Riaan Zoetmulder

Deep-Learning-Based Image Segmentation for Uncommon Ischemic Stroke: From Infants to Adults

# Deep-Learning-Based Image Segmentation for Uncommon Ischemic Stroke: From Infants to Adults

Riaan Zoetmulder

# Deep-Learning-Based Image Segmentation for Uncommon Ischemic Stroke:

## From Infants to Adults

**Riaan Zoetmulder**

Printing: Gildeprint || www.gildeprint.nl

Deep-Learning-Based Image Segmentation for Uncommon Ischemic Stroke: From Infants to Adults

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Aula der Universiteit
op vrijdag 17 februari 2023, te 14.00 uur

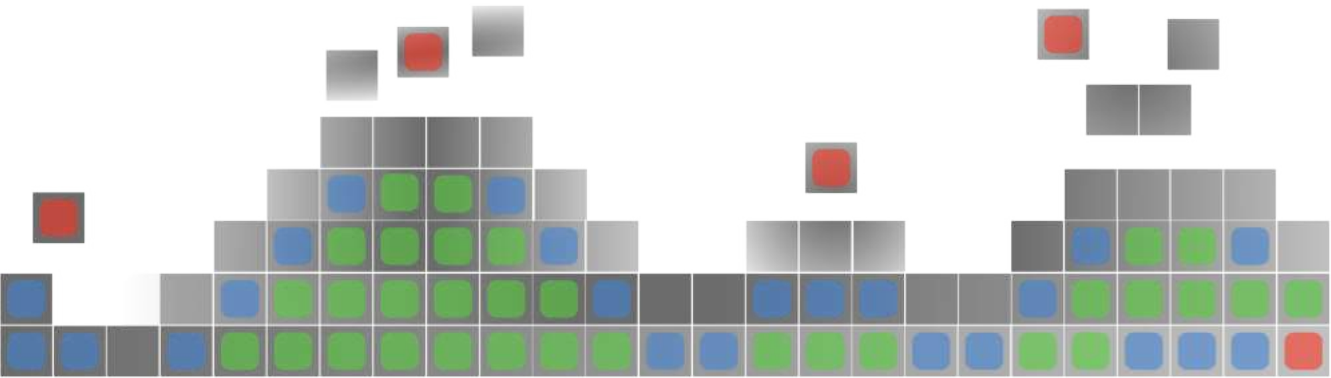
door Riaan Zoetmulder
geboren te Amsterdam

***Promotiecommissie***

| | | |
|---|---|---|
| *Promotores:* | prof. dr. H.A. Marquering | AMC-UvA |
| | prof. dr. I. Išgum | AMC-UvA |
| *Copromotores:* | dr. E. Gavves | Universiteit van Amsterdam |
| *Overige leden:* | prof. dr. ir. C.I. Sánchez Gutiérrez | Universiteit van Amsterdam |
| | dr. J. Coutinho | AMC-UvA |
| | prof. dr. ir. M. Staring | Universiteit Leiden |
| | prof. dr. C.G.M. Snoek | Universiteit van Amsterdam |
| | prof. dr. A. Abu-Hanna | AMC-UvA |
| | prof. dr. R.R. van Rijn | AMC-UvA |

Faculteit der Geneeskunde

# Contents



1 Introduction **2**

2 Domain- and Task-Specific Transfer learning **26**

2.1 Introduction . . . 28
2.2 Related Work . . . 30
2.3 Materials . . . 32
2.4 Methods: Source and Target Tasks . . . 35
2.5 Experiments & Results . . . 38
2.6 Discussion . . . 42
2.7 Conclusion . . . 46
2.8 Acknowledgements . . . 46
2.9 Supplementary Material . . . 47

3 Posterior Stroke Lesion Segmentation **58**

3.1 Introduction . . . 60
3.2 Materials & Methods . . . 61
3.3 Results . . . 66
3.4 Discussion . . . 67
3.5 Supplementary materials . . . 74

4 Thrombus Localization and Segmentation **84**

4.1 Introduction . . . 86
4.2 Materials & Methods . . . 87
4.3 Results . . . 94
4.4 Discussion . . . 98
4.5 Supplementary Materials . . . 101

5 Perinatal Arterial Ischemic Stroke **112**

5.1 Introduction . . . . . 114
5.2 Materials & Methods . . . . . 116
5.3 Experiments and Results . . . . . 123
5.4 Discussion . . . . . 131

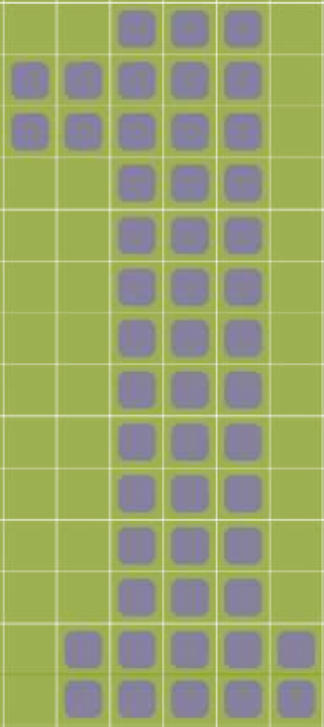
6 Summary **142**

7 General Discussion **148**

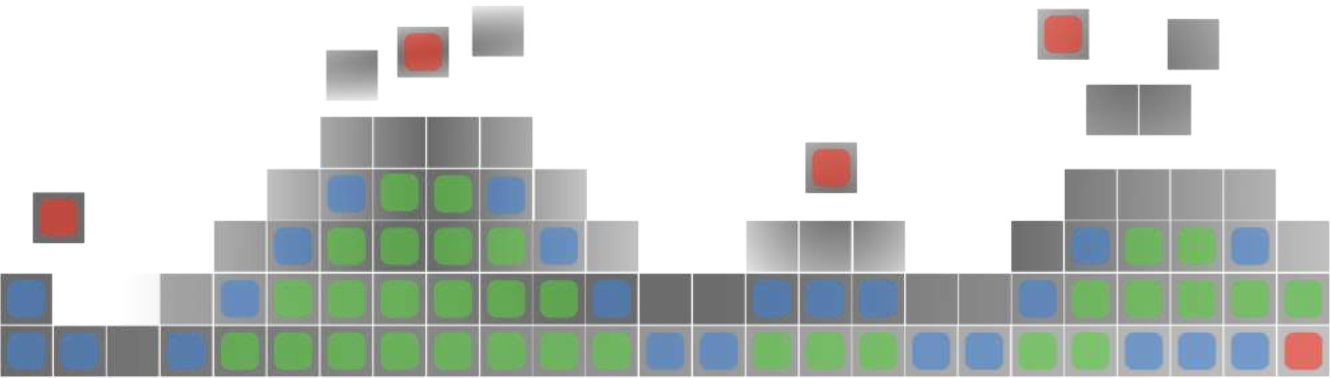
8 Nederlandse Samenvatting **162**

Appendix **169**

Abbreviations . . . . . 169
Portfolio . . . . . 171
Acknowledgements . . . . . 173
About the author . . . . . 174



# Chapter 1
## Introduction



1

> There are other things to fear... apart from death, old age and madness. For example, apoplexy, that lightning bolt which strikes you down without destroying you, yet after which all is finished. You are still yourself but you are no longer yourself: from a near angel like Ariel, you have become a dull mass which, like Caliban, is close to the beasts.
>
> *Alexandre Dumas*, The Count of Monte Cristo. p.478

Stroke is defined as an acute neurological deficit that stems from injury to the central nervous system that is caused by an interruption of the circulatory system [13; 59]. The brain, with its high metabolic activity, is especially vulnerable to injury due to ischemia [15; 47]. Ischemic large vessel stroke, which is caused by an occluding thrombus in one of the large vessels of the brain, accounts for up to 46 percent of all cases of stroke in adult patients [40; 62]. Hence, ischemic large vessel stroke in adult patients has a poor prognosis, with only 33 percent of patients reaching functional independence 90 days after the stroke incident [8]. This has lead to stroke in adults being associated with a high financial cost. The global financial cost associated with stroke in adult patients was estimated at US $ 851 Billion [51].

