Infarct Core Segmentation on Spectral CT Angiography using a Data-Uncertainty-Aware U-Net

Joris Vromans[1 *], Edwin Bennink[1], Jan Willem Dankbaar[1], Birgitta K. Velthuis[1], Hugo W.A.M. de Jong[1]

[1] Department of Radiology and Nuclear Medicine, University Medical Center Utrecht, P.O. Box 85500, 3508 GA, Utrecht, The Netherlands
j.vromans-2@umcutrecht.nl

Abstract. Spectral CT angiography with deep learning–based infarct core segmentation may aid in thrombectomy eligibility assessment in acute ischemic stroke. However, several inherent sources of data variability, such as acquisition timing and lack of a reliable ground truth, complicate core segmentation.

We therefore developed a data-uncertainty-aware U-Net, trained on iodine density (ID) and non-contrast CT images with aleatoric uncertainty estimation in its loss function, that is robust to uncertain input and noisy training labels. U-Net predictions were compared against relative ID threshold segmentations and CT perfusion (CTP)–derived references, with a radiologist scoring and ranking them across 39 cases.

U-Net segmentations significantly outperformed thresholded relative ID on Likert score (p = 3.4e-4, r = 0.60) and on ranking (p = 6.4e-5, r = 0.67). CTP references were only significantly better when the radiologist ranked the segmentations (p = 2.9e-2, r = 0.35). The Likert scores of the CTP references were not significantly different from those of the U-Net segmentations (p = 9.3e-2, r = 0.27). U-Net ranked first 10 out of 39 cases while the reference ranked first 25 out of 39 cases, showing good generalizability of the U-Net and imperfections of the CTP-derived references. Incorporating uncertainty prevented over-segmentation, leading to better volumetric similarity with the CTP references.

These results suggest a data-uncertainty-aware U-Net applied to spectral CT angiography ID maps could be a valuable alternative when CTP is unavailable.

Keywords: Acute Ischemic Stroke, Spectral CT Angiography, Iodine Density, Infarct Core, Segmentation, Data Uncertainty.

1 Introduction

Acute ischemic stroke (AIS) is a leading cause of morbidity and mortality worldwide, where rapid intervention is critical. Imaging protocols combine non-contrast CT (NCCT), CT angiography (CTA), and CT perfusion (CTP) imaging, with CTP used to distinguish irreversibly damaged infarct core from salvageable penumbra. Recent randomized controlled trials and updated clinical guidelines indicate that eligibility for endovascular thrombectomy may be determined from core volume alone [1-5]. Since CTP carries inherent drawbacks, including increased radiation dose, susceptibility to

motion artifacts, and incomplete cerebral coverage, more attention has been directed towards CTA for core volume estimation [6].

Conventional CTA reliably depicts arterial occlusions and collateral circulation, while dual-layer spectral CTA additionally provides quantitative iodine density (ID) maps that may function as a CTP surrogate. A limitation of single-phase CTA is that it captures only a single time-point at peak arterial enhancement, reducing sensitivity for perfusion abnormalities. Previous studies reported that CTA-based core segmentation generally overestimates core due to acquisition timing [6].

Vromans et al. previously explored ID-based core and penumbra segmentation by applying relative ID (rID) thresholds. While core volumes showed minor bias, large inter-patient variability rendered the rID thresholds insufficient for clinical adaptation. Nonetheless, these results confirmed that ID maps contain discriminative information for core delineation, motivating more advanced modeling [7].

In the current study, we attempted core delineation using a U-Net trained on ID and NCCT images. Several inherent sources of data variability—including CTA acquisition timing differences, motion artifacts, CT noise, and variability in CTP-derived reference segmentations—motivate an uncertainty-aware formulation [8-10]. Inspired by Kendall and Gal, we incorporate aleatoric uncertainty estimation into the loss function, which adds robustness to uncertain input and attenuates the influence of noisy training labels [11]. Beyond potentially improving segmentation performance, this approach could support safer clinical decision-making by identifying voxels where predictions are less reliable.

