







OPEN ACCESS

ORIGINAL RESEARCH

# Automatic segmentation of cerebral infarcts in follow-up computed tomography images with convolutional neural networks

Renan Sales Barros,<sup>1</sup> Manon L Tolhuisen,<sup>1,2</sup> Anna MM Boers,<sup>1,3</sup> Ivo Jansen,<sup>2</sup> Elena Ponomareva,<sup>3</sup> Diederik W J Dippel,<sup>4</sup> Aad van der Lugt,<sup>5</sup> Robert J van Oostenbrugge,<sup>6</sup> Wim H van Zwam ,<sup>7,8</sup> Olvert A Berkhemer,<sup>2,5</sup> Mayank Goyal ,<sup>9</sup> Andrew M Demchuk,<sup>10</sup> Bijoy K Menon,<sup>11</sup> Peter Mitchell,<sup>12</sup> Michael D Hill ,<sup>10</sup> Tudor G Jovin,<sup>13</sup> Antoni Davalos,<sup>14</sup> Bruce C V Campbell,<sup>15,16</sup> Jeffrey L Saver,<sup>17</sup> Yvo B W E M Roos,<sup>18</sup> Keith W. Muir,<sup>19</sup> Phil White,<sup>20,21</sup> Serge Bracard,<sup>22</sup> Francis Guillemin,<sup>22</sup> Silvia Delgado Olabarriaga,<sup>2</sup> Charles B L M Majoie,<sup>23</sup> Henk A Marquering <sup>1,2</sup>

For numbered affiliations see end of article.

## Correspondence to

Dr Henk A Marquering, Radiology, Academic Medical Center, 1105 AZ Amsterdam, The Netherlands; h.a.marquering@amc.uva.nl

Received 26 September 2019  
Revised 19 November 2019  
Accepted 20 November 2019  
Published Online First 23 December 2019

## ABSTRACT

**Background and purpose** Infarct volume is a valuable outcome measure in treatment trials of acute ischemic stroke and is strongly associated with functional outcome. Its manual volumetric assessment is, however, too demanding to be implemented in clinical practice.

**Objective** To assess the value of convolutional neural networks (CNNs) in the automatic segmentation of infarct volume in follow-up CT images in a large population of patients with acute ischemic stroke.

**Materials and methods** We included CT images of 1026 patients from a large pooling of patients with acute ischemic stroke. A reference standard for the infarct segmentation was generated by manual delineation. We introduce three CNN models for the segmentation of subtle, intermediate, and severe hypodense lesions. The fully automated infarct segmentation was defined as the combination of the results of these three CNNs. The results of the three-CNNs approach were compared with the results from a single CNN approach and with the reference standard segmentations.

**Results** The median infarct volume was 48 mL (IQR 15–125 mL). Comparison between the volumes of the three-CNNs approach and manually delineated infarct volumes showed excellent agreement, with an intraclass correlation coefficient (ICC) of 0.88. Even better agreement was found for severe and intermediate hypodense infarcts, with ICCs of 0.98 and 0.93, respectively. Although the number of patients used for training in the single CNN approach was much larger, the accuracy of the three-CNNs approach strongly outperformed the single CNN approach, which had an ICC of 0.34.

**Conclusion** Convolutional neural networks are valuable and accurate in the quantitative assessment of infarct volumes, for both subtle and severe hypodense infarcts in follow-up CT images. Our proposed three-CNNs approach strongly outperforms a more straightforward single CNN approach.

## INTRODUCTION

Measuring the volume of infarcts on non-contrast computed tomography (NCCT) scans provides a quantitative assessment of infarcted brain tissue resulting from ischemic stroke. Follow-up infarct volume measured after 24 hours from onset<sup>1</sup> is a valuable predictor of functional outcome. Infarct volume has been suggested as a surrogate endpoint for classic patient outcome scales in multiple randomized controlled trials.<sup>2</sup> By combining infarct volume with infarct location, a more precise prediction of patient outcome can be achieved.<sup>3</sup>

The reference standard for infarct segmentation is manual delineation by medical experts. However, manual delineation has several disadvantages as it is time-demanding, subjective, prone to errors, and costly.<sup>4</sup> Accordingly, manual delineation does not work well in large cohort studies.

