UvA-DARE (Digital Academic Repository)

Automated final lesion segmentation in posterior circulation acute ischemic stroke using deep learning

Zoetmulder, R.; Konduri, P.R.; Obdeijn, I.V.; Gavves, E.; Išgum, I.; Majoie, C.B.L.M.; Dippel, D.W.J.; Roos, Y.B.W.E.M.; Goyal, M.; Mitchell, P.J.; Campbell, B.C.V.; Lopes, D.K.; Reimann, G.; Jovin, T.G.; Saver, J.L.; Muir, K.W.; White, P.; Bracard, S.; Chen, B.; Brown, S.; Schonewille, W.J.; van der Hoeven, E.; Puetz, V.; Marquering, H.A.

Article

Automated Final Lesion Segmentation in Posterior Circulation Acute Ischemic Stroke Using Deep Learning

Riaan Zoetmulder [1,2,3], Praneeta R. Konduri [1,2], Iris V. Obdeijn [1], Efstratios Gavves [3], Ivana Išgum [1,2,3], Charles B.L.M. Majoie [2], Diederik W.J. Dippel [4], Yvo B.W.E.M. Roos [5], Mayank Goyal [6,7], Peter J. Mitchell [8], Bruce C. V. Campbell [9], Demetrius K. Lopes [10], Gernot Reimann [11], Tudor G. Jovin [12], Jeffrey L. Saver [13], Keith W. Muir [14], Phil White [15,16], Serge Bracard [17], Bailiang Chen [18], Scott Brown [19], Wouter J. Schonewille [20], Erik van der Hoeven [21], Volker Puetz [22] and Henk A. Marquering [1,2,*]

1   Department of Biomedical Engineering and Physics, Amsterdam UMC, Location AMC, 1105 Amsterdam, The Netherlands; r.zoetmulder@amsterdamumc.nl (R.Z.); p.r.konduri@amsterdamumc.nl (P.R.K.); i.v.obdeijn@student.utwente.nl (I.V.O.); i.isgum@amsterdamumc.nl (I.I.)
2   Department of Radiology and Nuclear Medicine, Amsterdam UMC, Location AMC, 1105 Amsterdam, The Netherlands; c.b.Majoie@amsterdamumc.nl
3   Informatics Institute, University of Amsterdam, 1097 Amsterdam, The Netherlands; E.Gavves@uva.nl
4   Department of Neurology, Erasmus MC University Medical Center, 3015 Rotterdam, The Netherlands; d.dippel@erasmusmc.nl
5   Department of Neurology, Amsterdam UMC, Location AMC, 1105 Amsterdam, The Netherlands; y.b.roos@amsterdamumc.nl
6   Radiology, Foothills Medical Centre, University of Calgary, Calgary, AB T2N 2T9, Canada; mgoyal@ucalgary.ca
7   Department of Clinical Neurosciences, Hotchkiss Brain Institute, University of Calgary, Calgary, AB T2N 4N1, Canada
8   Department of Radiology, The University of Melbourne & The Royal Melbourne Hospital, Melbourne, VIC 3050, Australia; Peter.Mitchell@mh.org.au
9   Department of Medicine and Neurology, Melbourne Brain Centre at the Royal Melbourne Hospital, University of Melbourne, Melbourne, VIC 3052, Australia; dr.bruce.campbell@gmail.com
10 Department of Neurological Surgery, Rush University Medical Center, Chicago, IL 60612, USA; brainaneurysm@mac.com
11 Department of Neurology, Community Hospital Klinikum Dortmund, 44137 Dortmund, Germany; Gernot.Rudel@klinikumdo.de
12 Cooper Neurological Institute, Cooper University Medical Center, Camden, NJ 08103, USA; tudorjovin@gmail.com
13 Department of Neurology and Comprehensive Stroke Center, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA; jsaver@mednet.ucla.edu
14 Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QB, UK; keith.muir@glasgow.ac.uk
15 Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK; phil.white@newcastle.ac.uk
16 Department of Neuroradiology, Newcastle upon Tyne Hospitals, Newcastle upon Tyne NE1 4LP, UK
17 INSERM U1254, IADI, University Hospital, Neuroradiology, 54511 Nancy, France; s.bracard@chru-nancy.fr
18 INSERM CIC-IT 1433, University Hospital, 54511 Nancy, France; B.CHEN@chru-nancy.fr
19 Altair Biostatistics, St Louis Park, MN 55416, USA; b_scott_brown@yahoo.com
20 St. Antonius Hospital, 3435 Nieuwegein, The Netherlands; w.schonewille@antoniusziekenhuis.nl
21 Department of Radiology, St. Antonius Hospital, P.O. Box 2500, 3430 Nieuwegein, The Netherlands; e.van.der.hoeven@antoniusziekenhuis.nl
22 Department of Neurology, Dresden University Stroke Centre, Technical University Dresden, Fetscherstraße 74, 01307 Dresden, Germany; volker.puetz@neuro.med.tu-dresden.de
*   Correspondence: h.a.marquering@amsterdamumc.nl

Abstract: Final lesion volume (FLV) is a surrogate outcome measure in anterior circulation stroke (ACS). In posterior circulation stroke (PCS), this relation is plausibly understudied due to a lack of methods that automatically quantify FLV. The applicability of deep learning approaches to PCS is limited due to its lower incidence compared to ACS. We evaluated strategies to develop a convolutional neural network (CNN) for PCS lesion segmentation by using image data from both ACS

and PCS patients. We included follow-up non-contrast computed tomography scans of 1018 patients with ACS and 107 patients with PCS. To assess whether an ACS lesion segmentation generalizes to PCS, a CNN was trained on ACS data (ACS-CNN). Second, to evaluate the performance of only including PCS patients, a CNN was trained on PCS data. Third, to evaluate the performance when combining the datasets, a CNN was trained on both datasets. Finally, to evaluate the performance of transfer learning, the ACS-CNN was fine-tuned using PCS patients. The transfer learning strategy outperformed the other strategies in volume agreement with an intra-class correlation of 0.88 (95% CI: 0.83–0.92) vs. 0.55 to 0.83 and a lesion detection rate of 87% vs. 41–77 for the other strategies. Hence, transfer learning improved the FLV quantification and detection rate of PCS lesions compared to the other strategies.

Keywords: posterior stroke; segmentation; transfer learning; deep learning; CT

1. Introduction

Infarct volume, representing the tissue damage after an acute ischemic stroke (AIS), is commonly considered as a surrogate endpoint for the primary functional outcome (modified Rankin Scale (mRS) after 90 days) [1]. Various trials have shown a strong association of final lesion volume (FLV) with functional outcome in patients suffering from a stroke due to a large vessel occlusion in the anterior circulation [1,2].

However, in patients with a stroke due to a posterior circulation stroke (PCS), the relation between FLV and outcome is understudied [3]. The low number of studies addressing this relation may be due to the combination of two reasons; the relatively low prevalence of PCS compared to anterior circulation stroke (ACS) and the lack of automated analysis of PCS lesion volume assessment.

With the huge effectiveness of endovascular treatment of anterior circulation stroke patients, treatment of posterior stroke has attained renewed interest in various studies and trials. For example, the recently completed BASICS trial [4] could not show a beneficial effect of endovascular treatment with functional outcome used as an outcome measure. Alternatively, secondary outcome measures such as FLV might show a beneficial effect of certain treatments since functional outcome, as addressed by the mRS, is a rather coarse outcome measure, which is also affected by many other confounders [5]. Developing methods that automatically segment lesions due to a posterior circulation stroke (PCS) would help investigate FLV as a surrogate outcome for this type of stroke.