Ischemic stroke occurs not only in adult patients, but also perinatally. Perinatal stroke is defined as a stroke that has occurred between 20 weeks of gestation up until 28 days after birth[20]. Perinatal Arterial Ischemic Stroke (PAIS) has the second highest incidence, after stroke in elderly patients [56]. The incidence rate of PAIS has been estimated to be between 1 per 3500 [1] and 1 per 7700 live births [25]. PAIS is associated with life time medical issues such as epilepsy, cerebral palsy, and cognitive and motor impairment [48].

The brain has two pathways that supply blood to it: The anterior and the posterior circulation. Anterior circulation large vessel ischemic stroke is more common than posterior circulation large vessel ischemic stroke, with 1 percent of large vessel ischemic strokes occurring in the posterior circulation. Poste-

rior circulation large artery stroke involves an occlusion of the basilar artery, intracranial vertebral arteries or posterior cerebellar arteries. It is associated with a high risk of mortality and disability [29; 41; 61; 67].

During an ischemic stroke, the area of the brain that is ischemic can be subdivided into two regions. The first region is named the ischemic core. It is composed of tissue that has died due to insufficient blood flow and can no longer be salvaged. The second region is named the penumbra [5]. Due to a diminished blood supply, the neurons in the penumbra are no longer evoking action potentials. However, due to alternative routes, some blood can still be supplied to the penumbra, which can keep the tissue alive for several hours. Therefore, identifying the ischemic penumbra can help select patients that have the most salvageable tissue and would benefit the most from treatment [47].

## Treatment

Approved treatments for adult patients suffering from acute ischemic stroke aim to reperfuse the ischemic area by removal of the occluding thrombus. Two treatments have proven to be both safe and effective for ischemic stroke. First, the occlusion can be dissolved by means of thrombolytic therapy. Thrombolytic therapy involves the intravenous administration of recombinant Tissue Plasminogen Activator (rTPA). It is effective if administered within 4.5 hours after stroke onset [19]. Second, the occlusion can be removed by means of mechanical thrombectomy. Mechanical thrombectomy involves navigating a stent-retriever or an aspiration catheter from a puncture in the femoral artery to the occluding thrombus in the brain and removing it. Mechanical thrombectomy in addition to thrombolytic therapy has become the recommended treatment for patients arriving to the hospital within 6 hours of the time the patient was last known to be well. When the patient arrives between 6 and 24 hours after the patient was last known to be well, mechanical thrombectomy can often still be performed. This is the case if there is a sufficiently large mismatch between the volume of the infarct core and the penumbra or the infarct core and the clinical deficit [2; 49; 54; 66].

For infant patients suffering from PAIS, there are currently no approved treatments. Thrombolytics and mechanical thrombectomy are not given due to a lack of evidence regarding their effectiveness in neonates [57]. All patients suffering from PAIS should be given neuroprotective measures and supportive care. Examples of supportive care are controlling seizures, ensuring adequate oxygenation and, correcting any anemia or dehydration, as well as reducing fever [20; 57].

The current focus of research that focuses on developing therapies for PAIS are novel stem-cell based neuro-regenerative treatments [68]. This is because neonates suffering from a PAIS have a greater ability to recover from the ischemic injury, due to their brains having greater plasticity than adult brains [10]. Moreover, animal models of the neonatal brain show a greater response to stem cell therapy than animal models of adult brains [69].

## Neuro-imaging

Brain imaging is recommended for patients suspected of suffering from an acute ischemic stroke [54]. Brain imaging serves multiple purposes during a workup of acute ischemic stroke [30; 46]. First, to differentiate between hemorrhagic and ischaemic stroke. Second, to determine the extent of the ischemic core. Third, to determine the size of the penumbra. Fourth, to exclude stroke mimics, such as migraines or tumors as the cause of the symptoms. Fifth, to assess the large arteries in the head and neck. Sixth, to guide interventions, such as selecting which patients are eligible for a thrombectomy.

Computed Tomography (CT) is the preferred imaging modality in adult patients with stroke due to widespread availability, short scan times and high sensitivity for differentiating hemorrhagic from ischemic stroke. Specifically, Non-Contrast Computed Tomography (NCCT) is used to detect hemorrhagic stroke and the presence of hyper-dense artery sign. During Computed Tomography Angiography (CTA) a bolus of contrast is administered to the patient to visualize the arteries and to quantify the vascular disease burden (the degree of stenosis, thrombus length etc.) caused by the occluding thrombus [17; 34].

Magnetic Resonance Imaging (MRI) can also aid in the diagnosis of stroke. MRI is as accurate as NCCT at the detection of acute hemorrhage and more accurate than CT at detecting chronic intracerebral hemorrhage [21; 32; 53]. Furthermore, diffusion weighted imaging (DWI), an MRI sequence that shows the diffusion of water, has been shown to be accurate at diagnosing acute ischemic stroke [11; 22; 24].

For neonatal patients with PAIS, MRI is preferred over CT. This is due to the low sensitivity of CT at detecting small or posterior circulation infarcts, the rarity of hyper-dense artery signs in neonatal stroke patients, and the radiation produced by CT scanners [9]. A radiological workup of PAIS patients consists of several MRI sequences. These sequences include Diffusion Weighted Imaging (DWI) and apparent diffusion coefficient, susceptibility weighted imaging, and magnetic resonance angiography. T2- and T1-weighted sequences are optionally added [9; 20; 33]. To evaluate the treatment efficacy of novel neuro-regenerative treatments information from images can be used, such as final lesion volume on follow-up MRI. [7].

Brain imaging data collected from stroke patients also has an important role in stroke research. Features extracted from these images, can be used as an alternative outcome measure to assess treatment efficacy. Examples are ischemic lesion volume on baseline and follow-up scans in adult patients [12; 70] in perinatal arterial ischemic stroke [7]. Image features, such as ischemic lesion volume, rely on its accurate segmentation. However, its manual segmentation is a time consuming task. Therefore, automated deep learning based segmentation methods have the potential to reduce the time required to create segmentations.

## Machine Learning and Deep Learning

Machine learning is the study and development of computer algorithms that are capable of learning. In this definition, learning is defined as the ability of the algorithm to improve its performance at tasks with experience [43].

For example, an algorithm that learns how to play chess would have its performance measured by its ability to win at the group of tasks that involve playing chess games and by gaining experience through playing games of chess against itself or others.

Deep learning is a sub-field of machine learning that utilizes layers of non-linear processing of information, also known as Artificial Neural Network (ANN), to solve learning problems [18]. The most basic neural network has five key components; weights, biases, an activation function, an output function, a loss function, and an optimizer. The goal of training a neural network is to minimize the loss function given the input data. The process by which the neural network is taught how to minimize the loss function is referred to as training. Before training, the weights and biases are initialized by assigning them random numbers. The input data is transformed by multiplication of the data with the weights and addition of the biases. Next, the result is transformed by applying the activation function. The resulting matrix of numbers is referred to as the features. This process is repeated a pre-specified number of times. When the final layer is reached, the output function is applied. The output function varies per image analysis task, for example classification uses the softmax output function and regression uses the linear activation function. Next, the value of the loss function is computed on the basis of the output. By using the optimizer the weights and biases are updated to reduce the value of the loss function. This process continues until the loss value has reached a stable value. See figure 1.1 for a schematic representation of this process.

In recent years, deep learning has been applied to solve multiple medical image analysis tasks [36]. Specifically, the network architecture that is commonly used to create solutions for medical image analysis are Convolutional Neural Networks (CNN) [27]. CNNs have their weights organised in convolutional kernels, which are slid over and multiplied with the underlying input data as opposed to matrices that are multiplied with the input data. A graphical representation of how convolutions work in a neural network can be found in figure 1.2. This allows local correlations between for example pixels in images to be more easily used. Examples of medical image analysis

