Unsupervised Deep Learning for Stroke Lesion Segmentation on Follow-up CT Based on Generative Adversarial Networks

H. van Voorst, P.R. Konduji, L.M. van Poppel, W. van der Steen, P.M. van der Sluijs, E.M.H. Slot, B.J. Emmer, W.H. van Zwan, Y.B.W.E.M. Roos, C.B.L.M. Majoie, G. Zaharchuk, M.W.A. Caan, and H.A. Marquering, on behalf of the CONTRAST Consortium Collaborators

ABSTRACT

BACKGROUND AND PURPOSE: Supervised deep learning is the state-of-the-art method for stroke lesion segmentation on NCCT. Supervised methods require manual lesion annotations for model development, while unsupervised deep learning methods such as generative adversarial networks do not. The aim of this study was to develop and evaluate a generative adversarial network to segment infarct and hemorrhagic stroke lesions on follow-up NCCT scans.

MATERIALS AND METHODS: Training data consisted of 820 patients with baseline and follow-up NCCT from 3 Dutch acute ischemic stroke trials. A generative adversarial network was optimized to transform a follow-up scan with a lesion to a generated baseline scan without a lesion by generating a difference map that was subtracted from the follow-up scan. The generated difference map was used to automatically extract lesion segmentations. Segmentation of primary hemorrhagic lesions, hemorrhagic transformation of ischemic stroke, and 24-hour and 1-week follow-up infarct lesions were evaluated relative to expert annotations with the Dice similarity coefficient, Bland-Altman analysis, and intraclass correlation coefficient.

RESULTS: The median Dice similarity coefficient was 0.31 (interquartile range, 0.08–0.59) and 0.59 (interquartile range, 0.29–0.74) for the 24-hour and 1-week infarct lesions, respectively. A much lower Dice similarity coefficient was measured for hemorrhagic transformation (median, 0.02; interquartile range, 0–0.14) and primary hemorrhage lesions (median, 0.08; interquartile range, 0.01–0.35). Predicted lesion volume and the intraclass correlation coefficient were good for the 24-hour (bias, 3 mL; limits of agreement, –64–59 mL; intraclass correlation coefficient, 0.83; 95% CI, 0.78–0.88) and excellent for the 1-week (bias, –4 mL; limits of agreement, –66–58 mL; intraclass correlation coefficient, 0.90; 95% CI, 0.83–0.93) follow-up infarct lesions.

CONCLUSIONS: An unsupervised generative adversarial network can be used to obtain automated infarct lesion segmentations with a moderate Dice similarity coefficient and good volumetric correspondence.

ABBREVIATIONS: AIS = acute ischemic stroke; BL = baseline; DSC = Dice similarity coefficient; FU = follow-up; FUBL-GAN = follow-up to baseline generative adversarial network; GAN = generative adversarial network; 24H = 24 hours; HT = hemorrhagic transformation; ICC = intraclass correlation coefficient; IQR = interquartile range; L1-loss = voxelwise absolute difference between generated baseline and real baseline NCCTs; L1+adv = L1 and adversarial loss; LoA = limits of agreement; nnUnet = no new Unet; PrH = primary hemorrhagic lesions; T1W = 1 week

Hemorrhagic transformation (HT) and malignant cerebral edema are severe complications after acute ischemic stroke (AIS), which frequently result in functional deterioration and death.1–3 Computer-guided visualization and segmentation of these hemorrhagic and infarct lesions can assist radiologists in detecting small lesions.4,5 Furthermore, lesion volume computed from a segmentation predicts

This study was funded by the CONTRAST consortium. The CONTRAST consortium acknowledges the support from the Netherlands Cardiovascular Research Initiative, an initiative of the Dutch Heart Foundation (CVON2015-01: CONTRAST) and from the Brain Foundation of the Netherlands (HA2015.01.06). The collaboration project is additionally financed by the Ministry of Economic Affairs by means of the Public-private partnerships Allowance made available by the Top Sector Life Sciences & Health to stimulate public-private partnerships (LSHON7016). This work was funded, in part, through unrestricted funding by Stryker, Medtronic, and Cerenovus.