Solutions for the automatic segmentation of FLV based on convolutional neural networks (CNNs) have been presented in the literature for CT and MR imaging [6,7]. However, these studies have only considered the FLV of patients with an AIS due to an occlusion of the anterior circulation [1,6]. To achieve good performance, CNNs typically require large amounts of labeled training data. However, PCS constitutes only 26% of AIS cases [8,9], and thus, training of CNNs for automatic PCS lesion segmentation is hindered by the limited availability of data. Furthermore, the applicability of methods developed for ACS FLV segmentation on posterior stroke lesion segmentation is unknown.

Several methods exist for dealing with a lack of data to train a CNN. One method that reduces the data needed to train CNNs by reusing knowledge is transfer learning [10]. To perform transfer learning, a CNN is pre-trained on a task for which large amounts of image data are available and fine-tuned on a different task for which little image data are available. Transfer learning has been successfully applied to solve various medical image analysis problems [11–14].

We evaluate strategies to create automated PCS lesion segmentation by using image data from patients with ACS and patients with PCS. We hypothesize that transfer learning utilizing data of ACS lesions improves automatic PCS lesion segmentation performance compared to alternative strategies: training a CNN on only ACS lesions, only on PCS lesions, or on the combination of ACS and PCS lesions.

2. Materials and Methods

2.1. Patient and Image Data

All involved patients in this retrospective study or their legal representatives provided written informed consent. The medical ethics committee of each participating hospital approved the use of the data after anonymization.

The Hermes dataset consists of 1665 patients who suffered from an ACS and was obtained from the HERMES collaboration [15], which investigated the effectiveness of endovascular therapy for treating ACS. This collaboration combined data from seven clinical randomized trials and collected data between December 2010 and December 2014. The scans contain a varying number of slices, which each have 512 rows and columns. The inclusion criteria are shown in Figure A1A. Patients were excluded if no follow-up non-contrast computed tomography (FU-NCCT) was made in the time window between 12 h and 2 weeks after stroke onset or if the preprocessing steps were unsuccessful. In total, 1018 patients out of the 1665 patients were included. Baseline characteristics of the included patients are shown in Table 1.

Table 1. Baseline characteristics, treatment and time data for patients with posterior circulation stroke and anterior circulation stroke. Prior to posterior stroke, transient ischemic attack (TIA), posterior circulation TIA, and atrial fibrillation (history or 12 lead electrocardiogram (ECG)) were not available (NAV) for the HERMES dataset. We use abbreviations for the National Institutes of Health Stroke Scale (NIHSS), modified Ranking Scale (mRS) and, Intravenous Thrombolysis (IVT).

Parameter	Posterior Stroke	Anterior Stroke
Clinical		
Age, years, mean (Standard Deviation)	65.65 (12.2)	66.1 (13.3)
Sex, F, No. [%]	34/107 [31.8]	458/1018 [45]
NIHSS at baseline, mean [median] (N)	21.4 [19] (107)	17 [17] (1015)
Prior Conditions		
Diabetes mellitus, No. [%]	28/107 [26.2]	169/1018 [16.6]
Hypertension, No. [%]	64/107 [59.8]	564/1018 [55.4]
Stroke, No. [%]	21/107 [19.6]	121/1018 [11.9]
Posterior circulation stroke, No. [%]	7/107 [6.5]	NAV
TIA, No. [%]	10/106 [9.4]	NAV
Posterior circulation TIA, No. [%]	2/106 [1.9]	NAV
Atrial fibrillation, No. [%]	13/107 [12.1]	314/1018 [30.8]
Atrial fibrillation (history or 12 lead ECG), No. [%]	23/107 [21.5]	NAV
Pre-Stroke mRS		
0, No. [%]	80/107 [74.8]	836/1017 [82.1]
1, No. [%]	11/107 [10.3]	129/1017 [12.7]
2, No. [%]	13/107 [12.1]	29/1017 [2.9]
3, No. [%]	3/107 [2.8]	23/1017 [2.3]
Treatment		
IVT, No. [%]	92/107 [86]	872/1018 [85.7]
Time		
Stroke onset to IVT, min., mean [Standard Deviation] (N)	176.9 [176.102] (90)	112.2 [57.2] (871)

The BASICS dataset consists of 168 patients who suffered from a PCS and was obtained from the BASICS trial [16,17], which investigated the effectiveness of endovascular therapy for treating patients with a PCS. This trial included patients from 23 centers, and the data were collected between 2011 and 2019. The scans are composed of a varying number of slices that each have 512 rows and columns. Imaging parameters are shown in Table A1. Inclusion criteria for our study are shown in Figure A1B.

Infarcts evolve over time, and the most recent FU-NCCT scans imaged the most evolved infarcts with the lowest densities. As a result of the more distinct presence of the infarcts in later scans, these scans are commonly used for the infarct volume assessment in clinical practice and clinical research. Hence, the latest FU-NCCT scan was used if multiple scans were available for the same patient. Patients were excluded if no FU-NCCT was made or if the follow-up image was of insufficient quality.

In total, 107 patients out of the 168 available patients were included. The mean and median time between symptom onset and the most recent follow-up NCCT are respectively, 98.5 h and 28 (IQR: 24–32). Baseline characteristics of the included BASICS patients are shown in Table 1. The infarcted regions, as expressed in affected PC-ASPECTS regions, are shown in Table A2 in the Appendix A.

2.2. Reference Segmentations

For patients with an ACS, reference segmentations were obtained by manual annotation by one of two experienced observers on the most recent FU-NCCT. The annotation procedure is outlined in [1]. In summary, a window width of 30 Hounsfield Units (HU) and a center level of 35 HU was set in ITK-Snap [18]. All hypo-dense regions on the ipsilateral hemisphere including edema were included in the segmentations. Infarcted tissue in the ipsilateral hemisphere with signs of an old infarct were excluded from the reference segmentations. Parenchymal hemorrhages adjacent to or within the affected area were included in the reference segmentations. Finally, reference segmentations were checked and, if necessary, corrected by one of three radiologists, each of whom had more than 5 years of experience.

Reference segmentations of lesions caused by a PCS on FU-NCCT scans were manually created by a single trained observer (IVO) and were checked by an experienced radiologist (CBLMM) who has more than 15 years of experience. Lesions were segmented by using the aforementioned window width and center level using ITK-Snap software [18]. PC-ASPECTS scores [19], which were scored by two experienced radiologists, were used when available to identify the infarcted territory.

2.3. Preprocessing

The intracranial region as a volume of interest was obtained automatically using a combination of preprocessing steps [6]. The bone was segmented using a threshold-based segmentation by selecting all voxels with an intensity of 170 HU or higher. Subsequently, the foramina, except the foramen magnum, were closed using morphological filters, and a region-growing algorithm was applied to select the intracranial volume. To obtain the final volume of interest, the region caudal to the foramen magnum was excluded.

To ensure the same size, orientation, and voxel sizes, all scans were aligned by automatically registering the images to a common space using rigid and affine transformations. Images were registered using the Mattes Mutual Information [20] with a gradient descent optimizer. In addition to registration, the scans were downsampled to allow the entire scan to be passed into the CNNs. After the preprocessing, each scan had a size of $256 \times 256 \times 32$, with a slice thickness of 5 mm.

The voxel intensities were clipped between -20 and 120 HU and subsequently normalized between minus one and one. The preprocessing was done using SimpleITK [21,22] and Python 2.7.