The aim of this study was (A) to compare ID-based core segmentation using (1) a fixed rID threshold method and (2) a novel data-uncertainty-aware U-Net against CTP-derived reference segmentations, and (B) to explore to what extent a data-uncertainty-aware U-Net can mitigate the effects of data-uncertainty, acknowledging some uncertainty should be dealt with at the time of acquisition.

2 Methods

2.1 Dataset

Patients were adults (>18 years) with confirmed AIS, stroke onset <9 hours, informed consent, and imaging acquired on a Philips iQon Spectral CT scanner (NCCT, CTA, and CTP) from the ENCLOSE study [12]. Following Aludin et al., additional inclusion criteria were: confirmed LVO in the MCA territory (ICA or MCA M1/M2) level); absence of contralateral high-grade (>70%) carotid stenosis; absence of CTP motion or contrast artifacts; absence of hemorrhage, tumor, prior ischemic lesion, or embolism in other territories; and infarct core volume >10 mL on CTP [13]. The ENCLOSE dataset comprised 405 patients, of which 39 fulfilled all inclusion criteria.

2.2 Preprocessing

ID maps were derived from spectral CTA using the Philips Spectral Research Toolkit (version 2.01). NCCT images and ID maps were resampled to 10 mm slice spacing and

registered to a MNI CT brain template. A 3D B-spline transformation was optimized on the conventional CTA acquisition, then applied to the ID maps. NCCT intensities were clipped to 0–80 HU and scaled between 0 and 1. ID maps were normalized per subject by dividing intensities by the 95th percentile intensity.

Following Aludin et al. and Vromans et al., CTP datasets were resampled to 10 mm slice spacing [7, 13]. CTP-derived core masks were generated using IntelliSpace Portal (ISP, advanced visualization workspace version 15, Philips) with standard thresholds (CBV <2.0mL/100g; rMTT >150%) [14]. These masks served as pragmatic reference standards, acknowledging inter-algorithm variability in CTP-derived core delineation. Core masks were registered to the CT brain template using the first CTP time frame as reference.

For all images, left hemispheres were mirrored onto the right to facilitate learning of ID asymmetries. Pre-processed slices had a size of 128 by 256 voxels.

2.3 Network architecture

A 2.5D U-Net was used, motivated by anisotropic voxel spacing (10 mm slice vs. 1 mm in-plane) [15]. The network had 12 input channels, representing the center and adjacent slices of the NCCT and ID map for each hemisphere; with fixed locations for the infarcted and contralateral hemispheres. The number of feature maps increased from 32 (shallowest) to 256 (deepest) in three image downsampling steps, doubling each step. The network was designed to output a core segmentation (μ, with sigmoid activation) as well as an uncertainty map ($\ln(\sigma^2)$).

2.4 Loss function

The gradual degradation of ischemic tissue and noisy labels motivated treating segmentation as a regression- rather than a binary problem. Kendall and Gal's Gaussian negative log-likelihood (NLL_G) approach (1) was adopted and inspired us to introduce uncertainty weighting within the class imbalance robust Soft Dice loss (2) [11]:

$$NLL_G = \frac{1}{BN}\sum_{b=1}^{B}\sum_{n=1}^{N}\left(\frac{(\mu_{b,n}-y_{b,n})^2}{2\sigma_{b,n}^2} + \frac{1}{2}\ln\left(\sigma_{b,n}^2\right)\right) \quad (1)$$

$$Soft\ Dice\ loss = 1 - \frac{1}{B}\sum_{b=1}^{B}\frac{2\cdot\sum_{n=1}^{N}\frac{\mu_{b,n}\cdot y_{b,n}}{\sigma_{b,n}^2}}{\sum_{n=1}^{N}\frac{\mu_{b,n}+y_{b,n}}{\sigma_{b,n}^2}} + \frac{\lambda}{BN}\sum_{b=1}^{B}\sum_{n=1}^{N}\ln\left(\sigma_{b,n}^2\right)^2 \quad (2)$$

With predicted probability (μ), target (y), learned uncertainty ($\ln(\sigma^2)$), regularization factor (λ =1e-3), batch size (B), and number of voxels (N).