Received January 27, 2022; accepted after revision June 2.

From the Departments of Radiology and Nuclear Medicine (H.v.V., P.R.K., L.M.v.P., B.J.E., C.B.L.M., H.A.M.), Biomedical Engineering and Physics (H.v.V., P.R.K., L.M.v.P., M.W.A.C., H.A.M.), and Neurology (Y.B.W.E.M.R.), Faculty of Medicine, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, the Netherlands; Departments of Neurology (W.v.d.S., P.M.v.d.S.), Radiology and Nuclear Medicine (W.v.d.S., P.M.v.d.S.), Erasmus University Medical Center, Rotterdam, the Netherlands; Department of Neurology and Neurosurgery (E.M.H.S.), University Medical Center Utrecht, Utrecht, the Netherlands; Department of Radiology and Nuclear Medicine (W.H.v.Z.), Maastricht University Medical Center, Maastricht, the Netherlands; and Department of Radiology (G.Z.), Stanford University, Stanford, California.

The funding sources were not involved in study design, monitoring, data collection, statistical analyses, interpretation of results, or manuscript writing.
The aim of this study was to accurately segment stroke lesions on follow-up NCCT scans with a GAN trained in an unsupervised manner. In line with Baumgartner et al., we developed a GAN to remove hemorrhagic and ischemic stroke lesions from follow-up NCCT scans by generating difference maps with a lesion and BL NCCT scans without a lesion.

**MATERIALS AND METHODS**

**GANs for BL NCCT Generation**

The GAN structure as adopted in this study consists of 2 competing deep learning models, referred to as generator and discriminator models. The generator model generates artificial images, while the discriminator model tries to distinguish the generated artificial image from the original images. In this study, the generator receives as input a follow-up NCCT scan with the lesion and generates a difference map. This difference map is subtracted from the input follow-up NCCT scan with the lesion to generate an artificial BL NCCT scan without the lesion. Because an infarct lesion is visually subtle in AIS BL NCCTs acquired in the acute stage (0–6 hours after symptom onset), the transformation from a follow-up scan at 24 hours (24H) or 1 week (1W) with a well-defined lesion to a BL scan entails essentially the removal of the lesion. Subsequently, the discriminator model classifies the presented images as being either an original BL or a generated BL NCCT. This classification is used to provide feedback to the generator model and to optimize the difference map.

The proposed GAN method is optimized with 2 types of loss functions: the voxelwise absolute difference between generated BL and real BL NCCTs (L1-loss) and the binary cross-entropy of the discriminator (adversarial-loss) for classifying generated and real BL NCCTs. Figure 1 presents the GAN model architecture we refer to as the follow-up to BL GAN (FU2BL-GAN).

**Patient Populations**

In this study, 820 patients were included between January 2018 and July 2021 in the training data set from the MR CLEAN-NOTIV ($n = 297$), MR CLEAN-MED ($n = 377$), and MR CLEAN-...
after endovascular treatment or randomization. Validation and test acquired 8
the training data files of randomized controlled trials and the validation
received for the use of data for substudies from patients included in
compliance with the declaration of Helsinki, informed consent has been
exclusion criteria of each of these randomized controlled trials have
transformation lesions after AIS (HT; N validation: 19; N test: 57);20
and primary hemorrhagic lesions (PrH; N validation: 11; N test: 24). The data from
LATE (n = 146) randomized controlled trials if BL and follow-up
NCCTs were available. Specific imaging protocols, inclusion, and
exclusion criteria of each of these randomized controlled trials have
been published previously.15-17 Scans of these 820 patients were
used to train the FU2BL-GAN. NCCT scans with lesion annotation from previously published studies by Konduri et al18 and Hssayeni et al19 were used to construct 4 randomly split dedicated validation and test sets (depicted in Fig 2): ischemic stroke lesions between 8-
and 72-hour (24H infarct; N validation: 46; N test: 141) and 72-hour to 2 weeks
after endovascular treatment or randomization. Validation and test sets were constructed with data from the studies by Konduri et al18 and Hssayeni et al19. 8H indicates 8 hours.

