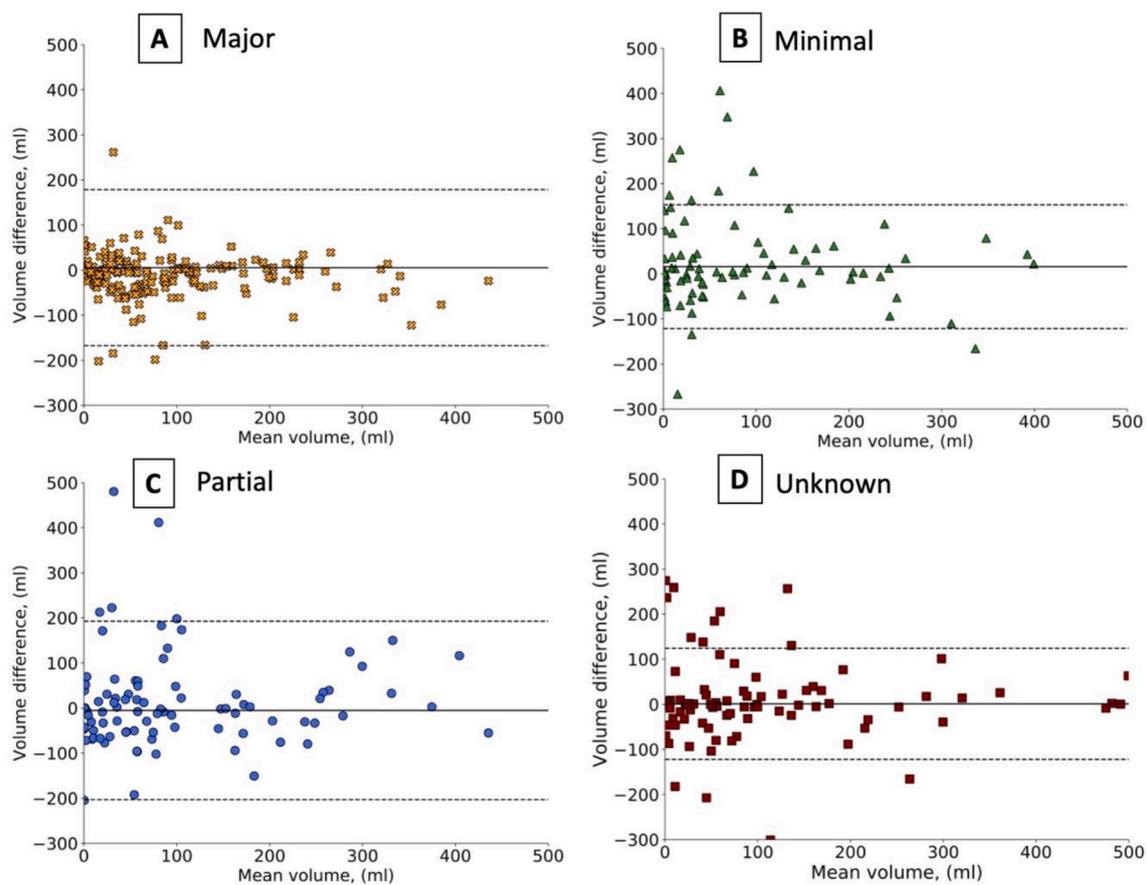
Fig. 5. Bland-Altman plots for patients in different reperfusion groups including major (A), minimal (B), partial (C), and unknown (D). The X-axis represents the mean volume, and the Y-axis represents the volume difference between the predicted and ground truth lesions. The solid line represents the bias, and the dashed lines represent the upper and lower 95% limits of agreement. Note that the error increases for larger lesions.