The effect of constraining the range of the uncertainty term ($\ln(\sigma^2)$) was investigated across six settings, from no uncertainty contribution ($\sigma^2 = 1$, $\ln(\sigma^2) = 0$) to broad limits with strong uncertainty influence ($-5 \leq \ln(\sigma^2) \leq 5$) with a step size of 1.

2.5 Training

The network was trained once for each case and uncertainty bound combination. In each training session, one case served as test, five as validation (selected at evenly spaced positions within the 10th–90th percentile of the reference core volume distribution), and the remaining 33 as training. Training began at learning rate 1e-4. Each iteration, a batch of 32 random samples was augmented (torchvision v0.27.1 - ElasticTransform: alpha = 50, sigma = 5, interpolation = bilinear). After each Adam optimizer update, validation loss was evaluated and the best-performing weights retained. The learning rate was multiplied by 0.5 after 16 consecutive non-improving iterations until the stopping criteria of 1e-6.

2.6 Uncertainty-based thresholding

The core threshold ($\mu - k \cdot \sigma_{norm} \geq 0.5$) follows from the constraint that a voxel with maximum probability ($\mu_{max} = 1$) at neutral uncertainty ($\sigma_{norm} = 1$) falls exactly on the decision boundary by defining k as $\frac{\mu_{max} - \mu_{boundary}}{\sigma_{neutral}} = 0.5$ with $\mu_{boundary}$ at 0.5. Uncertainty was normalized to compensate for bounding according to: $\sigma_{norm} = \frac{\sigma_{out} - \sigma_{min}}{\sigma_{neutral} - \sigma_{min}}$ with σ_{out} the predicted uncertainty, σ_{min} the minimum possible uncertainty within the uncertainty bound, and $\sigma_{neutral}$ the value at which σ leaves a voxel's data-fit contribution unchanged relative to the unweighted loss, i.e. equals 1. By incorporating uncertainty, the threshold excluded voxels that were likely false positives.

2.7 Metrics

Volumetric similarity (VS) with the CTP-derived reference ($1 - \frac{|V_{pred} - V_{ref}|}{V_{pred} + V_{ref}}$) was the primary performance metric, reflecting the importance of core volume in treatment decisions. The Dice score ($\frac{2 \cdot |pred \cap ref|}{|pred| + |ref|}$) was used as a secondary overlap metric.

2.8 Optimal uncertainty bound selection

For all six uncertainty bounds the mean validation VS was calculated. The bound with highest mean VS was used to predict core segmentations on corresponding test cases.

2.9 Evaluation

Whole-brain predictions were assessed using Dice, VS, Pearson correlation, volumetric bias, and limits of agreement with the CTP-derived reference. rID thresholding served as the baseline [7], with the optimal threshold determined by maximizing VS on all 39 cases.

Qualitative assessment was performed by a radiologist with >15 years of AIS experience. For each case, two rows of scrollable axial slices were presented: (top row)

CBV, CBF, NCCT, and ID map slices; (bottom row) U-Net, threshold-based, and CTP-derived core segmentations overlaid on CBV. CBV and CBF maps were generated with an in-house CTP processing package, which differed from ISP, to reduce bias toward the CTP reference [16]. Method labels were not shown and segmentation order was randomized across cases; the radiologist assigned each segmentation a rank and a Likert score using the following scale:

- Strongly agree: Segmentation represents core volume well; no edits required.
- Agree: Acceptable representation; minor edits within inter-observer variability would achieve strong agreement.
- Neither agree nor disagree: Minor but necessary edits required; these likely exceed inter-observer variability.
- Disagree: Major edits required.
- Strongly disagree: Unusable.