2.4. CNN for Automatic Posterior Circulation Lesion Segmentation

The preprocessed images were input to a CNN which consisted of three-dimensional convolutional kernels. The architecture of the CNN (shown in Figure A2A in the Appendix A) was inspired by U-Net [23] and ResNet [24]. The CNN consisted of a downsampling path and an upsampling path. The downsampling path started with an input block, consisting of a convolution with a kernel size of three and a stride of one, which was followed by a max

pooling layer with a pooling size of two and a stride of two. The dimensions of the input were $256 \times 265 \times 32$, and the input block generated features that consisted of eight channels.

Subsequent to the input block, three downsampling blocks were added, consisting of three 3D ResNet layers, as shown in Figure A2B in the Appendix A. The first two blocks were followed by average pooling with a stride and pooling size of two.

The upsampling path started with a transposed convolution using a stride of two. Next, two upsampling blocks followed by an output layer were added. Each upsampling block consisted of two ResNet layers, which were followed by a transposed convolution with a stride of two and a kernel size of three. Each upsampling block took the featu res from the previous block and the corresponding downsampling block and concatenated them. The output block consisted of two ResNet blocks followed by a convolutional layer. The CNNs were implemented using Tensorflow 1.5.

2.5. Experimental Setup

Four different training strategies for CNNs were evaluated: A CNN was randomly initialized and trained on images of patients in the HERMES dataset (ACS-CNN), BASICS dataset (PCS-CNN), and the HERMES and BASICS datasets combined (CD-CNN). The ACS-CNN was used to establish the generalization ability of a CNN trained on ACS to PCS lesion segmentation. The PCS-CNN served as a baseline for training with limited but representative data. The CD-CNN was used as a benchmark if both ACS and PCS data were available, but no transfer learning was used. Transfer learning reused the weights from the trained ACS-CNN to initialize all but the last block of the CNN, and it fine-tuned by updating all the weights, using images from the BASICS dataset (TL-CNN).

The CNNs used group normalization with four groups, the Leaky ReLU activation function, a batch size of two, and the Adam optimizer. The loss function used was the weighted binary cross-entropy. The initial learning rate for the ACS-CNN and CD-CNN was 10^{-3} and was decayed stepwise after 5, 10, 15, and 20 epochs to respectively, 5×10^{-4}, 2×10^{-4}, 10^{-4}, and 10^{-5}. These networks were trained for 25 epochs. The initial learning for the PCS-CNN and TL-CNN was 10^{-5} and was decayed after 25, 50, and 75 epochs to respectively, 5×10^{-6}, 2×10^{-6}, and 10^{-6}. These networks were trained and fine-tuned for 100 epochs. The weight decay was set to 10^{-5}.

Data augmentation was applied at training time. The images were rotated at a randomly chosen angle between zero and ten degrees along the axial plane in either direction or were randomly flipped along the sagittal plane.

We evaluated the performance of the ACS-CNN to check whether the model converges during pre-training. The ACS dataset was split randomly into scans for training (85%), validation (5%), and testing (10%). For this approach, we used the entire PCS dataset for evaluation. Thus, we used stratified 5-fold cross-validation. Given the number of available PCS patients, the first four testing splits consisted of 20% and the fifth consisted of 22% of the data. The training splits were of equal size and consisted of 78% of the PCS data.

2.6. Evaluation

The reliability between the automatically and manually segmented volumes was evaluated with the intraclass correlation coefficient (ICC) including the 95% Confidence Interval (95% CI). The ICC was interpreted in accordance to the American Psychological Association [25]. Following their guidelines, an ICC < 0.4 is defined as poor, an ICC between 0.4 and 0.6 is defined as fair, an ICC between 0.6 and 0.75 is defined as good, and an ICC greater than 0.75 is defined as excellent. In addition, a Bland–Altman analysis was performed to assess the bias and limits of agreement (LoA) in volume measurements. Statistical significance between ICCs was evaluated by using Fisher's *r*-to-*z* transformation.

We determined whether our model accurately detected lesions independent of size. Thus, we calculated the ratio of the total number of correctly detected lesions and the total number of lesions as determined by the ground truth in the dataset. A lesion was defined as detected if the percentage of overlapping voxels between the automatic and

reference segmentations was larger than a predefined minimum. However, in case of small thresholds, a non-zero overlap of automated and reference lesion segmentation could be caused by chance. To account for this issue, the required minimum percentage of overlapping voxels to count a lesion as detected was set to be either greater than zero percent or greater than a more conservative 20%. Next, the effect of lesion volume on the lesion detection rate was studied by excluding lesions of a progressively larger volume in the reference segmentations. This latter cutoff was set between zero and 4 mL with increments of 0.5 mL.

The segmentation performance of the automatic methods was evaluated by calculating the Dice coefficient as an overlap measure between the reference and the automatic segmentation. Normality of the distribution of the Dice coefficients was assessed before pairwise statistical testing by using the Shapiro–Wilk test. If the Dice coefficients were normally distributed, a paired t-test was used; otherwise, a Wilcoxon rank sum test was used. *p*-values were corrected for the family wise error rate using the Bonferroni correction. All statistical testing was done using the python library Pingouin, v.0.3.1 [26] (University of California, Berkeley, CA, USA).

3. Results

Baseline characteristics of the patients in BASICS and HERMES datasets were compared. Patients in the HERMES dataset had a similar age to patients in the BASICS dataset. Diabetes (26.2% vs. 16.6%, $p < 0.05$) and prior stroke (19.6% vs. 11.9%, $p < 0.05$) occurred more frequently in patients in the BASICS dataset. However, atrial fibrillation (12.1% vs. 30.8%, $p < 0.01$) occurred more frequently in patients in the HERMES dataset.

The median FLV in patients with PCS was 11 (IQR: 3.4–36) mL. The ICCs for volume assessments for the TL-CNN, PCS-CNN, CD-CNN, and ACS-CNN were 0.88 (95% CI: 0.83–0.92), 0.80 (95% CI: 0.72–0.86), 0.83 (95% CI: 0.76–0.88), and 0.55 (95% CI: 0.4–0.67), respectively. The ICC of the TL-CNN was significantly larger than the ICCs of the ACS-CNN ($p < 0.01$) and PCS-CNN ($p = 0.02$). The ICC of the ACS-CNN was significantly smaller than the ICCs of the PCS-CNN ($p < 0.01$) and CD-CNN ($p < 0.01$). In addition, the Bland–Altman analysis for each of the CNNs resulted in biases ranging from 0.8 mL for the TL-CNN to 13.5 mL for the ACS-CNN. The LoAs were the smallest for the TL-CNN with −29 to 30 mL and largest for the ACS-CNN with −32 to 59 mL. The bias and LoAs of the volume measurements are shown in Figure 1 and, Table A3 in the Appendix A.