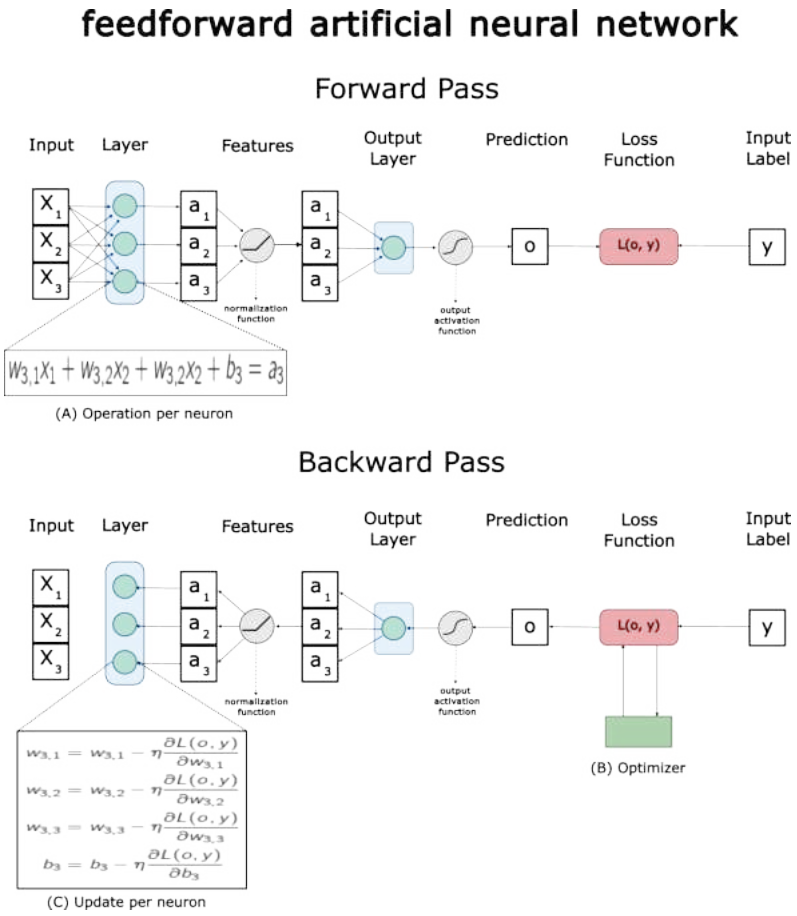



**Figure 1.1:** Forward (top) and backward pass (bottom) of a simple two-layered artificial neural network(ANN). The neural network consists of an input layer which is composed of three neurons. The input layer is followed by an activation function, which is followed by the output layer consisting of one neuron. Finally, the output layer is followed by a sigmoid activation function and the cross-entropy loss. During the forward pass the neural network is given the input data and the input label. **(A)** Each neuron in the input layer multiplies one of its weights ($w_{l,n}$) with one of the input values($x_n$), sums the result and adds a bias value ($b_n$) to calculate the features ($a_n$). The features are normalized and passed to the output layer. The output layer repeats the process and normalizes the result as is required for the specific task. The loss value is calculated using the input label and network result. During the backward pass the optimizer **(B)** is used to calculate the gradient updates to minimize the loss function. To achieve this the gradients are passed through the network such that all of the weights and biases are updated **(C)**.

tasks are segmentation of brain tissue on MRI [44], image registration [16] and lung nodule detection [28].

1

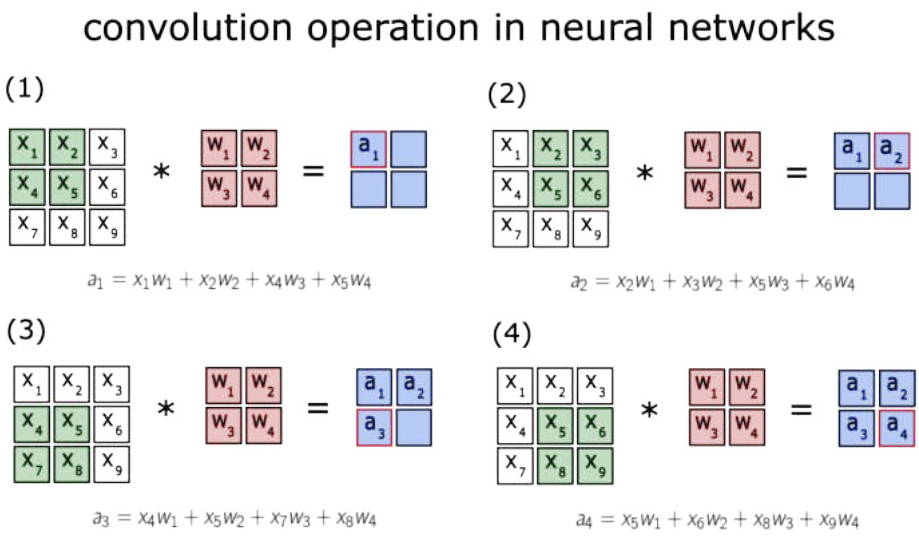



**Figure 1.2:** Example of how the convolution operation is used in a convolutional neural network (CNN). Shown in each of the four panels from left to right is the input data (green and white), the kernel that is parameterized by weights (red), and the resulting features (blue). The kernel is multiplied element-wise with the part of the input data that is shaded green. The formula below shows the resulting calculation for each feature. Steps 1 through four show how the kernel is moved over the input data to calculate all of the features. The weights in the kernel are updated such that spatial correlations in an image are captured by the network.

CNNs have been applied to image analysis tasks related to stroke in adult patients. Prior work has developed deep learning-based algorithms to identify patients with a large-vessel occlusion stroke in CTA images [4; 42; 60; 63] and to detect or segment thrombi in the anterior circulation on NCCT scans [35; 45; 64]. Other work has focused on developing algorithms to segment diffusion abnormalities during an acute ischemic stroke on DWI [37; 72]. Zhang et al. used a 3D DenseNet to segment the ischemic region on DWI [72]. In addition, Zhang et al. evaluated their method on data from the open-source Ischemic Stroke LEsion Segmentation (ISLES) challenge[39], making their method easily comparable to other existing methods. Liu et al. developed a method that outperformed existing methods, specifically on small lesions. Algorithms have also been developed to segment chronic stroke lesions on T1-weighted MRI scans by using variations of U-Net [58]. Tomita et al. used

a UNet which was trained by using zoomed-in sub volumes, the zoom factor was decreased as training progressed. They speculated that their training strategy improved performance due to improved regularization [65]. Qi et al. replaced the standard U-Net convolutional blocks with feature-similarity modules that more effectively utilize dependencies between distantly placed voxels [55]. Finally, Zhou et al. used a specific convolutional block to fuse 2D and 3D features and they used a novel mixing loss to more effectively address the class imbalance between foreground and background voxels [73]. Algorithms have also been developed to segment ischemic lesions due to an anterior circulation stroke on follow-up NCCT [6] and baseline CTA [50]. All previously mentioned studies have focused on segmentation of lesions on the same scan that was used to create the ground truth annotation. However, Yu et al. developed a method that predicts the follow-up lesion segmentation from baseline multi-sequence MRI [71]. Their method achieved a moderate overlap and a good volumetric agreement between the ground truth and predicted segmentation.

## Transfer Learning

A limitation of CNNs is that they require large amounts of annotated data to learn how to accurately segment objects in images. However, creating large amounts of annotations is a time-consuming task. Moreover, large amounts of scan data may not always be available for uncommon diseases. Hence, reducing the amount of annotated data that is required for the development of CNN based algorithms is an active area of research.

Transfer learning is a method that can be used to improve CNN performance in settings where available annotated data is scarce [52]. Transfer learning aims to re-use a CNN that has been pre-trained on a different source task with a large amount of available data on a new target task [36]. The task of a neural network refers to the specific application such as segmentation, classification or regression. In addition to the task, another relevant aspect of transfer learning is the source and target domain [52]. The domain refers to the type of data on which the neural network is trained or to which the neural network is applied.

Prior studies on transfer learning have evaluated the effect of different source tasks and domains on target medical MRI segmentation tasks in a limited manner. On the natural image source domain, studies have only assessed the source tasks classification [3; 26] and segmentation [31], but not self-supervised tasks. On the medical image source domain the only source tasks that were evaluated, were segmentation [14; 23; 38] and self-supervised tasks [74]. Classification source tasks were not yet evaluated. Moreover, all the studies evaluated the the effect the source tasks and domains had on target task performance on different data-sets, which makes a fair comparison difficult. Hence, a fair evaluation of the effect that the source task and domain have on the target medical segmentation tasks could allow for a effective application of transfer learning.

## Aim of this thesis

Both posterior circulation large vessel and perinatal arterial ischemic stroke are uncommon types of stroke. Hence, few data is available to train algorithms to perform complex segmentation tasks. The aim of this thesis is to investigate, develop, and evaluate deep learning-based algorithms for automatic segmentation of images of these types of stroke.