groups based on Bland-Altman plots (Fig. 5), we found a bias of about 20 ml for the minimal reperfusion group, suggesting that the DCNN model underestimated the final lesion size in this group. This was not surprising since the model trained in this study does not consider the PWI or reperfusion status of the patients during training, and larger lesion growth would be expected in patients with minimal reperfusion
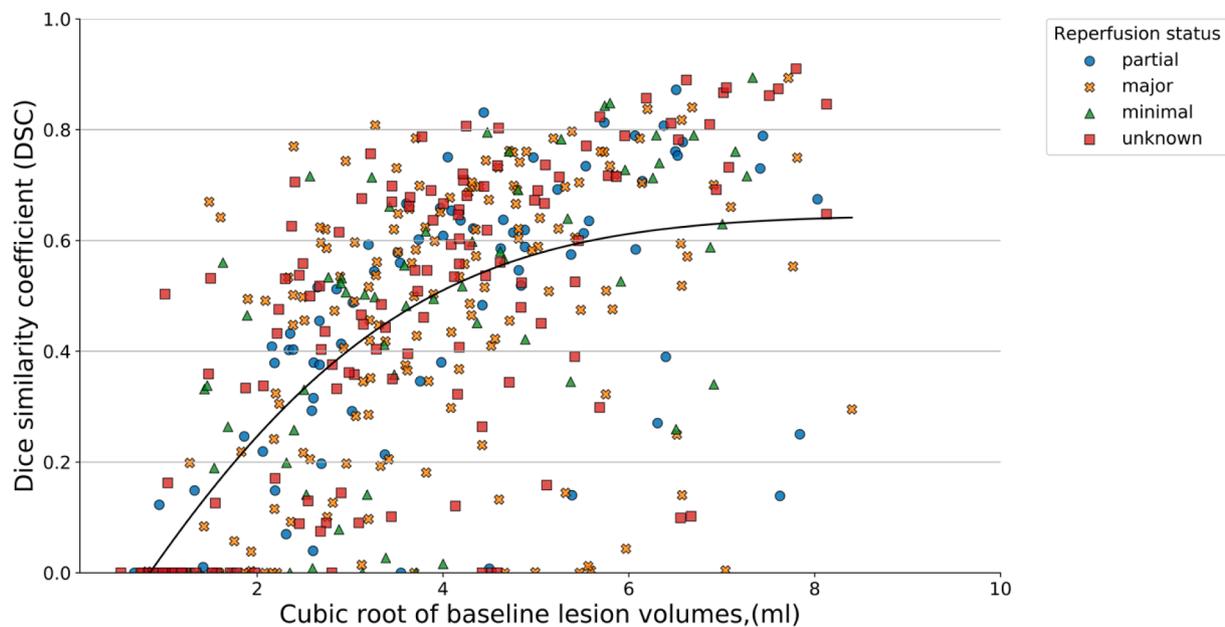
Fig. 6. The correlation between the DSC and cubic-root baseline lesion volumes ($\rho = 0.64$, $p < 0.01$). The best fit to an exponential function is indicated as a black curve.

(Wheeler et al., 2015). Considering this, the model could be pretrained on partial and unknown reperfusion groups and fine-tuned separately for minimal and major reperfusion groups in future investigations if the goal is to predict best and worst case outcomes from endovascular therapy, as in a recent study that included PWI information (Yu et al., 2021).

The DCNN model performed significantly better than simple ADC thresholding. As Fig. 3 shows, the DCNN model was more robust to false positives and false negatives, resulting in higher DSC (Table 4). The DCNN model produced fewer false negatives than the ADC threshold method, which confirms that the DCNN model does not just segment the baseline ADC lesions, but improves its prediction of the final lesion size and location. Moreover, the DCNN produced fewer false positives, indicating it was not susceptible to low-intensity ADC image artifacts that can affect thresholding methods relying solely on ADC intensity. It is possible that the model performs better than simple ADC thresholding by identifying subthreshold ADC decreases that may extend into the classically defined penumbra or patterns of ADC decrease within the visible regions that may provide information about growth patterns (Hevia Montiel et al., 2008; Oppenheim et al., 2001). Prior non-DL models such as the "region-growing principal" method have been previously described (Hevia Montiel et al., 2008; Rosso et al., 2009), with similar performance for predicting final lesion volume, but have not proven to be robust enough for clinical use. Finally, model performance was highly correlated with baseline lesion size. This was an expected finding which was also reported by previous studies that investigated stroke lesion prediction and/or segmentation (Nazari-Farsani et al., 2020; Perez Malla et al., 2019; Yu et al., 2020) and reflects the fundamental challenge of predicting very small lesions accurately. This implies that these methods could be used with more confidence in patients with larger baseline lesions.