The lesion detection rate of the TL-CNN was higher than for the other learning strategies (Figure 2), which shows that the lesion detection rate increases with increasing lesion volume and that the lesion detection rate decreases with increasing thresholds of overlapping voxels.
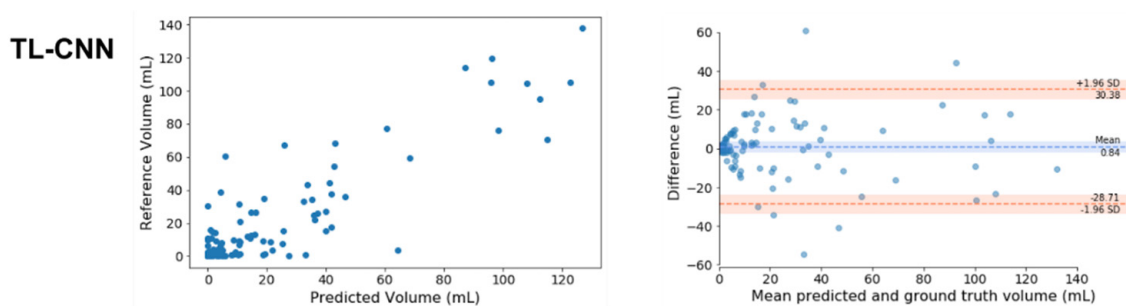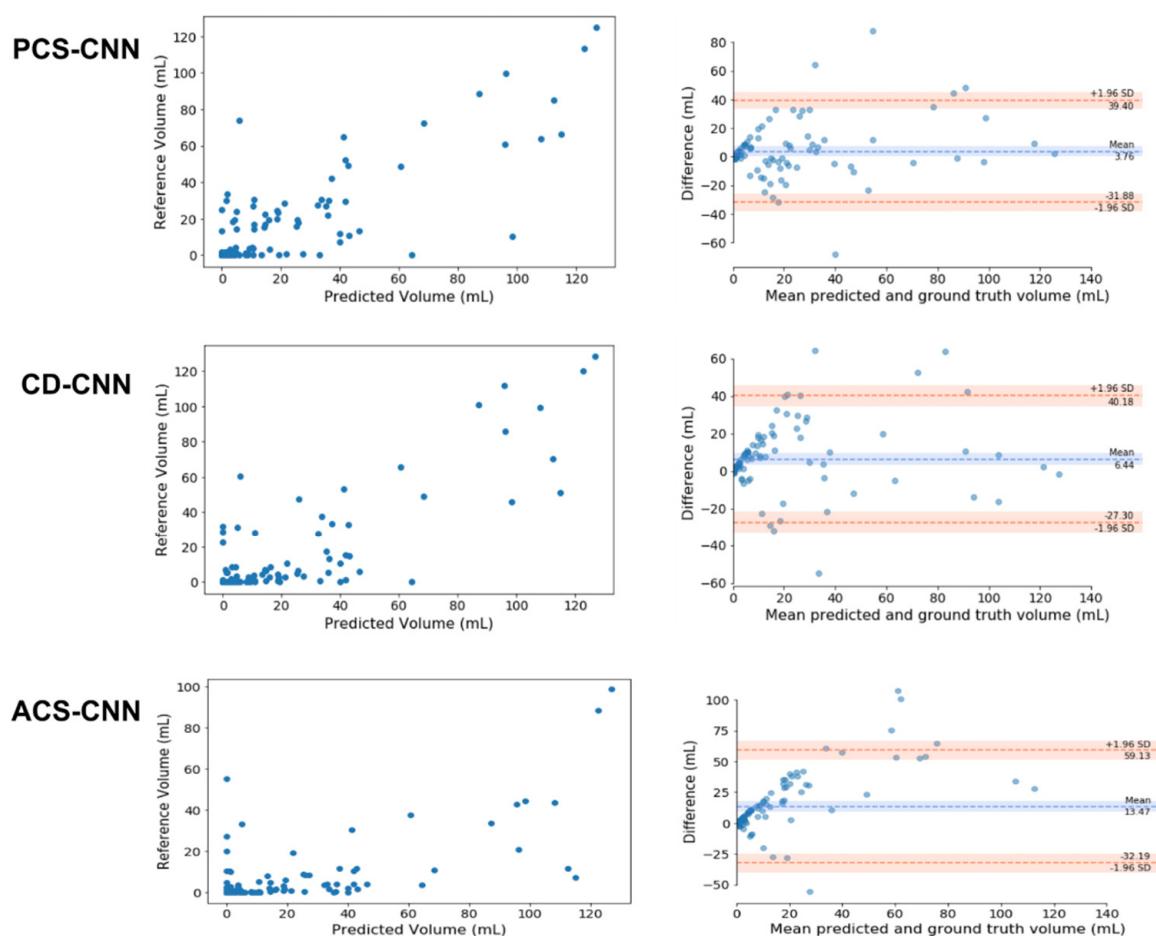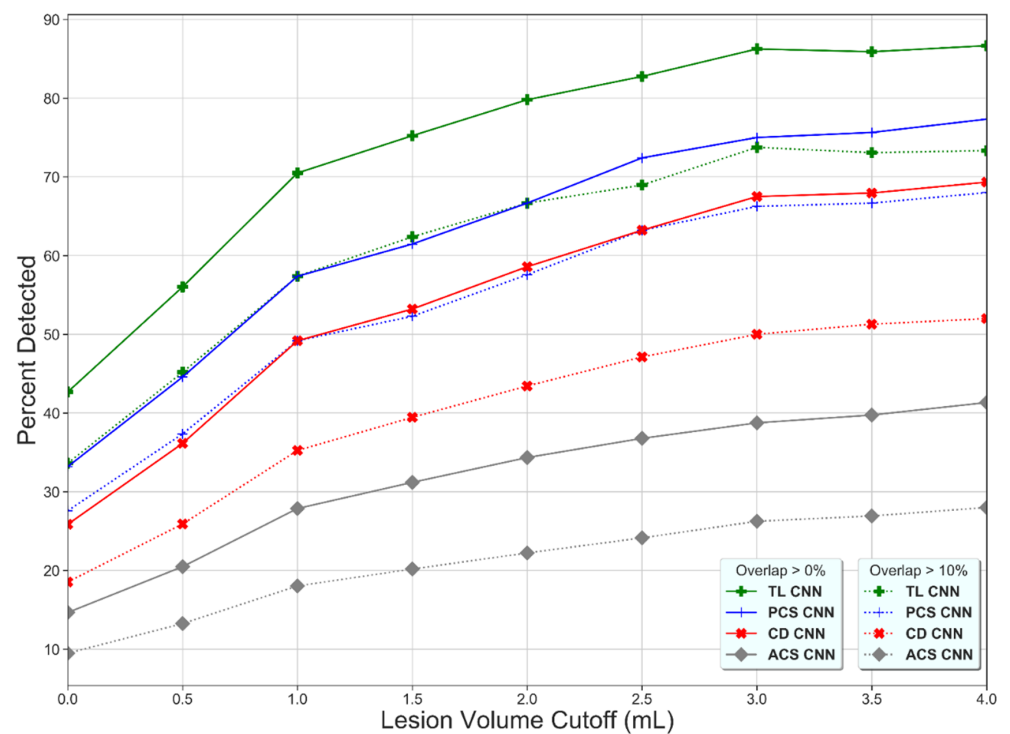

Figure 1. *Conts.*

Figure 1. Comparison of the automated and reference segmentation volume for the Transfer Learned CNN (TL-CNN), Posterior Circulation Stroke CNN (PCS-CNN), Combined Datasets CNN (CD-CNN), and Anterior Circulation Stroke CNN (ACS-CNN). Left column: Scatter plots comparing lesion volumes derived from the reference segmentations (*y*-axis) and from the automatic segmentations determined by the CNN (*x*-axis). Right column: Bland–Altman plots of the lesion volumes. The volumes corresponding to the reference and automatic segmentations are shown on the *x*-axis and the volume difference is shown on the *y*-axis.

The TL-CNN, PCS-CNN, CD-CNN, and ACS-CNN achieved a Dice coefficient of 0.25 ± 0.08, 0.21 ± 0.06, 0.16 ± 0.06, and 0.07 ± 0.03, on the overall PCS test set, respectively (Figure 3). The Dice coefficients were not normally distributed. Hence, a Wilcoxon rank sum test was used. Results of the Wilcoxon rank sum test are shown in Table A4. For the anterior circulation stroke lesions, the ACS-CNN achieved an average Dice coefficient of 0.60 ± 0.07.