## Thesis outline

In **chapter 2**, we evaluate transfer learning for medical image analysis tasks. Specifically, we focus on evaluating the overlap and detection accuracy of source-tasks and domain combinations for target segmentation tasks on MR scans. We apply transfer learning to two segmentation tasks related to posterior circulation large vessel ischemic stroke in **chapter 3** and **chapter 4**. In **chapter 3**, we develop and evaluate algorithms for automated segmentation of posterior circulation stroke lesions on follow-up scans. In **chapter 4**, we develop and evaluate an automatic method for localization and segmentation of thrombi in the posterior circulation stroke on baseline scans. In **chapter 5** a method is developed that automatically segments brain tissues and the

ischemic lesion per hemisphere on baseline and follow-up MR brain scans of patients suffering from a perinatal arterial ischemic stroke. Finally, in **chapter 6** we discuss the results.

# Bibliography



[1] N. Agrawal, S. C. Johnston, Y. W. Wu, S. Sidney, and H. J. Fullerton. Imaging data reveal a higher pediatric stroke incidence than prior us estimates. *Stroke*, 40(11):3415–3421, 11 2009.

[2] G. W. Albers, M. P. Marks, S. Kemp, S. Christensen, J. P. Tsai, S. Ortega-Gutierrez, R. A. McTaggart, M. T. Torbey, M. Kim-Tenser, T. Leslie-Mazwi, A. Sarraj, S. E. Kasner, S. A. Ansari, S. D. Yeatts, S. Hamilton, M. Mlynash, J. J. Heit, G. Zaharchuk, S. Kim, J. Carrozzella, Y. Y. Palesch, A. M. Demchuk, R. Bammer, P. W. Lavori, J. P. Broderick, and M. G. Lansberg. Thrombectomy for Stroke at 6 to 16 Hours with Selection by Perfusion Imaging. *New England Journal of Medicine*, 378(8):708–718, 2 2018.

[3] J. Amin, M. Sharif, M. Yasmin, T. Saba, M. A. Anjum, and S. L. Fernandes. A New Approach for Brain Tumor Segmentation and Classification Based on Score Level Fusion Using Transfer Learning. *Journal of Medical Systems*, 43(11):1–16, 11 2019.

[4] S. A. Amukotuwa, M. Straka, S. Dehkharghani, and R. Bammer. Fast automatic detection of large vessel occlusions on CT angiography. *Stroke*, 50(12):3431–3438, 12 2019.

[5] J. Astrup, B. K. Siesjö, and L. Symon. Thresholds in cerebral ischemia the ischemic penumbra. *Stroke*, 12(6):723–725, 1981.

[6] R. S. Barros, M. L. Tolhuisen, A. M. Boers, I. Jansen, E. Ponomareva, D. W. Dippel, A. Van Der Lugt, R. J. Van Oostenbrugge, W. H. Van Zwam, O. A. Berkhemer, M. Goyal, A. M. Demchuk, B. K. Menon, P. Mitchell, M. D. Hill, T. G. Jovin, A. Davalos, B. C. Campbell, J. L. Saver, Y. B. Roos, K. W. Muir, P. White, S. Bracard, F. Guillemin, S. D. Olabarriaga, C. B. Majoie, and H. A. Marquering. Automatic segmentation of cerebral infarcts in follow-up computed tomography images with convolutional neural networks. *Journal of NeuroInterventional Surgery*, 12(9):848–852, 9 2020.

[7] M. J. Benders, N. E. Van Der Aa, M. Roks, H. L. Van Straaten, I. Isgum, M. A. Viergever, F. Groenendaal, L. S. De Vries, and F. Van Bel. Feasibility

and safety of erythropoietin for neuroprotection after perinatal arterial ischemic stroke. *The Journal of pediatrics*, 164(3), 2014.

[8] O. Berkhemer, P. Fransen, D. Beumer, L. van den Berg, H. Lingsma, A. Yoo, W. Schonewille, J. Vos, P. Nederkoorn, M. Wermer, M. van Walderveen, J. Staals, J. Hofmeijer, J. van Oostayen, G. Lycklama à Nijeholt, J. Boiten, P. Brouwer, B. Emmer, S. de Bruijn, L. van Dijk, L. Kappelle, R. Lo, E. van Dijk, J. de Vries, P. de Kort, W. van Rooij, J. van den Berg, B. van Hasselt, L. Aerden, R. Dallinga, M. Visser, J. Bot, P. Vroomen, O. Eshghi, T. Schreuder, R. Heijboer, K. Keizer, A. Tielbeek, H. den Hertog, D. Gerrits, R. van den Berg-Vos, G. Karas, E. Steyerberg, H. Flach, H. Marquering, M. Sprengers, S. Jenniskens, L. Beenen, R. van den Berg, P. Koudstaal, W. van Zwam, Y. Roos, A. van der Lugt, R. van Oostenbrugge, C. Majoie, D. Dippel, v. Zwam, v. der Lugt, v. Oostenbrugge, and D. con. A Randomized Trial of Intraarterial Treatment for Acute Ischemic Stroke. *n engl j med*, 372:11–20, 2015.

[9] A. Biswas, K. Mankad, M. Shroff, P. Hanagandi, and P. Krishnan. Neuroimaging Perspectives of Perinatal Arterial Ischemic Stroke. *Pediatric Neurology*, 113:56–65, 12 2020.

[10] A. J. Bower. Plasticity in the adult and neonatal central nervous system. *British Journal of Neurosurgery*, 4(4):253–264, 1991.

[11] A. M. Brunser, A. Hoppe, S. Illanes, V. Díaz, P. Munoz, D. Cárcamo, V. Olavarria, M. Valenzuela, and P. Lavados. Accuracy of diffusion-weighted imaging in the diagnosis of stroke in patients with suspected cerebral infarct. *Stroke*, 44(4):1169–1171, 2013.

[12] A. Bucker, A. M. Boers, J. C. Bot, O. A. Berkhemer, H. F. Lingsma, A. J. Yoo, W. H. Van Zwam, R. J. Van Oostenbrugge, A. Van Der Lugt, D. W. Dippel, Y. B. Roos, C. B. Majoie, and H. A. Marquering. Associations of Ischemic Lesion Volume with Functional Outcome in Patients with Acute Ischemic Stroke: 24-Hour Versus 1-Week Imaging. *Stroke*, 48(5):1233–1240, 5 2017.

[13] B. C. Campbell and P. Khatri. Stroke, 2020.

[14] S. Chen, K. Ma, and Y. Zheng. MED3D: Transfer learning for 3D medical image analysis, 2019.

[15] D. W. Choi. Excitotoxic cell death. *Journal of Neurobiology*, 23(9):1261–1276, 11 1992.

[16] B. D. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Išgum. A deep learning framework for unsupervised affine and deformable image registration. *Medical Image Analysis*, 52:128–143, 2 2019.

[17] A. M. Demchuk, B. K. Menon, and M. Goyal. Comparing vessel imaging: Noncontrast computed tomography/computed tomographic angiography should be the new minimum standard in acute disabling stroke, 2016.

[18] L. Deng and D. Yu. Deep learning: Methods and applications, 2013.

[19] J. Emberson, K. R. Lees, P. Lyden, L. Blackwell, G. Albers, E. Bluhmki, T. Brott, G. Cohen, S. Davis, G. Donnan, J. Grotta, G. Howard, M. Kaste, M. Koga, R. Von Kummer, M. Lansberg, R. I. Lindley, G. Murray, J. M. Olivot, M. Parsons, B. Tilley, D. Toni, K. Toyoda, N. Wahlgren, J. Wardlaw, W. Whiteley, G. J. Del Zoppo, C. Baigent, P. Sandercock, and W. Hacke. Effect of treatment delay, age, and stroke severity on the effects of intravenous thrombolysis with alteplase for acute ischaemic stroke: A meta-analysis of individual patient data from randomised trials. *The Lancet*, 384(9958):1929–1935, 11 2014.

[20] D. M. Ferriero, H. J. Fullerton, T. J. Bernard, L. Billinghurst, S. R. Daniels, M. R. Debaun, G. Deveber, R. N. Ichord, L. C. Jordan, P. Massicotte, J. Meldau, E. S. Roach, and E. R. Smith. Management of Stroke in Neonates and Children: A Scientific Statement From the American Heart Association/American Stroke Association. *Stroke*, 50(3):E51–E96, 3 2019.