2.10 Statistical analysis

Differences in quantitative metrics between U-Net and thresholding were assessed by paired bootstrapped 95%-confidence intervals (CIs; 10,000 iterations). Likert scores and rankings were analyzed using the Friedman test; significant overall effects were followed by pairwise Wilcoxon signed-rank tests with Holm correction.

3 Results

3.1 Dataset

The ENCLOSE dataset comprised 405 patients, of which 39 fulfilled all predefined inclusion criteria [12]. An overview of the selection process is provided in Figure 1.

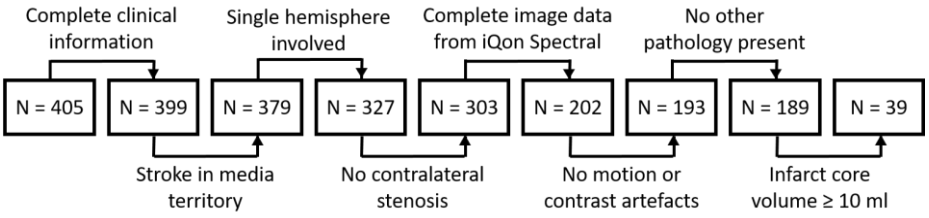

Fig. 1. Flow diagram illustrating the application of the dataset inclusion criteria, resulting in a final study cohort of 39 patients. Reproduced from Vromans et al. [7].

3.2 Optimal uncertainty bound

Mean VS with the CTP-derived infarct core reference on the validation sets for each uncertainty bound are shown in Table 1. The highest validation VS (0.75) was achieved for uncertainty bound [-2, 2].

Table 1. The mean volumetric similarity (VS) with the CTP-derived infarct core for the six uncertainty bounds computed from the validation sets.

$\ln(\sigma^2)$	0	[-1, 1]	[-2, 2]	[-3, 3]	[-4, 4]	[-5, 5]
VS	0.53	0.66	**0.75**	0.71	0.61	0.52

3.3 Predicted test-set infarct core volumes by method

Test-set core volume distributions by method, including all six uncertainty bounds, are shown in Figure 2. U-Net predictions followed a descending trend: greater uncertainty freedom yielded smaller, more conservative segmentations. The optimal threshold for the threshold-based baseline was a rID of 0.32.

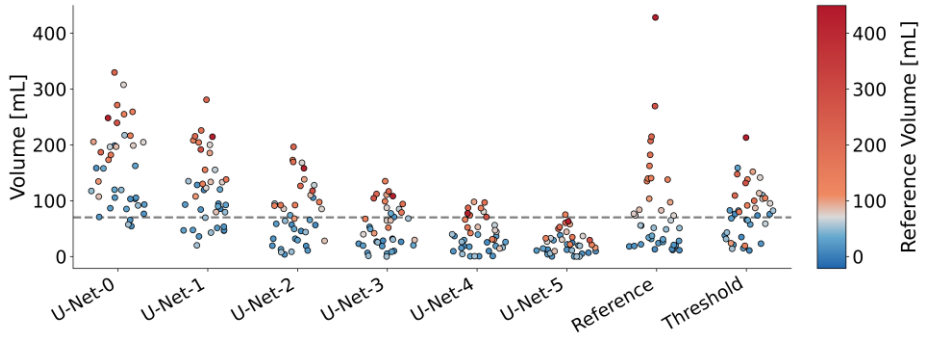

Fig. 2. Predicted test-set infarct core volumes by method. Uncertainty freedom increased from U-Net-0 (no uncertainty contribution ($\sigma^2 = 1$, $\ln(\sigma^2) = 0$)) to U-Net-5 (broad limits with strong uncertainty influence ($-5 \leq \ln(\sigma^2) \leq 5$)). U-Net volumes showed a descending trend where the more uncertainty freedom the U-Net had, the smaller the volumes became. The reference line at 70 mL represents the DEFUSE-3 threshold for endovascular thrombectomy eligibility [2].