Figure 2. Percentage of detected lesions (*y*-axis) as a function of the minimum volume requirement (*x*-axis) and the minimum percentage of overlapping voxels for the Transfer Learned CNN (TL-CNN), Posterior Circulation Stroke CNN (PCS-CNN), Combined Datasets CNN (CD-CNN), and Anterior Circulation Stroke CNN (ACS-CNN) (green, blue, red, and gray lines). For all methods, a higher lesion volume cutoff results in a higher percentage of detected lesions. The lower overlapping voxel requirement, the higher the percentage of detected lesions (dotted versus solid lines).
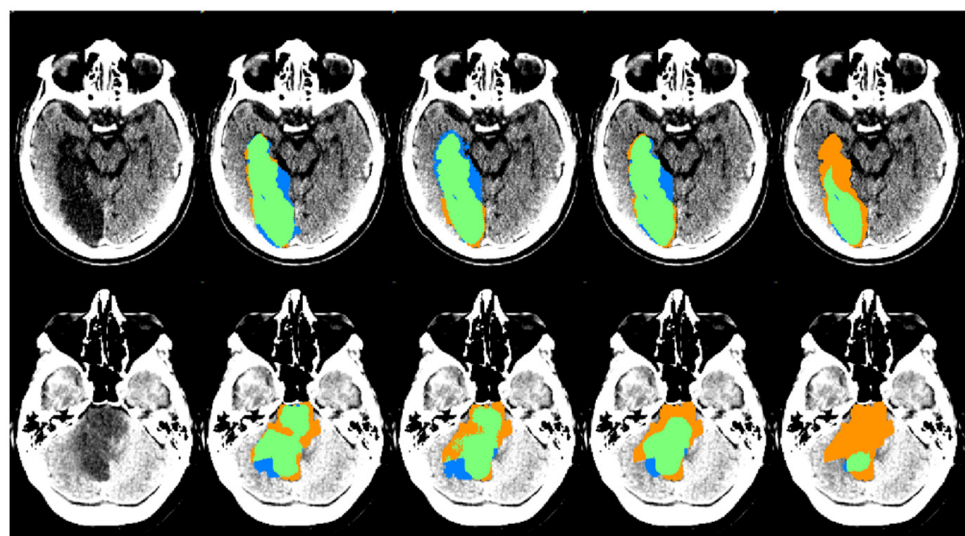
Figure 3. *Conts.*
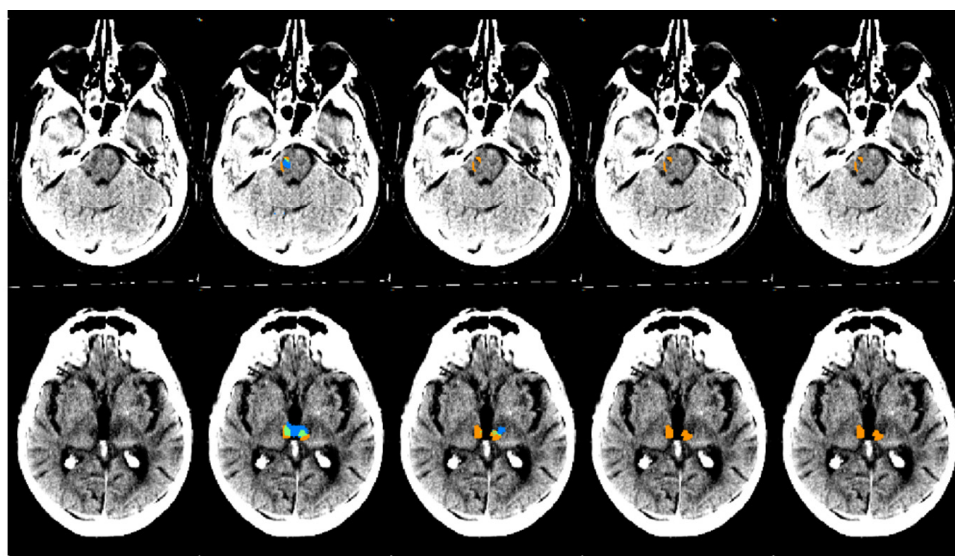
Figure 3. An example of automatic segmentation results obtained by the four CNNs on the PCS test set. From the left to the right column: the original scan, the automatic segmentation results from the TL-CNN, PCS-CNN, CD-CNN, and ACS-CNN are shown. The segmentation maps show true positives (green), false positives (blue), and false negatives (orange). The scans were plotted using a window center around 35, with a window width of 30.

4. Discussion

Our study found that transfer learning results in a high level of agreement between manually delineated and automatically quantified lesion volumes on the follow-up NCCTs of patients with a PCS. Furthermore, we found that transfer learning resulted in higher spatial accuracy and larger volume agreement of automatic PCS lesion segmentation compared to the other strategies. In addition, the TL-CNN models also detected a larger number of PCS lesions in comparison to the other strategies. Moreover, our results indicate that the ACS-CNN models, which were trained on only patients who suffered from an ACS, do not generalize to PCS lesion segmentation.

Our study is the first to address the automated segmentation of posterior circulation stroke lesions on sub-acute follow-up NCCT. Previous work include anterior stroke lesion segmentation in a variety of imaging modalities, such as baseline CTP [27], baseline CTA [28], follow-up NCCT [6], and baseline and follow-up DWI [29] using multiple approaches and addressing various types of stroke.

A previous study focused on developing a CNN-based method for automatic ACS lesion segmentation on FU-NCCT [6] with a higher spatial overlap accuracy than found in our study. This could be explained by the larger lesion volumes in their population (median FLV of 48 vs. 11 mL) and the larger dataset available for training. Furthermore, beam-hardening artifacts are common in the posterior fossa on NCCT. Beam-hardening artifacts may obscure lesions caused by a posterior circulation stroke, making these lesions harder to detect than lesions caused by anterior circulation stroke. Finally, the Dice coefficient is a global overlap metric, which is valuable when comparing large delineations. If the objects of interest are smaller, such as for the lesions in our population, the Dice coefficient may be too sensitive to small errors. Therefore, the Dice coefficient must be interpreted in addition to other metrics, such as the lesion detection rate and the correspondence in lesion volume, as expressed with the ICC between the automatically quantified and reference volumes, and not in isolation. The latter measure, lesion volume, is also clinically a more relevant measure than the spatial overlap. If we put the performance of our algorithms in context using the ICC, it shows that our algorithm can adequately predict FLV, which has a strong association with functional outcome in clinical practice. In addition, the bias and LoAs of the FLV were comparable to other studies assessing infarct volumes for similar problems.

One study found a bias between the automatically and manually segmented FLV on FU-NCCT of ACS patients with subtle hypo-densities of 29 mL with LoA of −91 to 149 mL, which are both considerably larger compared to our study [6]. Another study found a bias between the automatically and manually segmented FLV on baseline NCCT of ACS patients of 11 mL with LoA of −59 to 80 mL [30]. Our algorithm, which utilized transfer learning, resulted in a bias of 0.84 mL and LoA of −29 to 30 mL, showing that, on average, the bias in FLV is small. A weakness, as in many other studies, is the large LoA values in our study. Since the infarct volumes are clinically mostly used in prognostic models, it is unknown to what extent this patient-specific variation in FIV influences prognosis. However, the large LoAs suggest that patient-specific analysis still has to be improved before the algorithm could be applied in a clinical setting.

Other methods for automatic ACS lesion segmentation used information from the contralateral hemisphere to improve segmentation accuracy [28,31]. PCS lesions can affect both hemispheres; hence, comparing information between the ipsilateral and its contralateral hemisphere is unlikely to improve the accuracy.