We examined the largest over- and underestimations between the predictions and the ground truth. Most model underestimations occurred in the minimal and unknown reperfusion group, which makes sense given the expectation that the lesions would tend to grow more with poor reperfusion. Similarly, that overestimation of the model was primarily seen in subjects with major reperfusion group was not also surprising. It was difficult to assess the precise reason for the remainder of the outliers in the partial and unknown reperfusion groups, as they might have been affected by confounding factors such as reperfusion or collateral status of the patients. However, a concrete conclusion cannot be made as we do not have access to precise reperfusion or collateral flow information.

There are several limitations to this study. A limited number of patients was used to train and test our model. Although this dataset was gathered from multiple institutions with different scanners and inclusion of patients with a variety of clinical data, it is possible that the results may not generalize to specific stroke cohorts, particularly those composed of primarily non-LVO strokes. Second, the model in this study was trained only on DWI (and ADC images derived from DWI). We did investigate the effect of adding magnetic resonance angiography (MRA) images as an extra input channel to the model, but the model performance did not improve. We also tried combining the image data with clinical information including stroke sidedness, but again, this did not improve performance, possibly suggesting that information about stroke sidedness and clinical severity is already encoded to some extent in the DWI lesion's location and severity. Lastly, we did not attempt to create PWI maps from the DWI maps using image translation methods (Yu et al., 2022). This might improve performance and potentially help explaining lesion growth in some patients. Given the black-box nature of most AI systems (including this one), a method that helped explain why some lesions are predicted to grow while others are not would be valuable to inspire confidence and improve adoption of these types of models.

5. Conclusion

Using only baseline DWI without PWI, deep learning can predict final infarction volume with relatively good accuracy in stroke patients. Avoiding the need for PWI to assess stroke patients' final lesion volume may yield multiple benefits, among them shorter imaging studies and faster patient triage times.

6. Source of funding

7. Disclosure

Dr. Greg Zaharchuk reported receiving research support from GE Healthcare and Bayer AG, non-financial support from Nvidia Corporation, being on the scientific advisory board for Biogen, and being a cofounder in Subtle Medical, Inc, outside the submitted work. Dr. David Liebeskind is a consultant as Imaging Core Lab for Cerenovus, Genentech, Medtronic, Stryker, Rapid Medical. Dr. Gregory Albers is a consultant for and has equity in iSchemaView. Other co-authors declare no conflict of interest.

CRediT authorship contribution statement

Sanaz Nazari-Farsani: Writing – original draft, Writing – review & editing, Methodology, Software, Validation, Formal analysis, Visualization. **Yannan Yu:** Writing – original draft, Writing – review & editing, Methodology, Software, Visualization. **Rui Duarte Armindo:** Writing – original draft, Writing – review & editing, Data curation. **Maarten Lansberg:** Writing – original draft, Writing – review & editing, Data curation. **David S. Liebeskind:** Writing – original draft, Writing – review & editing, Data curation. **Gregory Albers:** Writing – original draft, Writing – review & editing, Data curation. **Soren Christensen:** Writing – original draft, Writing – review & editing, Data curation. **Craig S. Levin:** Writing – original draft, Writing – review & editing, Funding acquisition. **Greg Zaharchuk:** Conceptualization, Resources, Writing – original draft, Writing – review & editing, Funding acquisition, Supervision, Data curation, Investigation.