[21] J. B. Fiebach, P. D. Schellinger, A. Gass, T. Kucinski, M. Siebler, A. Villringer, P. Ölkers, J. G. Hirsch, S. Heiland, P. Wilde, O. Jansen, J. Röther, W. Hacke, and K. Sartor. Stroke Magnetic Resonance Imaging Is Accurate

in Hyperacute Intracerebral Hemorrhage: A Multicenter Study on the Validity of Stroke Imaging. *Stroke*, 35(2):502–506, 2 2004.

[22] J. B. Fiebach, P. D. Schellinger, O. Jansen, M. Meyer, P. Wilde, J. Bender, P. Schramm, E. Jüttler, J. Oehler, M. Hartmann, S. Hähnel, M. Knauth, W. Hacke, and K. Sartor. CT and diffusion-weighted MR imaging in randomized order: Diffusion-weighted imaging results in higher accuracy and lower interrater variability in the diagnosis of hyperacute ischemic stroke. *Stroke*, 33(9):2206–2210, 9 2002.

[23] M. Ghafoorian, A. Mehrtash, T. Kapur, N. Karssemeijer, E. Marchiori, M. Pesteie, C. R. Guttmann, F. E. de Leeuw, C. M. Tempany, B. van Ginneken, A. Fedorov, P. Abolmaesumi, B. Platel, and W. M. Wells. Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10435 LNCS, pages 516–524. Springer Verlag, 9 2017.

[24] R. G. González, P. W. Schaefer, F. S. Buonanno, L. H. Schwamm, R. F. Budzik, G. Rordorf, B. Wang, A. G. Sorensen, and W. J. Koroshetz. Diffusion-weighted MR imaging: Diagnostic accuracy in patients imaged within 6 hours of stroke symptom onset. *Radiology*, 210(1):155–162, 1 1999.

[25] S. Grunt, L. Mazenauer, S. E. Buerki, E. Boltshauser, A. C. Mori, A. N. Datta, J. Fluss, D. Mercati, E. Keller, O. Maier, C. Poloni, G. P. Ramelli, T. Schmitt-Mechelke, and M. Steinlin. Incidence and outcomes of symptomatic neonatal arterial ischemic stroke. *Pediatrics*, 135(5):e1220–e1228, 5 2015.

[26] T. Han, V. X. Nunes, L. F. De Freitas Souza, A. G. Marques, I. C. L. Silva, M. A. A. F. Junior, J. Sun, and P. P. Filho. Internet of medical things - Based on deep learning techniques for segmentation of lung and stroke regions in CT Scans. *IEEE Access*, 8:71117–71135, 2020.

[27] M. H. Hesamian, W. Jia, X. He, and P. Kennedy. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *Journal of Digital Imaging*, 32(4):582–596, 8 2019.

[28] X. Huang, J. Shan, and V. Vaidya. Lung nodule detection in CT using 3D convolutional neural networks. In *Proceedings - International Symposium on Biomedical Imaging*, pages 379–383. IEEE Computer Society, 6 2017.

[29] S. D. Israeli-Korn, Y. Schwammenthal, T. Yonash-Kimchi, M. Bakon, R. Tsabari, D. Orion, B. Bruk, N. Molshatzki, O. Merzeliak, J. Chapman, and D. Tanne. Ischemic stroke due to acute basilar artery occlusion: Proportion and outcomes. *Israel Medical Association Journal*, 12(11):671–675, 2010.

[30] S. Kamalian and M. H. Lev. Stroke Imaging. *Radiologic Clinics of North America*, 57(4):717–732, 7 2019.

[31] J. Kang and J. Gwak. Ensemble of Instance Segmentation Models for Polyp Segmentation in Colonoscopy Images. *IEEE Access*, 7:26440–26447, 2019.

[32] C. S. Kidwell, J. A. Chalela, J. L. Saver, S. Starkman, M. D. Hill, A. M. Demchuk, J. A. Butman, N. Patronas, J. R. Alger, L. L. Latour, M. L. Luby, A. E. Baird, M. C. Leary, M. Tremwel, B. Ovbiagele, A. Fredieu, S. Suzuki, J. P. Villabianca, S. Davis, B. Dunn, J. W. Todd, M. A. Ezzeddine, J. Haymore, J. K. Lynch, L. Davis, and S. Warach. Comparison of MRI and CT for detection of acute intracerebral hemorrhage. *Journal of the American Medical Association*, 292(15):1823–1830, 10 2004.

[33] S. Lee, D. M. Mirsky, L. A. Beslow, C. Amlie-Lefond, A. R. Danehy, L. Lehman, N. V. Stence, A. Vossough, M. Wintermark, and M. J. Rivkin. Pathways for Neuroimaging of Neonatal Stroke. *Pediatric Neurology*, 69:37–48, 4 2017.

[34] M. P. Lin and D. S. Liebeskind. Imaging of Ischemic Stroke. Technical Report 5, 2016.

[35] A. Lisowska, E. Beveridge, K. Muir, and I. Poole. Thrombus detection in CT brain scans using a convolutional neural network. In *BIOIMAGING 2017 - 4th International Conference on Bioimaging, Proceedings; Part of 10th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2017*, volume 2017-Janua, pages 24–33, 2017.

[36] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez. A survey on deep learning in medical image analysis, 12 2017.

[37] C.-F. Liu, J. Hsu, X. Xu, S. Ramachandran, V. Wang, M. I. Miller, A. E. Hillis, A. V. Faria, M. Wintermark, S. J. Warach, G. W. Albers, S. M. Davis, J. C. Grotta, W. Hacke, D.-W. Kang, C. Kidwell, W. J. Koroshetz, K. R. Lees, M. H. Lev, D. S. Liebeskind, A. G. Sorensen, V. N. Thijs, G. Thomalla, J. M. Wardlaw, and M. Luby. Deep learning-based detection and segmentation of diffusion abnormalities in acute ischemic stroke. *Communications Medicine 2021 1:1*, 1(1):1–18, 12 2021.

[38] M. Long, Y. Cao, J. Wang, and M. I. Jordan. Learning transferable features with deep adaptation networks. In *32nd International Conference on Machine Learning, ICML 2015*, volume 1, pages 97–105, 2015.

[39] O. Maier, B. H. Menze, J. von der Gablentz, L. Häni, M. P. Heinrich, M. Liebrand, S. Winzeck, A. Basit, P. Bentley, L. Chen, D. Christiaens, F. Dutil, K. Egger, C. Feng, B. Glocker, M. Götz, T. Haeck, H. L. Halme, M. Havaei, K. M. Iftekharuddin, P. M. Jodoin, K. Kamnitsas, E. Kellner, A. Korvenoja, H. Larochelle, C. Ledig, J. H. Lee, F. Maes, Q. Mahmood, K. H. Maier-Hein, R. McKinley, J. Muschelli, C. Pal, L. Pei, J. R. Rangarajan, S. M. Reza, D. Robben, D. Rueckert, E. Salli, P. Suetens, C. W. Wang, M. Wilms, J. S. Kirschke, U. M. Krämer, T. F. Münte, P. Schramm, R. Wiest, H. Handels, and M. Reyes. ISLES 2015 - A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Medical Image Analysis*, 35:250–269, 1 2017.

[40] K. Malhotra, J. Gornbein, and J. L. Saver. Ischemic strokes due to large-vessel occlusions contribute disproportionately to stroke-related dependence and death: A review. *Frontiers in Neurology*, 8(NOV), 11 2017.

[41] H. P. Mattle, M. Arnold, P. J. Lindsberg, W. J. Schonewille, and G. Schroth. Basilar artery occlusion, 11 2011.

[42] M. Meijs, F. J. Meijer, M. Prokop, B. v. Ginneken, and R. Manniesing. Image-level detection of arterial occlusions in 4D-CTA of acute stroke patients using deep learning. *Medical Image Analysis*, 66:101810, 12 2020.

[43] T. M. Mitchell. *Machine learning, International Edition*. McGraw-Hill Series in Computer Science. McGraw-Hill, 1997.

[44] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. De Vries, M. J. Benders, and I. Isgum. Automatic Segmentation of MR Brain Images with a Convolutional Neural Network. *IEEE Transactions on Medical Imaging*, 35(5):1252–1261, 5 2016.