Automated stroke lesion segmentation has also been developed for chronic stroke lesions on T1 MRI. Chronic stroke lesions have been segmented by using a random forest classifier to segment lesions in the left hemisphere by using hand-crafted image features [32]. Another study used a deep residual network to segment lesions on images of the ATLAS dataset, which contains manually traced lesions on 304 T1-weighted MRI images [33]. Both studies achieved higher similarity scores than our method, which could be explained by the larger FLVs and higher sensitivity provided by T1-weighted MRI images.

In previous research, transfer learning has also been successfully applied to improve the accuracy of various other medical image segmentation tasks. One study used CNNs pre-trained on eight different medical image segmentation tasks on various imaging modalities to improve automatic lung, liver, and liver tumor segmentation [34]. Unlike the aforementioned study, our study pre-trained on a single imaging modality and task. Another study pre-trained CNNs using self-supervised tasks to improve lung nodule, liver, and brain tumor segmentation [35]. In agreement with our study, the results of prior work indicate that for medical image segmentation, transfer learning can be beneficial.

Other work using transfer learning for medical image tasks included CNNs pre-trained on ImageNet [36] as a benchmark. These studies used transfer learning to improve the performance on medical image analysis tasks on 2D images. However, using ImageNet for transfer learning was less likely to be suitable for our study, because prior work has shown that ImageNet pre-training improves performance on medical image analysis tasks less than using a pre-trained 3D model [35].

This study has several limitations. First, most of the FU-NCCT scans of PCS patients included in this study were obtained 24 h after the onset of AIS. Creating manual reference segmentations of the lesions on these early FU-NCCT scans is more challenging owing to the subtle differences in HU values after 24 h. In addition, FLV segmentations for patients with PCS were performed by only one trained observer who used PC-ASPECTS scores to verify the lesion location and was supervised by a radiologist (CBLMM). Since the assessment was only performed by a single observer, inter-observer agreement of the reference standard could not be assessed. Second, this study suffered from a low number of available PCS patients. Therefore, an even lower number of patients would be included if the data were divided into training, validation, and test sets. This would have lowered the generalizability of the results to the PCS patient population. To overcome this, five-fold cross-validation was used to allow all the data to be used as testing data in the analysis and to assess the stability of the presented results. Third, the CNNs were not accurate at detecting lesions with a volume smaller than 2 mL. If a patient is suspected of having lesions with a small volume, results from the presented algorithm should be verified by an expert evaluation. Transfer learning allowed the CNNs to reuse information learned from ACS lesion segmentation to segment lesions caused by PCS. The resulting improvement in PCS lesion segmentation is likely due to the similarity between the pre-training and

the fine-tuning tasks [37]. However, in our approach, the detection and segmentation of small lesions and the segmentation of lesions that are connected to cerebrospinal fluid-filled areas is still suboptimal. Fourth, it is quite likely that the swelling due to stroke has caused herniation in multiple patients. In our study, we did not exclude patients based on herniation, also because CT cannot distinguish infarcts resulting from the acute stroke versus herniation.

Deep learning is potentially valuable for automating demanding tasks in the quantification of radiological imaging. It is well-known that deep learning requires large amounts of data to train algorithms, which may suggest a limited applicability of deep learning in less common diseases. This study also shows that deep learning models that are trained on a more general, less specific disease may not be sufficient. Here, we presented an alternative approach based on transfer learning and showed that deep learning models can be pre-trained on similar diseases and fine-tuned on the specific rarer disease.

To conclude, the presented transfer learning approach improves the automatic detection and segmentation of posterior circulation stroke lesions compared to the evaluated commonly used training strategies. The presented automated posterior stroke lesion segmentation method allows the inclusion of lesion volume as an image outcome measure and as a metric to predict outcome in large-scale clinical trials and potentially as a first step toward clinical application.

Appendix A

Figure A1. (**A**) Flowchart showing the exclusion criteria used for the HERMES dataset. Exclusion criteria were: No FU-NCCT available, which was acquired between 12 h and 2 weeks (*n* = 637), an intracranial region segmentation error (*n* = 3), and a registration error (*n* = 7). In total, 1018 patients were included. These were split into a training set (*n* = 876), a validation set (*n* = 50), and a testing set (*n* = 101). (**B**) Flowchart showing the exclusion criteria used for the BASICS dataset. Exclusion criteria were: No FU-NCCT available (*n* = 55) or an image quality that was too low (*n* = 6). In total, 107 patients were included. These were split into five training and testing sets for the five-fold cross-validation. Patients could only belong to one test set.

Table A1. Imaging parameters used to acquire the NCCT scans of the included patients from the BASICS trial. Imaging parameters were missing for two patients.

Manufacturer	Model	Patients	Exposure Time (mS)	Exposure (mA)	Tube Current	kVp
Philips	iCT 256	29	1913 (IQR: 1025–1913)	375 (IQR: 251–490)	256 (IQR: 244–256)	100 (IQR: 100–120)
GE	LightSpeed VCT	17	1000 (IQR: 1000–2000)	175 (IQR: 171–175)	351 (IQR: 179–351)	120 (IQR: 120–120)
SIEMENS	SOMATOM Definition Flash	13	2000 (IQR: 2000–2000)	340 (IQR: 285–430)	172 (IQR: 143–229)	120 (IQR: 100–120)
TOSHIBA	Aquilion	11	750 (IQR: 750–750)	187 (IQR: 187–225)	250 (IQR: 250–300)	120 (IQR: 120–120)
SIEMENS	Sensation 64	6	1000 (IQR: 1000–1000)	380 (IQR: 380–380)	352 (IQR: 323–380)	120 (IQR: 120–120)
Philips	Brilliance 64	5	1678 (IQR: 1000–1678)	351 (IQR: 250–351)	209 (IQR: 200–209)	120 (IQR: 120–120)
SIEMENS	SOMATOM Force	4	2000 (IQR: 1750–2000)	340 (IQR: 327–380)	170 (IQR: 163.5–196)	100 (IQR: 100–105)
SIEMENS	SOMATOM Definition AS+	4	1000 (IQR: 1000–1000)	250 (IQR: 233–290)	138 (IQR: 129–155)	120 (IQR: 115–120)
Philips	IQon—Spectral CT	4	1117 (IQR: 1117–1117)	200 (IQR: 200–200)	179 (IQR: 179–179)	120 (IQR: 120–120)
Other	Various	12	1000 (IQR: 1000–1000)	260 (IQR: 158–353)	260 (IQR: 220–325)	120 (IQR: 120–120)
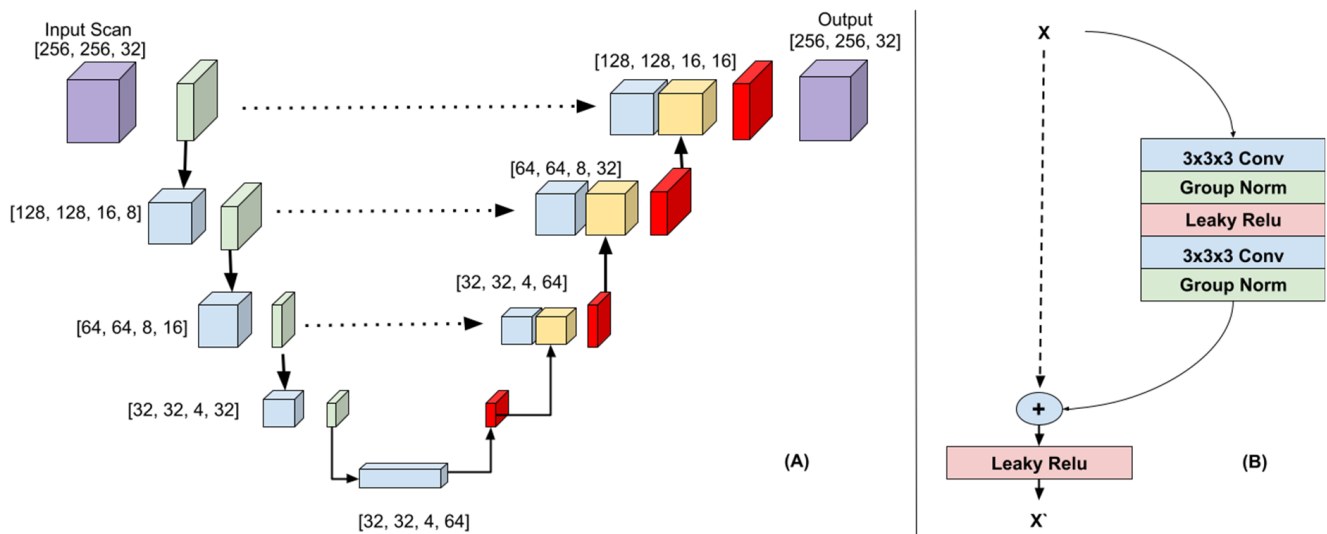
Figure A2. (**A**) 3D-Unet Architecture. The downsampling path (left) consisted of 3DResNet blocks with max pooling (green). The upsampling path (right) consisted of ResNet blocks followed by transposed convolutions (red). The features created in the downsampling path are colored blue and the features created in the upsampling path are colored yellow. The dotted arrows indicate the skip connections. The feature maps from the downsampling path were concatenated to the feature maps in the upsampling path. The input image and output probability map are colored purple. (**B**) 3D ResNet block.