Declaration of Competing Interest

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

Data availability

The data that has been used is confidential.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nicl.2022.103278.

References

Bernal, J., Kushibar, K., Asfaw, D.S., Valverde, S., Oliver, A., Martí, R., Lladó, X., 2019. Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. Artif. Intell. Med. 95, 64–81.

Bivard, A., Levi, C., Spratt, N., Parsons, M., 2013. Perfusion CT in Acute Stroke: A Comprehensive Analysis of Infarct and Penumbra. Radiology 267, 543–550. https://doi.org/10.1148/radiol.12120971.

Cheng, B., Knaack, C., Forkert, N.D., Schnabel, R., Gerloff, C., Thomalla, G., 2017. Stroke subtype classification by geometrical descriptors of lesion shape. PLoS One 12, e0185063.

Choi, Y., Kwon, Y., Lee, H., Kim, B.J., Paik, M.C., Won, J.-H., 2016. Ensemble of Deep Convolutional Neural Networks for Prognosis of Ischemic Stroke. In: Crimi, A., Menze, B., Maier, O., Reyes, M., Winzeck, S., Handels, H. (Eds.), Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp. 231–243.

Gillmann, C., Peter, L., Schmidt, C., Saur, D., Scheuermann, G., 2021. Visualizing Multimodal Deep Learning for Lesion Prediction. IEEE Comput. Graph. Appl. 41, 90–98. https://doi.org/10.1109/MCG.2021.3099881.

Hakim, A., Christensen, S., Winzeck, S., Lansberg, M.G., Parsons, M.W., Lucas, C., Robben, D., Wiest, R., Reyes, M., Zaharchuk, G., 2021. Predicting infarct core from computed tomography perfusion in acute ischemia with machine learning: lessons from the ISLES challenge. Stroke 52, 2328–2337.

Hevia Montiel, N., Rosso, C., Chupin, N., Deltour, S., Bardinet, E., Dormont, D., Samson, Y., Baillet, S., 2008. Automatic Prediction of Infarct Growth in Acute Ischemic Stroke from MR Apparent Diffusion Coefficient Maps. Acad. Radiol. 15, 77–83. https://doi.org/10.1016/j.acra.2007.07.007.

High, W.A., Ayers, R.A., Chandler, J., Zito, G., Cowper, S.E., 2007. Gadolinium is detectable within the tissue of patients with nephrogenic systemic fibrosis. J. Am. Acad. Dermatol. 56, 21–26.

Karthik, R., Gupta, U., Jha, A., Rajalakshmi, R., Menaka, R., 2019. A deep supervised approach for ischemic lesion segmentation from multimodal MRI using Fully Convolutional Network. Appl. Soft Comput. 84, 105685 https://doi.org/10.1016/j.asoc.2019.105685.

Kijowski, R., Liu, F., Caliva, F., Pedoia, V., 2020. Deep Learning for Lesion Detection, Progression, and Prediction of Musculoskeletal Disease. J. Magn. Reson. Imaging 52, 1607–1619. https://doi.org/10.1002/jmri.27001.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kuo, P.H., Kanal, E., Abu-Alfa, A.K., Cowper, S.E., 2007. Gadolinium-based MR Contrast Agents and Nephrogenic Systemic Fibrosis. Radiology 242, 647–649. https://doi.org/10.1148/radiol.2423061640.

Lansberg, M.G., Straka, M., Kemp, S., Mlynash, M., Wechsler, L.R., Jovin, T.G., Wilder, M.J., Lutsep, H.L., Czartoski, T.J., Bernstein, R.A., Chang, C.W.J., Warach, S., Fazekas, F., Inoue, M., Tipirneni, A., Hamilton, S.A., Zaharchuk, G., Marks, M.P., Bammer, R., Albers, G.W., 2012. MRI profile and response to endovascular reperfusion after stroke (DEFUSE 2): a prospective cohort study. Lancet Neurol. 11, 860–867. https://doi.org/10.1016/S1474-4422(12)70203-X.