Convolutional neural networks (CNNs) have outperformed many existing image analysis methods for image classification and image segmentation. CNNs have produced good segmentation results in multiple medical imaging domains, including segmentation of ischemic stroke lesions in magnetic resonance images of the brain.<sup>5–7</sup> In this study, we evaluated the usefulness of CNNs for automatic segmentation of infarcted brain tissue in follow-up NCCT scans from patients with an acute ischemic stroke.

## MATERIALS AND METHODS

### Image data

We used anonymized image data from the HERMES collaboration.<sup>8</sup> This collaboration combined clinical and image data from seven clinical trials that investigated the efficacy of endovascular therapy in patients with acute ischemic stroke. Central medical ethics committees and research boards of each participating hospital approved each trial and the use of anonymized image data in this retrospective study. All patients, or their legal representatives, provided written informed consent.



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

**To cite:** Sales Barros R, Tolhuisen ML, Boers AMM, et al. *J NeuroIntervent Surg* 2020;**12**:848–852.

We used image data only from patients with follow-up NCCT acquired between 12 hours and 2 weeks after stroke onset and for whom a reference infarct segmentation was available. A total of 1026 patients had follow-up NCCT imaging acquired within the selected time window and with an available reference segmentation. Thin-slice image data were reconstructed into scans with 5 mm slice thickness.

## Reference segmentations

The reference infarct segmentation on the follow-up NCCT scans was manually delineated by one of two experienced observers, as described by Boers *et al.*<sup>9</sup> In short, infarcts were identified as hypodense areas. Infarcted tissue in the ipsilateral hemisphere with characteristics of an old infarct were excluded from the reference segmentation. NCCT scans of patients who underwent decompressive hemicraniectomy were excluded. Parenchymal hemorrhages within or adjacent to the infarcted area were included in the reference segmentation. A standard window width of 30 Hounsfield units (HU) and center level of 35 HU were used to limit variation between observers. If multiple follow-up images were available, reference segmentation was performed in the latest acquired scan. The manual segmentations were checked by one of three expert radiologists and, when necessary, corrections were made.

## Preprocessing

To exclude trivial voxels that were of no interest, such as air or skull, we used automatic methods for intracranial region and cerebrospinal fluid (CSF) segmentation. First, we excluded all voxels outside the brain using an intracranial region segmentation. Subsequently, we also discarded all voxels selected by the CSF segmentation. All discarded voxels were neither used to train the CNN nor used for accuracy testing of the CNN.

The intracranial region segmentation uses the size range of the foramina of the skull, as reported by Berge *et al.*,<sup>10</sup> and typical HU values of the skull. This segmentation was performed according to the following steps:

- ▶ A threshold-based segmentation was performed to segment bones. We considered everything with intensity >160 HU as bone.
- ▶ A morphological dilatation with a 7 mm radius was used to close all foramina of the skull except the foramen magnum.
- ▶ The center of gravity of the segmented bone was used as a seed for a region growing inside the skull.
- ▶ A morphological dilatation with a 7 mm radius was applied to the region growing result to bring the segmented intracranial region close to the skull border.
- ▶ The foramen magnum was detected by evaluation of the segmented area in each individual slice from top to bottom. The foramen magnum slice was determined as the first slice with a segmented area <900 mm<sup>2</sup> after the slice with the maximum segmented area. All voxels below the foramen magnum slice were excluded from the segmentation.

The CSF segmentation was performed by selecting the voxels around the centroid of the segmented intracranial region as seeds for region growing. All voxels within a maximum distance of 15 mm from this centroid and with density values between -5 and 13 HU were used as seeds. The lower and upper thresholds of this region growing were also -5 and 13 HU.

We used a previously presented method for automated intracranial hemorrhage segmentation<sup>11</sup> to exclude the parenchymal hemorrhages of the CNN-based infarct segmentation. These hemorrhage voxels were not used to train the CNN. However,