[45] M. Mojtahedi, M. Kappelhof, E. Ponomareva, M. Tolhuisen, I. Jansen, A. A. Bruggeman, B. G. Dutra, L. Yo, N. Lecouffe, J. W. Hoving, H. van Voorst, J. Brouwer, N. A. Terreros, P. Konduri, F. J. Meijer, A. Appelman, K. M. Treurniet, J. M. Coutinho, Y. Roos, W. van Zwam, D. Dippel, E. Gavves, B. J. Emmer, C. Majoie, and H. Marquering. Fully Automated Thrombus Segmentation on CT Images of Patients with Acute Ischemic Stroke. *Diagnostics 2022, Vol. 12, Page 698*, 12(3):698, 3 2022.

[46] C. A. Molina and J. L. Saver. Extending reperfusion therapy for acute ischemic stroke: Emerging pharmacological, mechanical, and imaging strategies. *Stroke*, 36(10):2311–2320, 10 2005.

[47] M. A. Moskowitz, E. H. Lo, and C. Iadecola. The science of stroke: Mechanisms in search of treatments, 2010.

[48] K. B. Nelson and J. K. Lynch. Stroke in newborn infants. *The Lancet. Neurology*, 3(3):150–158, 3 2004.

[49] R. G. Nogueira, A. P. Jadhav, D. C. Haussen, A. Bonafe, R. F. Budzik, P. Bhuva, D. R. Yavagal, M. Ribo, C. Cognard, R. A. Hanel, C. A. Sila, A. E. Hassan, M. Millan, E. I. Levy, P. Mitchell, M. Chen, J. D. English, Q. A. Shah, F. L. Silver, V. M. Pereira, B. P. Mehta, B. W. Baxter, M. G. Abraham, P. Cardona, E. Veznedaroglu, F. R. Hellinger, L. Feng, J. F. Kirmani, D. K. Lopes, B. T. Jankowitz, M. R. Frankel, V. Costalat, N. A. Vora, A. J. Yoo, A. M. Malik, A. J. Furlan, M. Rubiera, A. Aghaebrahim, J.-M. Olivot, W. G. Tekle, R. Shields, T. Graves, R. J. Lewis, W. S. Smith, D. S. Liebeskind, J. L. Saver, and T. G. Jovin. Thrombectomy 6 to 24 Hours after Stroke with a Mismatch between Deficit and Infarct. *New England Journal of Medicine*, 378(1):11–21, 1 2018.

[50] O. Öman, T. Mäkelä, E. Salli, S. Savolainen, and M. Kangasniemi. 3D convolutional neural networks applied to CT angiography in the detection of acute ischemic stroke. *European radiology experimental*, 3(1):8, 2019.

[51] M. O. Owolabi, A. G. Thrift, A. Mahal, M. Ishida, S. Martins, W. D. Johnson, J. Pandian, F. Abd-Allah, J. Yaria, H. T. Phan, G. Roth, S. L. Gall, R. Beare, T. G. Phan, R. Mikulik, R. O. Akinyemi, B. Norrving, M. Brainin, V. L. Feigin, C. Abanto, S. F. Abera, A. Addissie, O. Adebayo, A. O. Adeleye, Y. Adilbekov, B. Adilbekova, T. A. Adoukonou, D. Aguiar de Sousa, T. Ajagbe, Z. Akhmetzhanova, A. Akpalu, J. Álvarez Ahlgren, S. Ameriso, S. Andonova, F. E. Awoniyi, M. Bakhiet, M. Barboza, H. Basri, P. Bath, O. Bello, D. Bereczki, S. Beretta, A. Berkowitz, A. Bernabé-Ortiz, J. Bernhardt, G. Berzina, M. Bisharyan, P. Bovet, H. Budincevic, D. Cadilhac, V. Caso, C. Chen, J. Chin, K. Chwojnicki, A. Conforto, V. T. Cruz, M. D'Amelio, K. Danielyan, S. Davis, V. Demarin, R. Dempsey, M. Dichgans, K. Dokova, G. Donnan, M. S. Elkind, M. Endres, U. Fischer, F. Gankpé, A. Gaye Saavedra, A. Gil, M. Giroud, E. Gnedovskaya, V. Hachinski, M. Hafdi, R. Hamadeh, T. K. Hamzat, G. Hankey, M. Heldner, E. A. Ibrahim, N. M. Ibrahim, M. Inoue, S. Jee, J. S. Jeng, Y. Kalkonde, S. Kamenova, B. Karaszewski, P. Kelly, T. Khan, S. Kiechl, A. Kondybayeva, J. Kõrv, M. Kravchenko, R. V. Krishnamurthi, J. Kruja, M. Lakkhanaloet, P. Langhorne, P. M. Lavados, Z. K. Law, A. Lawal, M. Lazo-Porras, D. Lebedynets, T. H. Lee, T. Leung, D. S. Liebeskind, P. Lindsay, P. López-Jaramillo, P. A. Lotufo, J. Machline-Carrion, A. Makanjuola, H. S. Markus, J. M. Marquez-Romero, M. Medina, S. Medukhanova, M. M. Mehndiratta, A. Merkin, E. Mirrakhimov, S. Mohl, M. Moscoso-Porras, A. Müller-Stierlin, S. Murphy, K. I. Musa, A. Nasreldein, R. G. Nogueira, C. Nolte, J. J. Noubiap, N. Novarro-Escudero, Y. Ogun, R. A. Oguntoye, M. I. Oraby, M. Osundina, B. Ovbiagele, D. N. Orken, A. . Ozdemir, S. Ozturk, M. Paccot, J. Phromjai, P. Piradov, T. Platz, T. Potpara, A. Ranta, F. Rathore, E. Richard, R. L. Sacco, R. Sahathevan, I. Santos Carquín, G. Saposnik, F. S. Sarfo, M. Sharma, K. Sheth, A. Shobhana, N. Suwanwela, I. Svyato, P. N. Sylaja, X. Tao, K. T. Thakur, D. Toni, M. A. Topcuoglu, J. Torales, A. Towfighi, T. C. Truelsen, A. Tsiskaridze, M. Tulloch-Reid, N. Useche, P. Vanacker, S. Vassilopoulou,

G. Vukorepa, V. Vuletic, K. W. Wahab, W. Wang, T. Wijeratne, C. Wolfe, Y. M. Yifru, A. Yock-Corrales, N. Yonemoto, L. Yperzeele, and P. Zhang. Primary stroke prevention worldwide: translating evidence into action. *The Lancet Public Health*, 7(1):e74–e85, 1 2022.

[52] S. J. Pan and Q. Yang. A survey on transfer learning, 2010.

[53] M. R. Patel, R. R. Edelman, and S. Warach. Detection of hyperacute primary intraparenchymal hemorrhage by magnetic resonance imaging. *Stroke*, 27(12):2321–2324, 1996.

[54] W. J. Powers, A. A. Rabinstein, T. Ackerson, O. M. Adeoye, N. C. Bambakidis, K. Becker, J. Biller, M. Brown, B. M. Demaerschalk, B. Hoh, E. C. Jauch, C. S. Kidwell, T. M. Leslie-Mazwi, B. Ovbiagele, P. A. Scott, K. N. Sheth, A. M. Southerland, D. V. Summers, and D. L. Tirschwell. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke a guideline for healthcare professionals from the American Heart Association/American Stroke A, 2019.

[55] K. Qi, H. Yang, C. Li, Z. Liu, M. Wang, Q. Liu, and S. Wang. X-Net: Brain Stroke Lesion Segmentation Based on Depthwise Separable Convolution and Long-Range Dependencies. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11766 LNCS, pages 247–255. Springer, 10 2019.

[56] T. N. Raju, K. B. Nelson, D. Ferriero, and J. K. Lynch. Ischemic perinatal stroke: Summary of a workshop sponsored by the National Institute of Child Health and Human Development and the National Institute of Neurological Disorders and Stroke. In *Pediatrics*, volume 120, pages 609–616. American Academy of Pediatrics, 9 2007.

[57] M. J. Rivkin, T. J. Bernard, M. M. Dowling, and C. Amlie-Lefond. Guidelines for Urgent Management of Stroke in Children. *Pediatric Neurology*, 56:8–17, 3 2016.

[58] O. Ronneberger, P. Fischer, and T. Brox. 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)*, volume 9351, pages 234–241, Munich, Germany, 2015. Springer Verlag.