3.4 Quantitative analysis

Uncertainty bound [-2, 2] yielded the highest test-set VS (0.70) and a Dice score of 0.36 (Table 2). The mean Dice score difference of 0.12 (CI: [0.08, 0.16]; p =1.5e-6) between U-Net-2 and threshold-based segmentations was statistically significant.

Table 2. The mean volumetric similarity (VS), Dice score, Pearson correlation coefficient (r) with p-value, volumetric bias (B), and limits of agreement (LA) with the CT perfusion-derived infarct core for each method on the test set.

Method	U-Net-0	U-Net-1	U-Net-2	U-Net-3	U-Net-4	U-Net-5	Threshold
VS	0.57	0.67	**0.70**	0.68	0.61	0.48	0.68
Dice	0.37	**0.38**	0.36	0.33	0.29	0.23	0.25
r	0.64	**0.73**	0.67	0.69	0.67	0.66	0.62
p-value	1.2e-5	**1.6e-7**	3.6e-6	9.7e-7	3.0e-6	5.6e-6	2.1e-5
B [mL]	76	35	-7	-31	-46	-60	**-4**
LA [mL]	±134	±**116**	±126	±129	±137	±146	±132

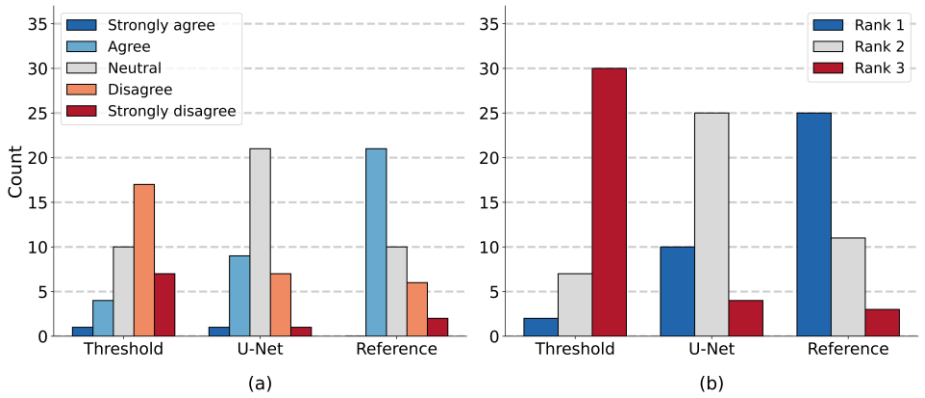

Fig. 3. Qualitative evaluation of the three segmentation methods by the radiologist. (a) Distribution of Likert-scale ratings, with CTP segmentations scoring slightly higher than U-Net, and both outperforming thresholding. (b) Distribution of assigned ranks (1 = best to 3 = worst), with CTP segmentations ranking best, followed by U-Net, and thresholding ranking worst.