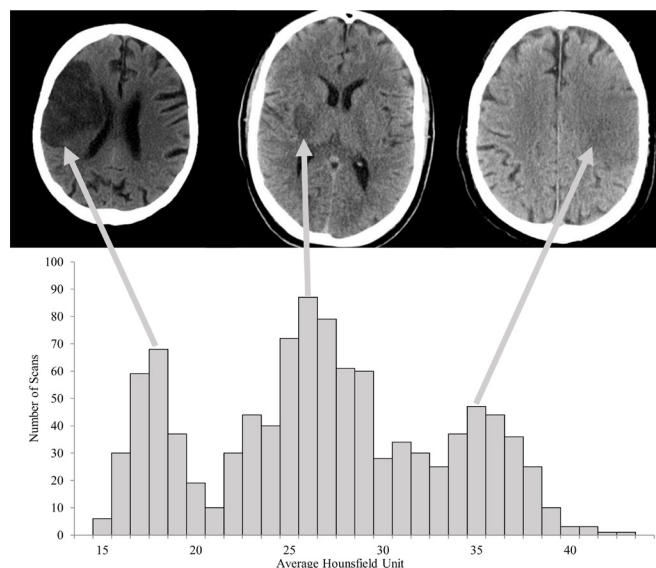
for infarct volume accuracy testing, any area that was classified as hemorrhage was added to the infarct segmentation.

## CNN-based infarct segmentation

The CNN architecture used in this study was developed in-house. Its hyperparameters were optimized for segmentation of a single foreground structure in head NCCT scans, which in this case was the infarcted brain tissue. Previously, the same CNN architecture was successfully used for intracranial hemorrhage segmentation.<sup>11</sup> This CNN architecture determines the probability of the voxel at the center of an image patch being foreground (infarcted tissue) or background (any other tissue). This probability was subsequently dichotomized using a cut-off value, which was optimized with the data in the validation set.

The CNN architecture has two convolutional layers followed by two fully connected dense layers. Each dense layer has 256 nodes. The size of the input patch was 19×19×3 voxels; 19×19 voxels in the axial plane and three slices high. Each slice of the input patch was processed as a different image channel. After each convolutional layer, there is a max-polling layer with a 2×2 kernel and a 2×2 stride. The first convolutional layer has 64 feature maps and the second has 128 feature maps. Both convolutional layers have kernels with size 5×5.

The hypodensity of the infarcted tissue in NCCT scans is related to breakdown of cells and its fluid content. As shown in figure 1, the infarcted areas in the three NCCT scans have different HU values. In figure 1, we also show the distribution of the average HU values of the infarct reference segmentations. In our population, the HU value distribution depicted three peaks, which we named subtle, intermediate, and severe hypodensity infarcts. Because of this observation, we trained three CNNs. Each of these CNNs was trained to classify a different hypodensity distribution of infarcted brain tissue. We grouped all patients according to the hypodensity of the delineated infarct. We used the average HU value of the infarction for this grouping. The average infarct intensity was computed after excluding the hemorrhage voxels of the reference segmentation.



**Figure 1** Histogram of average infarct intensities of the manually delineated infarcts. The left CT image at the top displays a relatively old infarct with a severe hypodensity; in the middle, an intermediate old infarct is shown; and the image on the right shows a relatively young infarct with a subtle hypodensity.

The thresholds that define each infarction class were (14, 22) HU for severe, (22, 32) HU for intermediate, and (32, 44) HU for subtle.

We used 570 randomly selected scans to train the three CNNs. We augmented the number of training infarct patches by flipping along the sagittal plane and by rotation. No data augmentation was applied to the non-infarct patches. We used an additional 60 scans to optimize the cut-off value for generating binary segmentations, 20 scans for each CNN. The union of the results of these three CNNs and the result of the intracranial hemorrhage segmentation was considered to be the automated generated infarct segmentation. The remaining 396 scans were used to test segmentation performance.

For comparison, we also trained a single CNN architecture for the segmentation of all types of infarction. The same methodology and data were used for this single CNN approach and the three-CNNs approach.

We used the Dice coefficient as an accuracy measure of the infarct segmentation performance in the test set. We calculated the intraclass correlation coefficients (ICCs) to compare the reference and the automatically generated infarct volumes. ICCs were interpreted according to the American Psychological Association *al*<sup>12</sup>:  $<0.4$  is poor;  $\geq 0.4$  to  $<0.6$  is fair,  $\geq 0.6$  to  $<0.75$  is good, and  $\geq 0.75$  is excellent. We opted not to compare our approach with U-Net or Mask R-CNN architectures. Both these architectures are more extensive than the proposed architecture and, in a straightforward approach, their input would be an entire NCCT slice. Since we used 5 mm reconstructions, and not all slices from a NCCT scan have infarction, we did not expect a satisfactory segmentation given the limited number of NCCT

**Table 1** Results of automated infarct segmentation for severe, intermediate, and subtle hypodense infarcts and the average over the whole test dataset for the three-CNNs approach. for comparison with the accuracy of the single CNN approach