Training Data and Training Protocol
All NCCT volumes were converted from DICOM to NIfTI format with dcm2niix available in MRicroGL, Version 1.2.20211006.22 Elastix, Version 5.0.0 (https://elastix.lumc.nl/) was used to coregister the follow-up and BL NCCTs of the training data;23 the scan with the thinnest slices was used as a moving image. Poor coregistration was detected by inspecting the overlay of the 2 images at the 30th, 50th, and 80th percentile sections. Up to 3 follow-up NCCTs were used per patient if clinical deterioration occurred within 8 hours after endovascular treatment or randomization (8 hours) and as part of the imaging protocols of 8–72 hours (24H) and 72 hours to 2 weeks (1W) after AIS.15-17 To ensure stable optimization and prevent overfitting, per training iteration, we used 1 follow-up NCCT and 1 corresponding BL NCCT section (512 × 512). Slices were sampled at random on the 10th and the 95th percentile sections. Furthermore, to emphasize the variation in attenuation between different tissues, the generator model received 2D slices from follow-up NCCTs with 2 channels on the basis of different Hounsfield unit ranges as input: The attenuation was clipped between both 0 and 100 HU for brain and infarct differentiation and 100 and 1000 HU for hemorrhage and skull differentiation. The images were subsequently normalized to a −1 to 1 range. BL NCCTs were only clipped between 0 and 100 HU and normalized to a −1 to 1 range. The discriminator network receives 2D slices of either generated BL or real BL NCCT scans. To make the FU2BL-GAN robust to differences in contrast and noise between the BL and follow-up NCCTs, we applied multiple intensity and noise-altering image augmentations (details available in the Online Supplemental Data). A batch size of 2 with a learning rate of 0.00002 for 500 epochs was used, subsequently linearly reduced to zero over the following 500 epochs (Nvidia TITAN V [https://www.nvidia.com/en-us/titan/titan-v/] with 12-gb RAM). The Online Supplemental Data contains a detailed description of the FU2BL-GAN architecture.

Lesion Segmentation
To obtain lesion segmentations, we passed validation and test set NCCTs through the generator model to generate difference maps. Due to computational constraints, the validation set difference maps were computed every 10th training epoch. Subsequently, segmentations were obtained by applying a threshold to the difference maps. The resulting Dice similarity coefficient (DSC) of the segmentations relative to the ground truth was used to determine the optimal threshold for the difference map —0.2 to +0.3 with steps of 0.01 (equivalent to 0.5 HU). An automatically computed brain mask based on intensity thresholds and region growing was used to remove false-positive segmentations that were not allocated inside the skull.7 In the Online Supplemental Data, validation set results are depicted. Finally, the optimal epoch and threshold were used to obtain segmentations for the test sets.
Evaluation and Outcome Metrics

Reported results were based on test set segmentations and were reported relative to expert-based ground truth segmentations. The DSC and Hausdorff distance in millimeters were used to compute spatial correspondence. Results from the FU2BL-GAN approach trained with L1 and adversarial loss (L1 + adv) were compared statistically with the Wilcoxon rank-sum test with a simpler approach trained with L1-loss only (L1) only. Furthermore, the results of the FU2BL-GAN were compared with two 2D Unets trained on segmentations from the 24H and 1W infarct validation sets using the no new Unet (nnUnet) framework as a conventional supervised learning BL.24 Volumetric correspondence between the ground truth and predicted segmentations were analyzed with Bland-Altman plots with bias (mean between methods) and limits of agreement (LoA, ±1.96 SDs from the bias) and the intraclass correlation coefficient (ICC) with 95% CIs. The 2-way mixed-effects approach for consistency of a single fixed rate was used to describe differences between the FU2BL-GAN-based segmentations and the expert-based ground truth lesion segmentations. A subgroup analysis was performed for lesions of >10 mL to address the effect of lesion size on our outcome metrics. Results were reported as median with interquartile range (IQR) or mean with 95% CIs.