Table A2. Lesion location in the posterior fossa, which was scored manually by using the PC-ASPECTS.

Lesion Location	Count/Total
No lesion	15/107
Left thalamus	33/107
Left cerebellum	40/107
Left PCA territory	19/107
Right thalamus	26/107
Right cerebellum	40/107
Right PCA territory	19/107
Midbrain	34/107
Pons	46/107
Other	4/107

Table A3. The ICC, Dice coefficients, bias, and limits of agreement between the automatically quantified and manually segmented volumes, respectively for the Transfer Learned (TL-CNN), Anterior Circulation Stroke (ACS-CNN), Combined Dataset (CD-CNN), and Posterior Circulation Stroke (PCS-CNN) Convolutional Neural Networks, tested on the PCS test set.

Method	ICC	Dice	Bias	Limits of Agreement
TL-CNN	0.88 (95% CI: 0.83–0.92)	0.25 ± 0.07	0.84 mL	−28.7 to 30.4 mL
PCS-CNN	0.80 (95% CI: 0.72–0.86)	0.21 ± 0.06	3.8 mL	−31.9 to 39.4 mL
CD-CNN	0.83 (95% CI: 0.76–0.88)	0.16 ± 0.06	6.4 mL	−27.3 to 40.2 mL
ACS-CNN	0.55 (95% CI: 0.4–0.67)	0.07 ± 0.03	13.5 mL	−32.2 to 59.1 mL

Table A4. Wilcoxon rank sum test on pairwise differences between the Dice coefficient and the bias of the volume differences. The W-statistic and *p*-value are shown in this table. The TL-CNN produced a significantly greater Dice coefficient than the other methods. The PCS-CNN produced a significantly greater Dice coefficient than the CD-CNN and ACS-CNN, and the CD-CNN produced a significantly greater Dice coefficient than the ACS-CNN. The TL-CNN produced a significantly smaller bias than the CD-CNN and ACS-CNN, the PCS-CNN produced a significantly smaller bias than the CD-CNN and ACS-CNN, and finally, the CD-CNN produced a significantly smaller bias than the ACS-CNN.

Method 1	Method 2	Dice Coefficient		Bias	
		W	*p*-Value	W	*p*-Value
TL-CNN	PCS-CNN	766	<0.05	2205	0.28
TL-CNN	CD-CNN	216	<0.01	1018	<0.01
TL-CNN	ACS-CNN	62	<0.01	938	<0.01
PCS-CNN	CD-CNN	535	<0.01	1958	<0.05
PCS-CNN	ACS-CNN	114	<0.01	1350	<0.01
CD-CNN	ACS-CNN	88	<0.01	1443	<0.01

References

1. Boers, A.M.; Jansen, I.G.; Beenen, L.F.; Devlin, T.G.; San Roman, L.; Heo, J.H.; Ribó, M.; Brown, S.; Almekhlafi, M.A.; Liebeskind, D.S.; 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*, 1137–1142. [CrossRef]
2. Zaidi, S.F.; Aghaebrahim, A.; Urra, X.; Jumaa, M.A.; Jankowitz, B.; Hammer, M.; Nogueira, R.; Horowitz, M.; Reddy, V.; Jovin, T.G. Final infarct volume is a stronger predictor of outcome than recanalization in patients with proximal middle cerebral artery occlusion treated with endovascular therapy. *Stroke* **2012**, *43*, 3238–3244. [CrossRef] [PubMed]
3. Villringer, K.; Florczak-Rzepka, M.; Grittner, U.; Brunecker, P.; Tepe, H.; Nolte, C.H.; Fiebach, J.B. Characteristics associated with outcome in patients with first-ever posterior fossa stroke. *Eur. J. Neurol.* **2018**, *25*, 818–824. [CrossRef] [PubMed]
4. Schonewille, W.J.; Wijman, C.A.C.; Michel, P.; Rueckert, C.M.; Weimar, C.; Mattle, H.P.; Engelter, S.T.; Tanne, D.; Muir, K.W.; Molina, C.A.; et al. Treatment and outcomes of acute basilar artery occlusion in the Basilar Artery International Cooperation Study (BASICS): A prospective registry study. *Lancet Neurol.* **2009**, *8*, 724–730. [CrossRef]
5. Banks, J.L.; Marotta, C.A. Outcomes validity and reliability of the modified rankin scale: Implications for stroke clinical trials—A literature review and synthesis. *Stroke* **2007**, *38*, 1091–1096. [CrossRef]
6. Barros, R.S.; Tolhuisen, M.L.; Boers, A.M.M.; Jansen, I.; Ponomareva, E.; Dippel, D.W.J.; Van Der Lugt, A.; Van Oostenbrugge, R.J.; Van Zwam, W.H.; Berkhemer, O.A.; et al. Automatic segmentation of cerebral infarcts in follow-up computed tomography images with convolutional neural networks. *J. Neurointerv. Surg.* **2020**, *12*, 848–852. [CrossRef] [PubMed]
7. Kim, Y.C.; Lee, J.E.; Yu, I.; Song, H.N.; Baek, I.Y.; Seong, J.K.; Jeong, H.G.; Kim, B.J.; Nam, H.S.; Chung, J.W.; et al. Evaluation of Diffusion Lesion Volume Measurements in Acute Ischemic Stroke Using Encoder-Decoder Convolutional Network. *Stroke* **2019**, *50*, 1444–1451. [CrossRef]
8. Bogousslavsky, J.; Van Melle, G.; Regli, F. *Original Contributions The Lausanne Stroke Registry: Analysis of 1000 Consecutive Patients With First Stroke.* Available online: https://www.ahajournals.org/doi/abs/10.1161/01.str.19.9.1083 (accessed on 24 August 2021).
9. Moulin, T.; Tatu, L.; Vuillier, F.; Berger, E.; Chavot, D.; Rumbach, L. Role of a stroke data bank in evaluating cerebral infarction subtypes: Patterns and outcome of 1,776 consecutive patients from the Besancon Stroke Registry. *Cerebrovasc. Dis.* **2000**, *10*, 261–271. [CrossRef]
10. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [CrossRef]
11. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef] [PubMed]
12. Sanford, T.H.; Zhang, L.; Harmon, S.A.; Sackett, J.; Yang, D.; Roth, H.; Xu, Z.; Kesani, D.; Mehralivand, S.; Baroni, R.H.; et al. Data Augmentation and Transfer Learning to Improve Generalizability of an Automated Prostate Segmentation Model. *Am. J. Roentgenol.* **2020**, *215*, 1403–1410. [CrossRef]
13. Nishio, M.; Fujimoto, K.; Matsuo, H.; Muramatsu, C.; Sakamoto, R.; Fujita, H. Lung Cancer Segmentation With Transfer Learning: Usefulness of a Pretrained Model Constructed From an Artificial Dataset Generated Using a Generative Adversarial Network. *Front. Artif. Intell.* **2021**, *4*, 694815. [CrossRef]
14. Byra, M.; Wu, M.; Zhang, X.; Jang, H.; Ma, Y.-J.; Chang, E.Y.; Shah, S.; Du, J. Knee menisci segmentation and relaxometry of 3D ultrashort echo time cones MR imaging using attention U-Net with transfer learning. *Magn. Reson. Med.* **2020**, *83*, 1109–1122. [CrossRef]
15. Goyal, M.; Menon, B.K.; Van Zwam, W.H.; Dippel, D.W.J.; Mitchell, P.J.; Demchuk, A.M.; Dávalos, A.; Majoie, C.B.L.M.; Van Der Lugt, A.; De Miquel, M.A.; et al. Endovascular thrombectomy after large-vessel ischaemic stroke: A meta-analysis of individual patient data from five randomised trials. *Lancet* **2016**, *387*, 1723–1731. [CrossRef]