		ICC	Dice	Test set size
Three-CNNs approach	Severe	0.98	0.78±0.09	67
	Intermediate	0.93	0.61±0.21	204
	Subtle	0.66	0.37±0.26	125
	All infarctions	0.88	0.57±0.26	396
Single CNN approach	All infarctions	0.34	0.18±0.23	396

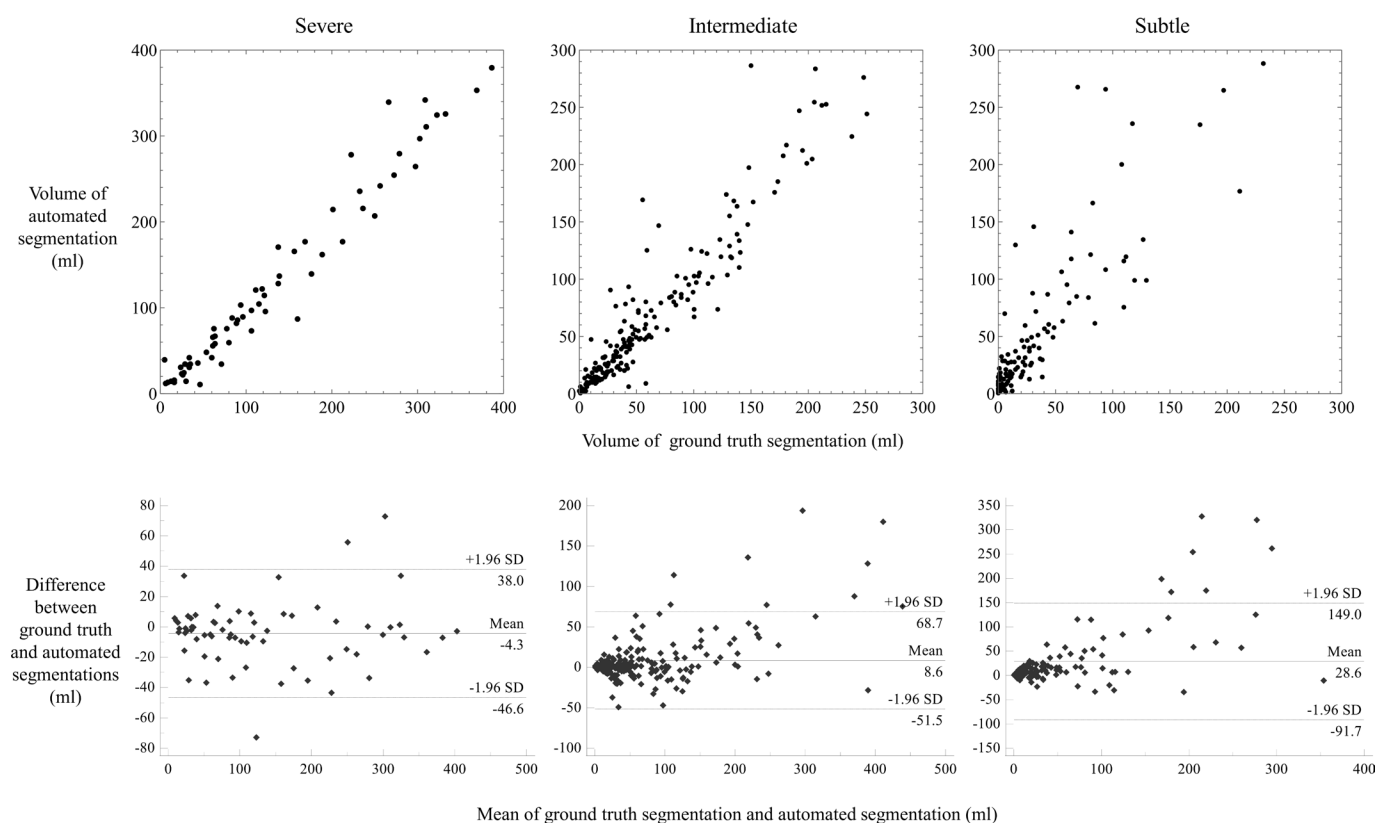
CNN, convolutional neural network; ICC, intraclass correlation coefficient.

slices with infarcted brain tissue that would be used as training samples.