RESULTS

Ischemic and hemorrhagic lesions in our test sets were relatively small; the distribution of volumes was skewed toward smaller lesions. Ground truth lesion volume of test sets had a median of 35 mL (IQR, 16–78 mL) in the 24H and 66 mL (IQR, 29–125 mL) in the 1W infarct NCCTs, respectively. For the HT and PrH test sets respectively, the mean lesion size was 6 mL (IQR, 2–12 mL) and 6 mL (IQR, 1–12 mL). Characteristics of the training data can be found in the Online Supplemental Data. Training characteristics and the optimal difference map thresholds are available in the Online Supplemental Data.

Quantitative Results

As depicted in Figs 3 and the Online Supplemental Data, DSC and lesion volume were positively related. The median DSC of the FU2BL-GAN was 0.31 (IQR, 0.08–0.59) in the 24H infarct test set, 0.59 (IQR, 0.29–0.74) for the 1W infarct test set, 0.02 (IQR, 0.0–0.14) for the HT test set, and 0.08 (IQR, 0.01–0.35) for the PrH test set. The FU2BL-GAN (L1 + adv) had a statistically significant higher DSC than the model trained with only L1-loss (L1) for all test sets but a significantly lower DSC compared with the nnUnet approach (Fig 3). The subgroup of lesions of >10 mL (Fig 3B) had a higher DSC than the overall population (Fig 3A), especially for the HT (median, 0.46; IQR, 0.07–0.51) and PrH (median, 0.44; IQR, 0.24–0.55) test sets but also for the follow-up infarct (24H infarct: median, 0.41; IQR, 0.15–0.62; 1W infarct: median, 0.60; IQR, 0.35–0.75) test sets. For all the infarct and hemorrhage test sets, the Hausdorff distances of both the FU2BL-GAN and L1 approach were poor with a median varying between 83 and 87 mm (Online Supplemental Data). Bland-Altman plots of the 4 test sets are depicted in Fig 4. Bias and LoA for the FUB2BL-GAN of the 24H (bias, –3 mL; LoA, –64–59 mL) and 1W (bias, –4 mL; LoA, –66–58 mL) test sets were low, representing good correspondence of the segmentations with ground truth annotations. However, both the HT (bias, 22 mL; LoA, –49–92 mL) and PrH (bias, 23 mL; LoA, –10–57 mL) segmentations overestimated lesion volume and had several outliers that affected volumetric correspondence. The ICC for the FUB2BL-GAN was excellent in the 1W infarct test set (ICC, 0.90; 95% CI, 0.83–0.93), good in the 24H infarct (ICC, 0.83; 95% CI, 0.78–0.88) and PrH (ICC, 0.84; 95% CI, 0.66–0.93) test sets, and poor in the HT (ICC, 0.11; 95% CI, –0.15–0.36) test sets. However, the nnUnet approach resulted in a lower bias and LoA, a higher ICC, and a lower Hausdorff distance than the FU2BL-GAN for both the 24H (bias, –14 mL; LoA, –57–29 mL; ICC, 0.91; 95% CI, 0.88–0.92; Hausdorff distance, 28 mm; IQR, 18–42 mm) and 1W (bias, –6 mL; LoA, –36–23 mL; Hausdorff distance, 36 mm; IQR, 23–45 mm).
ICC, 0.98; 95% CI, 0.97–0.99; Hausdorff distance, 21 mm; IQR, 14–37 mm) infarct lesions test sets (Online Supplemental Data).