[59] R. L. Sacco, S. E. Kasner, J. P. Broderick, L. R. Caplan, J. J. Connors, A. Culebras, M. S. Elkind, M. G. George, A. D. Hamdan, R. T. Higashida, B. L. Hoh, L. S. Janis, C. S. Kase, D. O. Kleindorfer, J. M. Lee, M. E. Moseley, E. D. Peterson, T. N. Turan, A. L. Valderrama, and H. V. Vinters. An Updated Definition of Stroke for the 21st Century. *Stroke*, 44(7):2064–2089, 2013.

[60] S. A. Sheth, V. Lopez-Rivera, A. Barman, J. C. Grotta, A. J. Yoo, S. Lee, M. E. Inam, S. I. Savitz, and L. Giancardo. Machine Learning-Enabled Automated Determination of Acute Ischemic Core From Computed Tomography Angiography. *Stroke*, 50(11):3093–3100, 11 2019.

[61] W. S. Smith. Intra-arterial thrombolytic therapy for acute basilar occlusion: Pro. In *Stroke*, volume 38, pages 701–703. Lippincott Williams & Wilkins, 2 2007.

[62] W. S. Smith, M. H. Lev, J. D. English, E. C. Camargo, M. Chou, S. C. Johnston, G. Gonzalez, P. W. Schaefer, W. P. Dillon, W. J. Koroshetz, and K. L. Furie. Significance of large vessel intracranial occlusion causing acute ischemic stroke and tia. *Stroke*, 40(12):3834–3840, 12 2009.

[63] M. T. Stib, J. Vasquez, M. P. Dong, Y. H. Kim, S. S. Subzwari, H. J. Triedman, A. Wang, H.-L. C. Wang, A. D. Yao, M. Jayaraman, J. L. Boxerman, C. Eickhoff, U. Cetintemel, G. L. Baird, and R. A. McTaggart. Detecting Large Vessel Occlusion at Multiphase CT Angiography by Using a Deep Convolutional Neural Network. *Radiology*, 297(3):640–649, 12 2020.

[64] M. L. Tolhuisen, E. Ponomareva, A. M. M. Boers, I. G. H. Jansen, M. S. Koopman, R. Sales Barros, O. A. Berkhemer, W. H. van Zwam, A. van der Lugt, C. B. L. M. Majoie, and H. A. Marquering. A Convolutional Neural Network for Anterior Intra-Arterial Thrombus Detection and Segmentation on Non-Contrast Computed Tomography of Patients with Acute Ischemic Stroke. *Applied Sciences*, 10(14):4861, 7 2020.

[65] N. Tomita, S. Jiang, M. E. Maeder, and S. Hassanpour. Automatic post-stroke lesion segmentation on MR images using 3D residual convolutional neural network. *NeuroImage: Clinical*, 27, 1 2020.

[66] G. Turc, P. Bhogal, U. Fischer, P. Khatri, K. Lobotesis, M. Mazighi, P. D. Schellinger, D. Toni, J. De Vries, P. White, and J. Fiehler. European Stroke Organisation (ESO) - European Society for Minimally Invasive Neurological Therapy (ESMINT) Guidelines on Mechanical Thrombectomy in Acute Ischemic Stroke. *Journal of NeuroInterventional Surgery*, 0:1–30, 3 2019.

[67] B. Voetsch, L. D. DeWitt, M. S. Pessin, and L. R. Caplan. Basilar Artery Occlusive Disease in the New England Medical Center Posterior Circulation Registry. *Archives of Neurology*, 61(4):496–504, 4 2004.

[68] N. Wagenaar, C. G. De Theije, L. S. De Vries, F. Groenendaal, M. J. Benders, and C. H. Nijboer. Promoting neuroregeneration after perinatal arterial ischemic stroke: neurotrophic factors and mesenchymal stem cells. *Pediatric Research 2018 83:1*, 83(1):372–384, 11 2017.

[69] S. J. Warach, M. Luby, G. W. Albers, R. Bammer, A. Bivard, B. C. Campbell, C. Derdeyn, J. J. Heit, P. Khatri, M. G. Lansberg, D. S. Liebeskind, C. B. Majoie, M. P. Marks, B. K. Menon, K. W. Muir, M. W. Parsons, A. Vagal, A. J. Yoo, A. V. Alexandrov, J. C. Baron, D. J. Fiorella, A. J. Furlan, J. Puig, P. D. Schellinger, and M. Wintermark. Acute Stroke Imaging Research Roadmap III Imaging Selection and Outcomes in Acute Stroke Reperfusion Clinical Trials: Consensus Recommendations and Further Research Priorities. *Stroke*, 47(5):1389–1398, 5 2016.

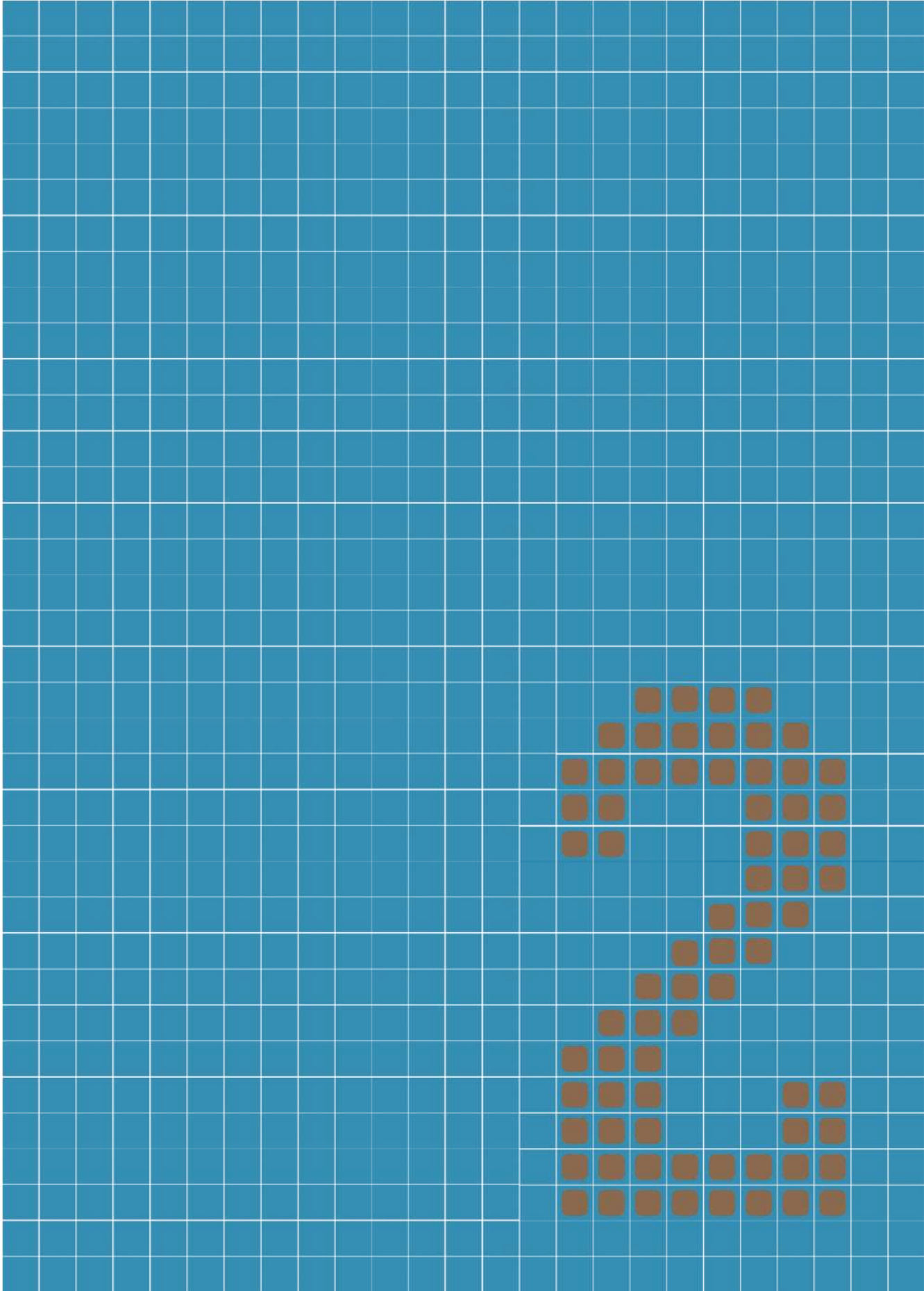
[70] A. J. Yoo, Z. A. Chaudhry, R. G. Nogueira, M. H. Lev, P. W. Schaefer, L. H. Schwamm, J. A. Hirsch, and R. G. González. Infarct volume is a pivotal biomarker after intra-arterial stroke therapy. *Stroke*, 43(5):1323–1330, 5 2012.