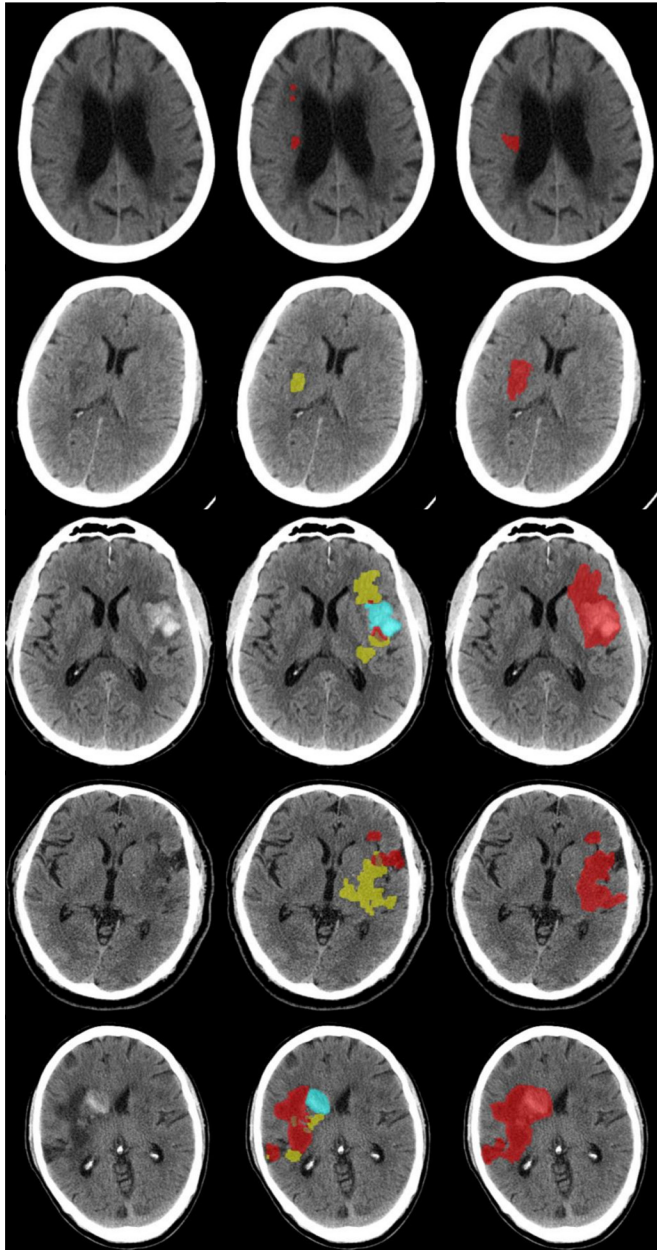
## RESULTS

The median infarct volume was 48 (IQR 15–125) mL overall, with 29 (IQR 11–86), 46 (IQR 18–101), and 89 (IQR 35–210) mL for patients with a subtle, intermediate, and severe hypodense infarct, respectively.

The comparison between manually delineated infarct volumes and the volumes from the three-CNNs approach showed an excellent agreement with an ICC of 0.88. Even better agreement was observed for severe and intermediate hypodense infarcts with ICCs of 0.98 and 0.93, respectively. Agreement was good for subtle hypodense infarcts, with an ICC of 0.66. In [figure 2](#),



**Figure 2** Top: Comparison of the infarct volume of the results from the three-CNNs approach (y axis) with the reference to infarct volume (x axis). Bottom: Bland-Altman plots of the infarct volumes. The difference in the volume determination is given along the y axis, and the average of the automated and reference infarct volume is depicted along the x axis. The different columns show separate severe, intermediate, and subtle hypodensity infarcts.



**Figure 3** Sample results. from left to right we have input image, union of the segmentation results, and reference segmentation. For simplicity, in the center column we rendered the hemorrhages (blue) over the subtle infarcts (yellow), subtle infarcts over standard infarcts (orange), and standard infarcts over severe infarcts (red). The Dice coefficients from top to bottom were 0.10, 0.26, 0.40, 0.55, and 0.70. In the left column the original images are shown. The right shows the merged segmentations.

the agreement between the infarct volumes is shown. Agreement of the single CNN approach was poor, with an ICC of 0.34.

The average Dice coefficient achieved by the three-CNNs approach was  $0.57 \pm 0.26$ . The average Dice coefficients for each category were  $0.78 \pm 0.09$ ,  $0.61 \pm 0.21$ , and  $0.37 \pm 0.26$ , for the severe, intermediate, and subtle hypodense infarcts, respectively. The method based on a single CNN achieved an average Dice coefficient of  $0.18 \pm 0.23$ . Table 1 shows a summary of the segmentation performance measures. In figure 3, we show some sample results from the three-CNNs approach.