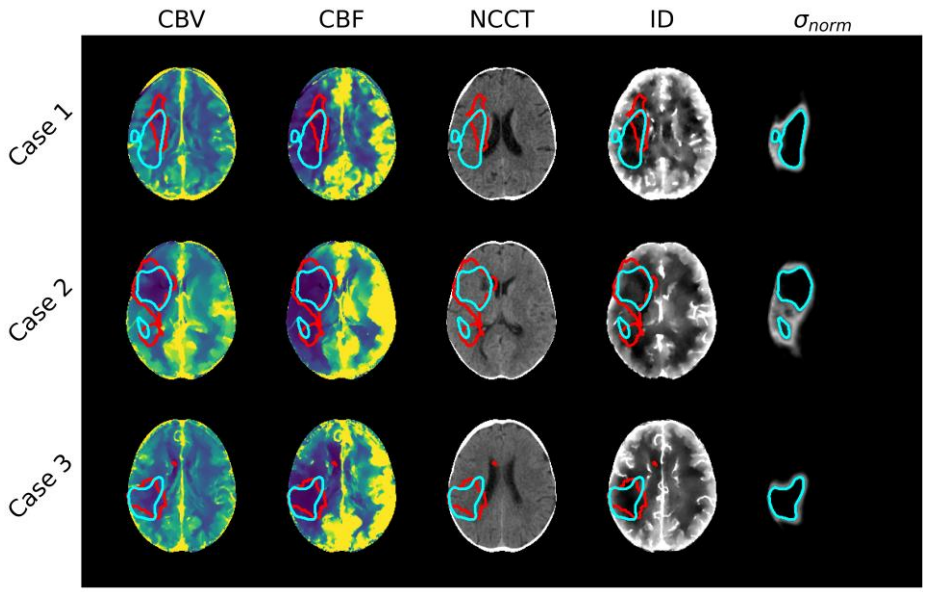

Fig. 4. Cerebral blood volume (CBV), cerebral blood flow (CBF), non-contrast CT (NCCT), iodine density (ID), and uncertainty (σ_{norm}, visualization range 0 to 1) maps for three different subjects with the reference (red) and U-Net (cyan) segmentations overlayed. A large deviation from reference can be seen in case 1 where the segmented ischemic defect in the ID map more represents the CBF defect. Case 2 shows uncertainty of the network for the CBF defect. Case 3 shows a good match between CBV and ID defect and good correspondence between the prediction and reference.

3.5 Qualitative analysis

U-Net-2 scored significantly better than rID-thresholding (p = 3.4e-4, r = 0.60) on the Likert score and not significantly worse than CTP (p = 9.3e-2, r = 0.27). U-Net ranking was significantly worse than CTP (p = 2.9e-2, r = 0.35) but significantly better than rID-thresholding (p = 6.4e-5, r = 0.67) (Figure 3).

Figure 4 shows representative cases illustrating the correspondence between predicted and reference segmentations relative to the underlying CTP and ID maps.

4 Discussion

The uncertainty-aware U-Net infarct core segmentations with uncertainty bound [-2, 2] significantly outperformed rID-threshold segmentations (p = 3.4e-4, r = 0.60) on the Likert score and ranking (p = 6.4e-5, r = 0.67) with large effects. CTP segmentations were only significantly better when the radiologist ranked the methods (p = 2.9e-2, r = 0.35) with a medium effect. On the Likert score, the CTP segmentations were not significantly better than U-Net segmentations (p = 9.3e-2, r = 0.27). U-Net ranked first 10 out of 39 times while the reference ranked first 25 out of 39 times, showing good generalizability and imperfections of the CTP-derived reference. These are promising results for core segmentation on spectral CTA ID maps. However, visualization of the predictions overlayed on ID- and CTP parameter maps illustrated the challenge of core segmentation on CTA: likely due to CTA acquisition timing, the visible perfusion defect may more closely resemble a CBF rather than a CBV defect, as shown in Figure 4 [6]. This mismatch is not addressable at the image analysis stage but must instead be addressed at the time of image acquisition through protocol optimization (e.g. an additional early bolus to saturate the collateral vessels or an acquisition between the mid-arterial and late venous phase) as previous studies have also concluded [6].

Likert scores indicated that most U-Net segmentations required minor edits that fell outside of inter-observer variability, while for CTP the required edits fell mostly within inter-observer variability. Minor required edits may also have minor effect on total volumetric accuracy.

Quantitatively, U-Net with uncertainty bound [-2, 2] significantly improved Dice score over rID-thresholding (difference: 0.12; CI: [0.08, 0.16]; p = 1.5e-6). VS for U-Net and rID-threshold segmentations were similar, 0.70 and 0.68 respectively. However, superior spatial overlap carries direct clinical relevance beyond volumetric agreement: accurate localization of the infarct is useful for risk stratification and treatment planning. Nevertheless, all quantitative metrics should be interpreted in light of the imperfect nature of the CTP-derived reference standard.