Lucas, C., Kemmling, A., Bouteldja, N., Aulmann, L.F., Madany Mamlouk, A., Heinrich, M.P., 2018. Learning to Predict Ischemic Stroke Growth on Acute CT Perfusion Data by Interpolating Low-Dimensional Shape Representations. Front. Neurol. 9, 989. https://doi.org/10.3389/fneur.2018.00989.

Mathews, C., Mohamed, A., 2022. Nested U-Net with Enhanced Attention Gate and Compound Loss for Semantic Segmentation of Brain Tumor from Multimodal MRI. Int. J. Intell. Eng. Syst.

Nazari-Farsani, S., Nyman, M., Karjalainen, T., Bucci, M., Isojärvi, J., Nummenmaa, L., 2020. Automated segmentation of acute stroke lesions using a data-driven anomaly detection on diffusion weighted MRI. J. Neurosci. Methods 333, 108575.

Nighoghossian, N., Hermier, M., Adeleine, P., Derex, L., Dugor, J.F., Philippeau, F., Ylmaz, H., Honnorat, J., Dardel, P., Berthezène, Y., Froment, J.C., Trouillas, P., 2003. Baseline Magnetic Resonance Imaging Parameters and Stroke Outcome in Patients Treated by Intravenous Tissue Plasminogen Activator. Stroke 34, 458–463. https://doi.org/10.1161/01.STR.0000053850.64877.AF.

Ogata, T., Christensen, S., Nagakane, Y., Ma, H., Campbell, B.C.V., Churilov, L., Lansberg, M.G., Straka, M., De Silva, D.A., Mlynash, M., Bammer, R., Olivot, J.-M., Desmond, P.M., Albers, G.W., Davis, S.M., Donnan, G.A., 2013. The Effects of Alteplase 3 to 6 Hours After Stroke in the EPITHET–DEFUSE Combined Dataset. Stroke 44, 87–93. https://doi.org/10.1161/STROKEAHA.112.668301.

Oktay, O., Schlemper, J., Folgoc, L. le, Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., 2018. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

Oppenheim, C., Grandin, C., Samson, Y., Smith, A., Duprez, T., Marsault, C., Cosnard, G., 2001. Is There an Apparent Diffusion Coefficient Threshold in Predicting Tissue Viability in Hyperacute Stroke? Stroke 32, 2486–2491. https://doi.org/10.1161/hs1101.098331.

Pinto, A., Mckinley, R., Alves, V., Wiest, R., Silva, C.A., Reyes, M., 2018. Stroke Lesion Outcome Prediction Based on MRI Imaging Combined With Clinical Information. Front. Neurol. 9, 1060. https://doi.org/10.3389/fneur.2018.01060.

Robben, D., Boers, A.M.M., Marquering, H.A., Langezaal, L.L.C.M., Roos, Y.B., van Oostenbrugge, R.J., van Zwam, W.H., Dippel, D.W.J., Majoie, C.B.L.M., van der Lugt, A., 2020. Prediction of final infarct volume from native CT perfusion and treatment parameters using deep learning. Med. Image Anal. 59, 101589.