## DISCUSSION

We have shown that CNNs are valuable in the automated cerebral infarct segmentation in follow-up CT images of patients with acute ischemic stroke, with excellent agreement with volumetric assessments of expert observers. Owing to the wide variety of the severity of hypodensities, we proposed using the combination of three CNNs, which strongly outperformed a single CNN approach.

Infarct location and infarct volume have been strongly associated with outcome of patients with ischemic stroke in several studies.<sup>3,13</sup> Reliably segmenting cerebral infarcts is challenging because of pathophysiological heterogeneity, presence of pre-existing pathologies such as old infarcts, leukoaraiosis, atrophy, intrinsic differences in attenuation of grey and white matter, and hemorrhagic transformation. Thus, to be able to develop robust automated methods for cerebral infarct segmentation, heterogeneous image data are required. The proposed method was evaluated in a large cohort of patients from seven multicenter randomized trials enrolling in multiple countries. The follow-up NCCT scans used in our study also had a (pragmatically) wide range of follow-up time after stroke onset, ranging from 12 hours to 2 weeks. Despite these variations, the proposed approach based on three different CNNs produced accurate cerebral infarct segmentations. The volume of these segmentations had good or excellent correlation with the reference infarct volume. We have shown that accuracy for old, severe hypodense infarcts was higher than for subtle hypodense infarcts. Note that, although we presented the results in a selective manner, exactly the same procedure was applied for the infarct segmentations in all the three different infarct categories.

A number of previous studies on automatic infarct core segmentation in various image modalities have been presented. Multiple CNN-based techniques have been introduced recently. On baseline CT perfusion, state-of-the-art infarct segmentation was obtained by a CNN architecture proposed by Liu *et al*,<sup>14</sup> achieving an average Dice coefficient of  $0.51 \pm 0.31$ . On MRI the CNN architecture proposed by Kamnitsas *et al*<sup>6</sup> reported an average Dice coefficient of  $0.66 \pm 0.24$ . Maier *et al*<sup>7</sup> tested several methods with different types of MR images. Their best reported result was achieved by a CNN with an average Dice coefficient of  $0.73 \pm 0.18$ . The current state-of-the-art method for infarct segmentation on MR images is the CNN proposed by Zhang *et al*,<sup>5</sup> which achieved an average Dice coefficient of 0.79 in a test set with 90 images. Although good segmentation results were achieved in CT perfusion and MR images, NCCT scans are still the predominant method for assessment of follow-up infarct in patients with ischemic stroke. Therefore, we focused on using NCCT as input for the proposed cerebral infarct segmentation method.

On NCCT scans, two semiautomated methods are available for infarct segmentation. The semiautomated method by Bardera *et al*<sup>15</sup> was evaluated with 18 patients and reported a Pearson's correlation coefficient of 0.98 and 0.97 compared with the manual segmentations from two different observers. The semiautomated method by Kuang *et al*<sup>16</sup> was evaluated with 16 patients and reported an average Dice coefficient of  $0.76 \pm 0.10$ . By contrast, our method is both fully automated, which avoids the variability introduced by the user inputs, and has been tested on a far larger number of patients.

Other fully automated methods for infarct segmentation on NCCT are available. The method by Boers *et al*<sup>17</sup> reported an average Dice coefficient of  $0.74 \pm 0.13$  in a test set with 34 images. The average onset to follow-up scanning time in the study by Boers *et al* was  $4.1 \pm 2.3$  days. The average Dice

coefficient between human observers in the study by Boers *et al* was 0.84 ranging from 0.63 to 0.94, which was somewhat higher than the agreement we achieved. However, it should be noted that the manual delineation was performed for old, hypodense infarcts only. The method by Vos *et al*<sup>18</sup> reported an average Dice coefficient of  $0.74 \pm 0.09$  in a test set with 30 images. In the study by Vos *et al*, the average time between onset and scan acquisition was 3 days ranging between 2 and 5 days. More recently, the method by Gillebert *et al*<sup>19</sup> was evaluated with 12 patients with ischemic stroke and reported Dice coefficients ranging from 0.27 to 0.71. The scans used to evaluate the method by Gillebert *et al* had an average acquisition time after onset of 40 hours. Their method was evaluated in a limited set of selected images to illustrate different types of ischemic stroke lesions. In contrast with the methods of Boers *et al*, Vos *et al*, and Gillebert *et al*, our method has been thoroughly evaluated with a large and diverse test set.