**Qualitative Visual Results**

Figure 5 depicts visual examples of each test set in the first 4 rows, while the last 2 rows depict examples with poor segmentation performance. In contrast to the examples shown in the first 3 rows, the input NCCT of the PrH test set is the acute-phase NCCT with hemorrhagic lesions. For this case, the generated scan can be regarded as a prehemorrhagic stroke NCCT scan. Although lesions visually appear to be removed accurately, the generator model was not able to completely reconstruct 24H infarcted brain tissue similar to the BL NCCTs (columns 1 versus 3). False-positive hemorrhage segmentations were present when the input NCCT scan had beam-hardening artifacts in the brain close to the skull or when the overall scan attenuation was higher (arrows in PrH and HT column 2). False-negative hemorrhage segmentations were present if the hemorrhage was small and the attenuation increase was low (row 6, poor HT). False-positive infarct segmentations occurred close to the ventricles and other locations, where CSF results in a hypoattenuated region (row 5). False-negative infarct segmentation errors mainly occurred in the 24H infarct data set because the infarct lesion was not yet significantly hypoattenuated (row 5).

**DISCUSSION**

Our study shows that when one uses a GAN deep learning structure, it is possible to obtain follow-up ischemic and large hemorrhagic lesion segmentations without using manually annotated training data. Although the visual quality of generated BL scans was not always optimal, lesion segmentation quality was often not affected. External validation in 4 test sets revealed reasonable segmentation quality in terms of DSC and good-to-excellent volumetric correspondence with the ground truth for follow-up infarct lesions in NCCT at 24H and 1W follow-up after AIS. In terms of DSC and volumetric correspondence, our work performs on a par with previous work on supervised deep-for-follow-up infarct lesion segmentation (DSC median, 0.57 [SD:0.26]; ICC, 0.88).9 However, the presented unsupervised FU2BL-GAN did not outperform the supervised nnUnet benchmark model with respect to all outcome measures. Kuang et al25 also used a GAN to segment infarct lesions but achieved much higher segmentation quality (DSC mean, 0.70 [SD, 0.12]). However, the approach by Kuang et al required a training set with manual lesion annotations because the adversarial (GAN) loss was used in addition to the supervised loss functions. DSC and volumetric correspondence for segmenting the HT and PrH lesions were worse than those of existing supervised methods.4,5,8,10 Poor detection and segmentation of hemorrhagic lesions are likely due to the small lesion size in our test sets and an under-representation of hemorrhages in the training data.

The unsupervised approach to training is a major advantage compared with conventional supervised deep learning methods. With the growing availability of unlabeled and weakly labeled imaging data bases, unsupervised GAN-based approaches can be

![FIG 4. Bland-Altman plots of predicted lesion size for the FU2BL-GAN. A, 24H infarct follow-up. B, 1W infarct follow-up. C, HT. D, PrH.](image-url)
used without the manual annotation effort for automated lesion segmentation. However, the downside of the presented approach is the requirement of paired training images, coregistered images with and without lesions from the same patient. Such high-quality registration is often difficult to achieve when considering medical imaging because most organs, tissues, and body parts deform or move between acquisition moments. Because the brain only slightly deforms and moves between acquisition moments, the use of a GAN-based lesion segmentation method similar to the presented FU2BL-GAN seems promising for other brain pathologies.

One of the main shortcomings of the presented FU2BL-GAN is that it can only be trained on CT slices sampled at random. Because not every section in an NCCT volume of a patient contains an initial AIS lesion and only a minority of the volumes contain a hemorrhagic lesion, the FU2BL-GAN likely experienced an under-representation of brain lesions. This under-representation during the training of NCCT slices with a lesion, especially with a hemorrhagic lesion, compared with slices without lesions, is known to result in poorer segmentation performance; in technical literature, this is often referred to as the “class imbalance problem.” In contrast, supervised deep learning methods often use adjusted sampling techniques that require ground-truth annotations; most class (nonlesion tissue) is undersampled relative to the minority class (the lesion) to balance class representation. A valuable improvement in our FU2BL-GAN would be to manually classify slices for lesion presence and volumes for the presence of a hemorrhage. Although these section- or volume-level annotations would take some time to acquire, such sparse annotation methods are still less time-consuming than manually segmenting lesions required for supervised deep learning. Alternatively, automated NCCT-section classification algorithms for infarct or hemorrhage presence can be used to classify NCCT slices on the basis of lesion presence.