Evaluation of the six uncertainty bound settings demonstrated that the degree of uncertainty regularization meaningfully influences segmentation performance and careful tuning counters over-segmentation as seen by U-Net-0 without uncertainty regularization and other CTA-based core segmentation studies [6].

In conclusion, this study provides proof-of-concept that spectral CTA ID maps, analyzed with a data-uncertainty-aware U-Net, support automated core segmentation of quality comparable to CTP-derived references. These results motivate further

investigation in larger, multicenter cohorts to establish the clinical utility of core segmentations from a data-uncertainty-aware network on spectral CTA as a clinically valuable alternative when CTP is not available. Future work should additionally examine whether tailoring the CTA acquisition protocol can reduce the performance gap with CTP. Moreover, extending the uncertainty framework to include epistemic uncertainty (e.g., variation in retrained model parameters) would provide a complementary confidence measure to further support safe integration of learned uncertainty maps into clinical decision workflows.

Acknowledgments. This work was supported by Health Holland (LSH) [TKI2211].

Disclosure of Interests. The authors declare no competing interests.

References

1. Nogueira, R. G. et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. The New. Engl. J. Medicine (2018).
2. Albers, G. W. et al. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. The New. Engl. J. Medicine (2018).
3. Olthuis, S. G. H. et al. Endovascular treatment versus no endovascular treatment after 6–24 h in patients with ischaemic stroke and collateral flow on CT angiography. The Lancet (2023).
4. Dong, Z. et al. Simplified stroke imaging selection modality for endovascular thrombectomy in the extended time window: systematic review and meta-analysis. J Neurointerv Surg (2023).
5. Powers, W. J. et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke (2019).
6. Hokkinen, L., Mäkelä, T., Savolainen, S. & Kangasniemi, M. Factors influencing the reliability of a CT angiography-based deep learning method for infarct volume estimation. BJR Open (2024).
7. Vromans, J., Bennink, E., Dankbaar, J.W., Velthuis, B.K., de Jong, H.W.A.M. Assessment of threshold-based infarct core and penumbra segmentations from spectral CT angiography iodine density maps. Scientific Reports (2026).
8. Kamalian, S. et al. CT cerebral blood flow maps optimally correlate with admission diffusion-weighted imaging in acute stroke but thresholds vary by postprocessing platform. Stroke (2011).
9. Koopman, M. S. et al. Comparison of three commonly used CT perfusion software packages in patients with acute ischemic stroke. J Neurointerv Surg (2019).
10. Kim, N. et al. Optimal cerebral blood flow thresholds for ischemic core estimation using computed tomography perfusion and diffusion-weighted imaging. Ann Neurol (2024).
11. Kendall, A. & Gal, Y., A. What uncertainties do we need in Bayesian deep learning for computer vision? NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems (2017).
12. Kauw, F. et al. Early detection of small volume stroke and thromboembolic sources with computed tomography: rationale and design of the ENCLOSE study. Eur Stroke J (2020).

13. Aludin, S. et al. Spectral imaging and analysis of monophasic CT angiography to assess infarct core and penumbra in acute stroke. Scientific Reports (2024).
14. Wintermark, M. et al. Perfusion-CT assessment of infarct core and penumbra: receiver operating characteristic curve analysis in 130 patients suspected of acute hemispheric stroke. Stroke (2006).
15. Ronneberger, O., Fischer, P., Brox, T. U-Net: convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science (2015).
16. Bennink, E., Oosterbroek, J., Kudo, K., Viergever, M.A., Velthuis, B.K., de Jong, H.W.A.M. Fast nonlinear regression method for CT brain perfusion analysis. J Med Imaging (Bellingham) (2016).