The data used in our study included follow-up scans as early as 12 hours after stroke onset. Infarcts in these early follow-up scans might be subtle and harder to segment. Thus, it was expected that our method would achieve a lower accuracy in such scans. Moreover, the manual delineation in these scans is more difficult, resulting in more variation among experts. This may also strongly contribute to the lower agreement of the automated method with the reference standard. In some cases (also in figure 3), the network in charge of segmenting subtle infarcts overestimates the infarct region by including subtle hypodense areas which are not part of the infarction. Another common source of misclassifications by our proposed method is the inclusion of cerebral sulci in the results of the network trained to segment severe infarctions (figure 3).

A major limitation is the highly selective nature of the HERMES population. All patients had anterior circulation stroke confirmed by CT angiography, mostly within 6 hours of onset. Patients were excluded from most studies if they had prior disability or low Alberta Stroke Program Early CT scores. As a result, many of the background abnormalities typical in populations with acute stroke were less prevalent in our population. Moreover, average age was around 69, and very elderly patients were under-represented. Despite variation among study populations, these still represent a much more homogeneous group than patients with stroke as a whole.

Overall, the proposed method achieved an excellent correlation with the reference infarct volume. This suggests that our method can be used in clinical trials, replacing tedious manual delineations. Its value in functional outcome prediction for patients with ischemic stroke and its value as a secondary outcome measure in treatment trials still has to be established.

#### Author affiliations

<sup>1</sup>Department of Biomedical Engineering and Physics, Amsterdam UMC, location AMC, Amsterdam, the Netherlands

<sup>2</sup>Department of Radiology and Nuclear Medicine, Amsterdam UMC, location AMC, Amsterdam, the Netherlands

<sup>3</sup>Nico-lab, Amsterdam, Netherlands

<sup>4</sup>Department of Neurology, Erasmus MC - University Medical Center, Rotterdam, Netherlands

<sup>5</sup>Department of Radiology and Nuclear Medicine, Erasmus MC - University Medical Center Rotterdam, Rotterdam, Netherlands

<sup>6</sup>Department of Neurology, School for Cardiovascular Diseases (CARIM), Maastricht University Medical Center, Maastricht, the Netherlands

<sup>7</sup>Department of Radiology and Nuclear Medicine, Maastricht University Medical Center, Maastricht, Netherlands

<sup>8</sup>Cardiovascular Research Institute Maastricht (CARIM), Maastricht, the Netherlands

<sup>9</sup>Department of Diagnostic Imaging, University of Calgary, Calgary, Alberta, Canada

<sup>10</sup>Department of Clinical Neurosciences, University of Calgary, Calgary, Alberta, Canada

<sup>11</sup>Calgary Stroke Program, University of Calgary, Calgary, Alberta, Canada

<sup>12</sup>Department of Radiology, Royal Melbourne Hospital, Melbourne, Victoria, Australia

<sup>13</sup>Department of Neurology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

<sup>14</sup>Department of Neurology, Hospital Universitari Germans Trias i Pujol, Barcelona, Spain, Badalona, Spain

<sup>15</sup>Department of Medicine, University of Melbourne, Parkville, Victoria, Australia

<sup>16</sup>Department of Neurology, Royal Melbourne Hospital, Melbourne, Victoria, Australia

<sup>17</sup>Department of Neurology, UCLA, Los Angeles, California, USA

<sup>18</sup>Department of Neurology, Amsterdam UMC, location AMC, Amsterdam, the Netherlands

<sup>19</sup>Institute of Neuroscience & Psychology, University of Glasgow, Queen Elizabeth University Hospital, Glasgow, Scotland, UK

<sup>20</sup>Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

<sup>21</sup>Department of Neuroradiology, Newcastle upon Tyne Hospitals, Newcastle upon Tyne, UK

<sup>22</sup>CIC1433-Epidémiologie Clinique, Inserm, Centre Hospitalier Régional et Universitaire de Nancy, Université de Lorraine, Nancy, France

<sup>23</sup>Department of Radiology and Nuclear Medicine, Amsterdam UMC, location AMC, Amsterdam, the Netherlands

**Twitter** Michael D Hill @mihill68

**Contributors** All authors have made a substantial contribution to all categories established by the ICMJE guidelines on authorship.