Rosso, C., Hevia-Montiel, N., Deltour, S., Bardinet, E., Dormont, D., Crozier, S., Baillet, S., Samson, Y., 2009. Prediction of Infarct Growth Based on Apparent Diffusion Coefficients: Penumbral Assessment without Intravenous Contrast Material. Radiology 250, 184–192. https://doi.org/10.1148/radiol.2493080107.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Thamm, T., Guo, J., Rosenberg, J., Liang, T., Marks, M.P., Christensen, S., Do, H.M., Kemp, S.M., Adair, E., Eyngorn, I., Mlynash, M., Jovin, T.G., Keogh, B.P., Chen, H.J., Lansberg, M.G., Albers, G.W., Zaharchuk, G., 2019. Contralateral Hemispheric Cerebral Blood Flow Measured With Arterial Spin Labeling Can Predict Outcome in Acute Stroke. Stroke 50, 3408–3415. https://doi.org/10.1161/STROKEAHA.119.026499.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A. pietro, Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C.N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D.A., Hagen, D.R., Pasechnik, D. v, Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G.A., Ingold, G.-L., Allen, G.E., Lee, G.R., Audren, H., Probst, I., Dietrich, J.P., Silterra, J., Webber, J.T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J.L., de Miranda Cardoso, J.V., Reimer, J., Harrington, J., Rodríguez, J. L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N.J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T.J.,

Robitaille, T.P., Spura, T., Jones, T.R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y.O., Vázquez-Baeza, Y., Contributors, S. 1. 0, 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

Wheeler, H.M., Mlynash, M., Inoue, M., Tipirnini, A., Liggins, J., Bammer, R., Lansberg, M.G., Kemp, S., Zaharchuk, G., Straka, M., Albers, G.W., Investigators, D., 2015. The growth rate of early DWI lesions is highly variable and associated with penumbral salvage and clinical outcomes following endovascular reperfusion. Int. J. Stroke 10, 723–729. https://doi.org/10.1111/ijs.12436.

Xue, Y., Farhat, F.G., Boukrina, O., Barrett, A.M., Binder, J.R., Roshan, U.W., Graves, W. W., 2020. A multi-path 2.5 dimensional convolutional neural network system for segmenting stroke lesions in brain MRI images. Neuroimage Clin. 25, 102118 https://doi.org/10.1016/j.nicl.2019.102118.

Perez Malla, C.U., Valdes Hernandez, M.D.C., Rachmadi, M.F., Komura, T., n.d. Evaluation of Enhanced Learning Techniques for Segmenting Ischaemic Stroke Lesions in Brain Magnetic Resonance Perfusion Images Using a Convolutional Neural Network Scheme. Frontiers in neuroinformatics JID – 101477957 PMC – PMC6548861 OTO – NOTNLM.

Yu, Y., Xie, Y., Thamm, T., Gong, E., Ouyang, J., Huang, C., Christensen, S., Marks, M.P., Lansberg, M.G., Albers, G.W., Zaharchuk, G., 2020. Use of Deep Learning to Predict Final Ischemic Stroke Lesions From Initial Magnetic Resonance Imaging. JAMA Netw. Open 3, e200772–e. https://doi.org/10.1001/jamanetworkopen.2020.0772.

Yu, Y., Xie, Y., Thamm, T., Gong, E., Ouyang, J., Christensen, S., Marks, M.P., Lansberg, M.G., Albers, G.W., Zaharchuk, G., 2021. Tissue at Risk and Ischemic Core Estimation Using Deep Learning in Acute Stroke. Am. J. Neuroradiol. 42, 1030–1037. https://doi.org/10.3174/ajnr.A7081.

Yu, Y., Gong, E., Ouyang, J., Christensen, S., Scalzo, F., Liebeskind, D.S., Lansberg, M.G., Albers, G., Zaharchuk, G., 2022. Abstract 8: Hypoperfusion Lesion And Target Mismatch Prediction In Acute Ischemic Stroke From Baseline Mr Diffusion Imaging Using A 3d Convolutional Neural Network. Stroke 53, A8–A. https://doi.org/10.1161/str.53.suppl_1.8.

Zaharchuk, G., Marks, M.P., Do, H.M., Bammer, R., Lansberg, M., Kemp, S., Albers, G.W., and iCAS Investigators, 2015. Abstract W MP16: Introducing the Imaging the Collaterals in Acute Stroke (iCAS) Multicenter MRI Trial. Stroke 46, AWMP16–AWMP16. https://doi.org/10.1161/str.46.suppl_1.wmp16.