Subsequently, this information can be used to select training data for further improvement of the FU2BL-GAN.

Although the test sets used in this study are from multiple centers, it remains largely unclear what scanners, settings, and post-processing methods were used. Furthermore, Konduri et al reported extensive exclusion criteria related to the image quality and noise level, excluding 93 of 280 patients in their data set. These factors influence the ability to generalize results from this study and require additional external validation on subgroups and other data sets.

CONCLUSIONS
The presented FU2BL-GAN is an unsupervised deep learning approach trained without manual lesion annotations to segment stroke lesions. With the FU2BL-GAN, it is feasible to obtain automated infarct lesion segmentations with moderate DSC and good volumetric correspondence.

ACKNOWLEDGMENTS
CONTRAST Clinical Trial Collaborators
Research Leaders
Diederik Dippel (MD, PhD),1 Charles Majoie (MD, PhD)3
Consortium Coordinator
Rick van Nuland, (PhD)24
Imaging Assessment Committee
Charles Majoie (MD, PhD)—Chair,3 Aad van der Lugt (MD, PhD)—Chair,1 Adriaan van Es, (MD, PhD),1,2 Pieter-Jan van Doormaal (MD),1 René van den Berg, (MD, PhD),2 Ludo Beenen (MD),3 Bart Emmer (MD, PhD),3 Stefan Roosendaal (MD, PhD),3 Wim van Zwam (MD, PhD),4 Alda Annechien Postma (MD, PhD),25 Lonneke Yo (MD, PhD),6 Menno Krietemeijer (MD),6

FIG 5. Visual results of the FU2BL-GAN. The first column contains the input NCCT with lesion used as input for the generator model to generate a difference map (column 2). The difference map is subtracted from the input NCCT (column 1) to obtain a generated BL scan (column 3). The negative (blue) and positive (red) values of the difference map correspond to the deviation of the difference map from zero. A higher deviation from zero implies a higher attenuation adjustment of the follow-up NCCT to generate the BL NCCT without a lesion. Column 4 contains the ground truth lesion annotations. Arrows show false-positive hemorrhage (rows 3 and 4), false-negative infarct (row 5, upper arrow), false-positive infarct (row 5, lower arrow), and false-negative hemorrhage segmentation (arrow, row 6).
REFERENCES

1. Katramados AM, Hacèine-Beïl Y, Varelas PN. What to look for on post-stroke neuroimaging. *Neuroimaging Clin N Am* 2018;28:649–62 CrossRef Medline

2. Park TH, Lee JK, Park MS, et al. Neurologic deterioration in patients with acute ischemic stroke or transient ischemic attack. *Neurology* 2020;95:e2178–91 CrossRef Medline

3. Seners P, Turc G, Oppenheim C, et al. Incidence, causes and predictors of neurological deterioration occurring within 24 h following acute ischaemic stroke: a systematic review with pathophysiological implications. *J Neurol Neurosurg Psychiatry* 2015;86:87–94 CrossRef Medline

4. Karthik R, Menaka R, Johnson A, et al. Neuroimaging and deep learning for brain stroke detection: a review of recent advancements and future prospects. *Comput Methods Programs Biomed* 2020;197:105728 CrossRef Medline

5. Yeo M, Tahayori B, Kok HK, et al. Review of deep learning algorithms for the automatic detection of intracranial hemorrhages on computed tomography head imaging. *J Neurointerv Surg* 2020;13:369–78 CrossRef Medline
6. Boers AM, Jansen IG, Beenen LF, et al. Association of follow-up infarct volume with functional outcome in acute ischemic stroke: a pooled analysis of seven randomized trials. *J Neurointerv Surg* 2018;10:1137–42

7. Hemphill JC, Greenberg SM, Anderson CS, et al; Council on Clinical Cardiology. Guidelines for the Management of Spontaneous Intracerebral Hemorrhage: a Guideline for Healthcare Professionals from the American Heart Association/American Stroke Association. *Stroke* 2015;46:2032–60