**Funding** This study was funded by ITEA3 (grant number: 10004).

**Competing interests** RSB, AMMB, HAM, and CBLMM are cofounders and shareholder of Nico Laboratory. EP is a shareholder of Nico Laboratory.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Wim H van Zwam <http://orcid.org/0000-0003-1631-7056>

Mayank Goyal <http://orcid.org/0000-0001-9060-2109>

Michael D Hill <http://orcid.org/0000-0002-6269-1543>

Henk A Marquering 0000-0002-1414-6313

#### REFERENCES

- Berkhemer OA, Kamalian S, González RG, *et al*. Imaging biomarkers for intra-arterial stroke therapy. *Cardiovasc Eng Technol* 2013;4:339–51.
- Warach SJ, Luby M, Albers GW, *et al*. Acute stroke imaging research roadmap III imaging selection and outcomes in acute stroke reperfusion clinical trials. *Stroke* 2016;47:1389–98.
- Ernst M, Boers AMM, Aigner A, *et al*. Association of computed tomography ischemic lesion location with functional outcome in acute large vessel occlusion ischemic stroke. *Stroke* 2017;48:2426–33.
- Doyle S, Forbes F, Jaillard A. Sub-acute and Chronic Ischemic Stroke Lesion MRI Segmentation. In: *Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries*. Springer International Publishing, 2018: 111–22.
- Zhang R, Zhao L, Lou W, *et al*. Automatic segmentation of acute ischemic stroke from DWI using 3-D fully convolutional DenseNets. *IEEE Trans Med Imaging* 2018;37:2149–60.
- Kamnitsas K, Ledig C, Newcombe VFJ, *et al*. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* 2017;36:61–78.
- Maier O, Schröder C, Forkert ND, *et al*. Classifiers for ischemic stroke lesion segmentation: a comparison study. *PLoS One* 2015;10:e0145118.
- Goyal M, Menon BK, van Zwam WH, *et al*. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *The Lancet* 2016;387:1723–31.
- Boers AMM, Jansen IGH, Brown S, *et al*. Mediation of the relationship between endovascular therapy and functional outcome by follow-up infarct volume in patients with acute ischemic stroke. *JAMA Neurol* 2019;76:194.
- Berge JK, Bergman RA. Variations in size and in symmetry of foramina of the human skull. *Clin Anat* 2001;14:406–13.
- Barros RS, der SWEvan, Boers AMM, *et al*. (Submitted) automated detection and segmentation of subarachnoid hemorrhages with Convolutional neural networks. *Radiol Artif Intell* 2019;In press.

- 12 Association AP. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 1995;6:284–90.
- 13 Bivard A, Levi C, Lin L, *et al*. Validating a predictive model of acute advanced imaging biomarkers in ischemic stroke. *Stroke* 2017;48:645–50.
- 14 Liu L, Yang S, Meng L. Multi-scale Deep Convolutional Neural Network for Stroke Lesions Segmentation on CT Images. In: Crimi A, Bakas S, Kuijf H, eds. *Brain lesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Cham: Springer International Publishing, 2019: 283–91.
- 15 Bardera A, Boada I, Feixas M, *et al*. Semi-automated method for brain hematoma and edema quantification using computed tomography. *Comput Med Imaging Graph* 2009;33:304–11.
- 16 Kuang H, Najm M, Menon BK. Joint Segmentation of Intracerebral Hemorrhage and Infarct from Non-Contrast CT Images of Post-treatment Acute Ischemic Stroke Patients. In: Frangi AF, Schnabel JA, Davatzikos C, eds. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2018*. . Springer International Publishing, 2018: 9901. 681–8.
- 17 Boers AM, Marquering HA, Jochem JJ, *et al*. Automated cerebral infarct volume measurement in follow-up noncontrast CT scans of patients with acute ischemic stroke. *AJNR Am J Neuroradiol* 2013;34:1522–7.
- 18 Vos PC, Weaver NA, Biesbroek JM. Automatic detection and segmentation of ischemic lesions in computed tomography images of stroke patients. In: *Medical imaging 2013: computer-aided diagnosis*. 8670, 2013.
- 19 Gillebert CR, Humphreys GW, Mantini D. Automated delineation of stroke lesions using brain CT images. *Neuroimage Clin* 2014;4:540–8.