16. Van Der Hoeven, E.J.; Schonewille, W.J.; Vos, J.A.; Algra, A.; Audebert, H.J.; Berge, E.; Ciccone, A.; Mazighi, M.; Michel, P.; Muir, K.W.; et al. The Basilar Artery International Cooperation Study (BASICS): Study protocol for a randomised controlled trial. *Trials* **2013**, *14*, 200. [CrossRef] [PubMed]

17. Langezaal, L.C.M.; van der Hoeven, E.J.R.J.; Mont'Alverne, F.J.A.; de Carvalho, J.J.F.; Lima, F.O.; Dippel, D.W.J.; van der Lugt, A.; Lo, R.T.H.; Boiten, J.; Lycklama à Nijeholt, G.J.; et al. Endovascular Therapy for Stroke Due to Basilar-Artery Occlusion. *N. Engl. J. Med.* **2021**, *384*, 1910–1920. [CrossRef] [PubMed]

18. Yushkevich, P.A.; Piven, J.; Hazlett, H.C.; Smith, R.G.; Ho, S.; Gee, J.C.; Gerig, G. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* **2006**, *31*, 1116–1128. [CrossRef]

19. Puetz, V.; Sylaja, P.N.; Coutts, S.B.; Hill, M.D.; Dzialowski, I.; Mueller, P.; Becker, U.; Urban, G.; O'Reilly, C.; Barber, P.A.; et al. Extent of hypoattenuation on CT angiography source images predicts functional outcome in patients with basilar artery occlusion. *Stroke* **2008**, *39*, 2485–2490. [CrossRef]

20. Mattes, D.; Haynor, D.R.; Vesselle, H.; Lewellyn, T.K.; Eubank, W. Nonrigid multimodality image registration. In Proceedings of the Medical Imaging 2001: Image Processing, San Diego, CA, USA, 17–22 February 2001; SPIE: Bellingham, WA, USA, 2001; Volume 4322, pp. 1609–1620.

21. Lowekamp, B.C.; Chen, D.T.; Ibáñez, L.; Blezek, D. The design of simpleITK. *Front. Neuroinform.* **2013**, *7*, 45. [CrossRef] [PubMed]

22. Yaniv, Z.; Lowekamp, B.C.; Johnson, H.J.; Beare, R. SimpleITK Image-Analysis Notebooks: A Collaborative Environment for Education and Reproducible Research. *J. Digit. Imaging* **2018**, *31*, 290–303. [CrossRef]

23. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 9351, pp. 234–241.

24. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.

25. Cicchetti, D.V. Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychol. Assess.* **1994**, *6*, 284–290. [CrossRef]

26. Vallat, R. Pingouin: Statistics in Python. *J. Open Source Softw.* **2018**, *3*, 1026. [CrossRef]

27. Clèrigues, A.; Valverde, S.; Bernal, J.; Freixenet, J.; Oliver, A.; Lladó, X. Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks. *Comput. Biol. Med.* **2019**, *115*, 103487. [CrossRef] [PubMed]

28. Öman, O.; Mäkelä, T.; Salli, E.; Savolainen, S.; Kangasniemi, M. 3D convolutional neural networks applied to CT angiography in the detection of acute ischemic stroke. *Eur. Radiol. Exp.* **2019**, *3*, 8. [CrossRef]

29. Yu, Y.; Xie, Y.; Thamm, T.; Gong, E.; Ouyang, J.; Huang, C.; Christensen, S.; Marks, M.P.; Lansberg, M.G.; Albers, G.W.; et al. Use of Deep Learning to Predict Final Ischemic Stroke Lesions From Initial Magnetic Resonance Imaging. *JAMA Netw. Open* **2020**, *3*, e200772. [CrossRef]

30. Qiu, W.; Kuang, H.; Teleg, E.; Ospel, J.M.; Sohn, S.I.; Almekhlafi, M.; Goyal, M.; Hill, M.D.; Demchuk, A.M.; Menon, B.K. Machine Learning for Detecting Early Infarction in Acute Stroke with Non–Contrast-enhanced CT. *Radiology* **2020**, *294*, 638–644. [CrossRef] [PubMed]

31. Barman, A.; Inam, M.E.; Lee, S.; Savitz, S.; Sheth, S.; Giancardo, L. Determining ischemic stroke from ct-angiography imaging using symmetry-sensitive convolutional networks. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging, Venice, Italy, 8–11 April 2019; IEEE Computer Society: Washington, DC, USA, 2019; pp. 1873–1877.

32. Pustina, D.; Branch Coslett, H.; Turkeltaub, P.E.; Tustison, N.; Schwartz, M.F.; Avants, B. Automated Segmentation of Chronic Stroke Lesions Using LINDA: Lesion Identification With Neighborhood Data Analysis. *Hum. Brain Mapp.* **2016**, *37*, 1405–1421. [CrossRef]

33. Tomita, N.; Jiang, S.; Maeder, M.E.; Hassanpour, S. Automatic post-stroke lesion segmentation on MR images using 3D residual convolutional neural network. *NeuroImage Clin.* **2020**, *27*, 102276. [CrossRef]

34. Chen, S.; Ma, K.; Zheng, Y. MED3D: Transfer learning for 3D medical image analysis. *arXiv* **2019**, arXiv:1904.00625v4.

35. Zhou, Z.; Sodha, V.; Rahman Siddiquee, M.M.; Feng, R.; Tajbakhsh, N.; Gotway, M.B.; Liang, J. Models genesis: Generic autodidactic models for 3d medical image analysis. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019*; Springer: Berlin/Heidelberg, Germany, 2019; Volume 11767, pp. 384–393.

36. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2014**, *115*, 211–252. [CrossRef]

37. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; Volume 4, pp. 3320–3328.