8. Barros RS, van der Steen WE, Boers AMM, et al. Automated segmentation of subarachnoid hemorrhages with convolutional neural networks. *Informatics Med Unlocked* 2020;19:100321

9. Barros RS, Tolhuisen ML, Boers AM, et al. Automatic segmentation of cerebral infarcts in follow-up computed tomography images with convolutional neural networks. *J Neurointerv Surg* 2020;12:48–52

10. Zhao X, Chen K, Wu G, et al. Deep learning shows good reliability for automatic segmentation and volume measurement of brain hemorrhage, intraventricular extension, and peripheral edema. *Eur Radiol* 2021;31:5012–20

11. Goodfellow I, Bengio Y, Courville A. *Deep learning*. MIT Press 2016. deeplearningbook.org. Accessed July 11, 2022

12. Yi X, Walia E, Babyn P. *GoodfellowI, BengioY, CourvilleA.* Visual Feature Attribution using Wasserstein GANs. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, Utah. June 18–23, 2018:8309–19

13. Treurniet KM, LeCouffe NE, Kappelhof M, et al; MR CLEAN-NO IV Investigators. MR CLEAN-NO IV: intravenous treatment followed by endovascular treatment versus direct endovascular treatment for acute ischemic stroke caused by a proximal intracranial occlusion—study protocol for a randomized clinical trial. *Trials* 2021;22:1–15

14. Chalos V, Van De Graaf RA, Roonzenbeek B, et al; MR CLEAN-MED investigators. Multicenter randomized clinical trial of endovascular treatment for acute ischemic stroke: the effect of periprocedural medication—acetylsalicylic acid, unfractionated heparin, both, or neither (MR CLEAN-MED). Rationale and study design. *Trials* 2020;21:1–17

15. Pirson FA, Hinsenveld WH, Goldhoorn R-JB, et al; MR CLEAN-LATE investigators. MR CLEAN-LATE study protocol. *Trials* 2021;22:160

16. Konduri P, van Kranendonk K, Boers A, et al; MR CLEAN Trial Investigators (Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands). The role of edema in subacute lesion progression after treatment of acute ischemic stroke. *Front Neurol* 2021;12:705221

17. Hssayeni MD, Crock MS, Salman AD, et al. Intracranial hemorrhage segmentation using a deep convolutional model. *Data* 2020;5:14–18

18. Van Kranendonk KR, Treurniet KM, Boers AM, et al; MR CLEAN Investigators. Hemorrhagic transformation is associated with poor functional outcome in patients with acute ischemic stroke due to a large vessel occlusion. *J Neurointerv Surg* 2019;11:464–68

19. Berkhemer OA, Fransen PS, Beumer D, et al. MR CLEAN Investigators. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med* 2015;372:11–20

20. dcm2niix. 2021. https://github.com/rordenlab/dcm2niix. Accessed November 3, 2021

21. Klein S, Staring M, Murphy K, et al. *Elastix: a toolbox for intensity-based medical image registration*. *IEEE Trans Med Imaging* 2010;29:196–205

22. Kuang H, Menon BK, Qiu W. Automated stroke lesion segmentation in non-contrast CT scans using dense multi-path contextual generative adversarial network. *Phys Med Biol* 2020;65:215013

23. Lin TY, Goyal P, Girshick R, et al. *nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation*. *Nat Methods* 2021;18:203–11

24. Kwon H, Menon BK, Qiu W. Automated stroke lesion segmentation in non-contrast CT scans using dense multi-path contextual generative adversarial network. *Phys Med Biol* 2020;65:215013

25. Flanders AE, Prevedello LM, Shih G, et al; RSNA-ASNR 2019 Brain Hemorrhage CT Annotators. Construction of a Machine Learning Dataset through Collaboration: The RSNA 2019 Brain CT Hemorrhage Challenge. *Radiology Artif Intell* 2020;2:e209002