[71] Y. Yu, Y. Xie, T. Thamm, E. Gong, J. Ouyang, C. Huang, S. Christensen, M. P. Marks, M. G. Lansberg, G. W. Albers, and G. Zaharchuk. Use of Deep Learning to Predict Final Ischemic Stroke Lesions From Initial Magnetic Resonance Imaging. *JAMA network open*, 3(3):e200772, 3 2020.

[72] R. Zhang, L. Zhao, W. Lou, J. M. Abrigo, V. C. Mok, W. C. Chu, D. Wang, and L. Shi. Automatic Segmentation of Acute Ischemic Stroke From DWI Using 3-D Fully Convolutional DenseNets. *IEEE Transactions on Medical Imaging*, 37(9):2149–2160, 9 2018.

[73] Y. Zhou, W. Huang, P. Dong, Y. Xia, and S. Wang. D-UNet: A dimension-fusion U shape network for chronic stroke lesion segmentation, 8 2019.

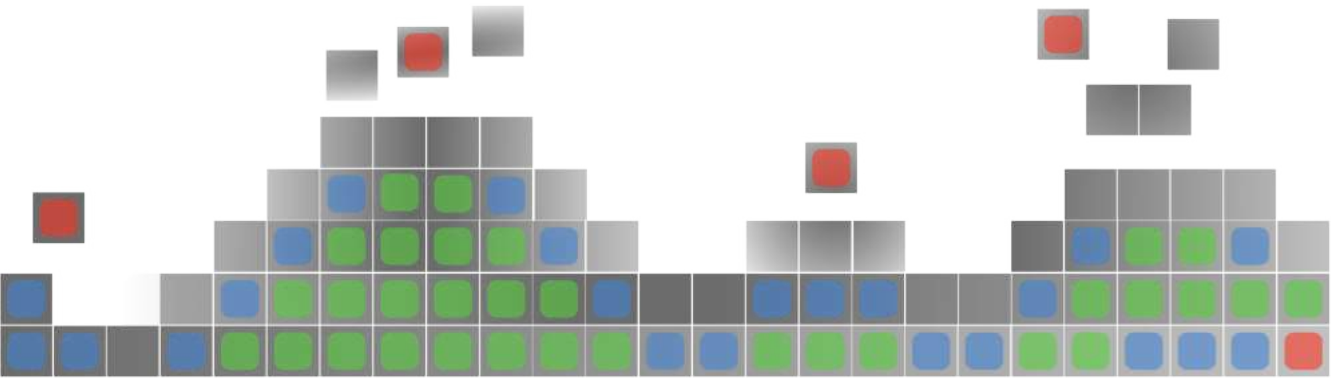
[74] Z. Zhou, V. Sodha, M. M. Rahman Siddiquee, R. Feng, N. Tajbakhsh, M. B. Gotway, and J. Liang. Models genesis: generic autodidactic models for 3d medical image analysis. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11767 LNCS, pages 384–393, 2019.



# Chapter 2

# Domain- and Task-Specific Transfer Learning For Medical Segmentation Tasks

**Based on:** R. Zoetmulder, E. Gavves, M. Caan, and H.A. Marquering. "Domain- and Task-Specific Transfer Learning For Medical Segmentation Tasks", Computer Methods and Programs in Biomedicine, vol 214 (2022).



# Abstract

**Background and objectives:** Transfer learning is a valuable approach to perform medical image segmentation in settings with limited cases available for training convolutional neural networks (CNN). Both the source task and the source domain influence transfer learning performance on a given target medical image segmentation task. This study aims to assess transfer learning-based medical segmentation task performance for various source task and domain combinations.

**Methods:** CNNs were pre-trained on classification, segmentation, and self-supervised tasks on two domains: natural images and T1 brain MRI. Next, these CNNs were fine-tuned on three target T1 brain MRI segmentation tasks: stroke lesion, MS lesions, and brain anatomy segmentation. In all experiments, the CNN architecture and transfer learning strategy were the same. The segmentation accuracy on all target tasks was evaluated using the mIOU or Dice coefficients. The detection accuracy was evaluated for the stroke and MS lesion target tasks only.

**Results:** CNNs pre-trained on a segmentation task on the same domain as the target tasks resulted in higher or similar segmentation accuracy compared to other source task and domain combinations. Pre-training a CNN on ImageNet resulted in a comparable, but not consistently higher lesion detection rate, despite the amount of training data used being 10 times larger.

**Conclusions:** This study suggests that optimal transfer learning for medical segmentation is achieved with a similar task and domain for pre-training. As a result, CNNs can be effectively pre-trained on smaller datasets by selecting a source domain and task similar to the target domain and task.

## 2.1 Introduction



Convolutional Neural Networks (CNN) have become the standard approach for medical image segmentation [93]. Accurate CNN-based segmentation approaches typically require a large amount of manually annotated data for training. However, manual annotation of medical images is commonly a time-consuming task, which may require specialized expertise. Reducing the demand for large annotated datasets is therefore an active area of research.

In this study, we focus on transfer learning, which is a broadly applicable strategy to reduce the need for annotated data. Transfer learning aims to reuse a CNN trained on a large dataset rather than directly training a CNN from scratch [105]. In this approach, the weights that are obtained by pre-training are subsequently used to initialize a CNN and perform a different medical image analysis task on a different dataset. The source domain and the source task during pre-training are two relevant aspects of transfer learning [114]. The source domain refers to the type of data used, and the source task refers to the specific application used to pre-train. Analogously, the target domain and target task refer to the type of data and specific application of the main goal. The pre-trained weights can be used on a target task using two strategies: feature extraction or fine-tuning. In feature extraction, the transferred weights are fixed when learning the target task. In fine-tuning, the transferred weights are updated to perform the target task.

Prior studies on transfer learning for medical segmentation target tasks have mostly used two source domains; natural images and medical images. Previous studies that used natural images as the source domain have used two source tasks: ImageNet classification [121] and image segmentation. Two examples with ImageNet classification as the source task are brain-tumor segmentation [77] and stroke lesion segmentation [91]. An example of a study that used natural image segmentation as the source task is colorectal polyp segmentation [98]. The value of pre-training on natural image datasets for medical image analysis target tasks is not clear enough yet. This is because natural image datasets differ from medical image datasets in three important ways. Firstly, medical classification and segmentation tasks often contain a few classes [76; 109] whereas natural image classification and segmentation

tasks can contain hundreds of different classes [121]. Secondly, natural images are made in heterogeneous settings, whereas medical images are acquired in controlled settings. Hence, the variation in terms of object orientation is larger in natural images than in medical images. Thirdly, natural color images commonly have three channels representing the colors red, green and blue. Whereas scans in medical datasets often do not consist of three channels. This is for example the case if the medical dataset consists of Magnetic Resonance (MR) or CT scans. Recent work already showed that pre-training on a gray scale version of ImageNet improves transfer learning performance [131]. Transfer learning with medical images as the source domain include segmentation and self-supervised tasks as the source tasks. For example, segmentation source tasks have been used to improve white matter lesions segmentation [87], neonatal brain tissue type segmentation [136], and lung nodule and liver tumor segmentation [80]. Self-supervised source tasks have been used to improve lung nodule segmentation [140].

The choice of the source task has been shown to influence the target task performance. On natural images as the target domain, it has been shown that selecting source tasks that were more similar to their target tasks resulted in better performance on the target task [135]. However, for medical image segmentation target tasks, this has not been established.

If we categorize imaging tasks as self-supervised, classification and segmentation tasks, for the natural image source domain, studies have used classification [77; 91] and segmentation [98], but not self-supervised source tasks. Differently, for the medical image source domain, studies have used segmentation [80; 87; 136] and self-supervised [140], but not classification source tasks. It can therefore be concluded that the effect of the source domain and tasks on the target medical segmentation accuracy has not yet extensively been evaluated.

In the current study, we empirically investigate the effect of the choice of source task and domain on the performance of multiple medical segmentation target tasks: stroke lesion, MS lesion, and brain anatomy segmentation on T1 MR. Furthermore, we aim to compare the optimal source-target task/domain combination with a common benchmark in transfer learning research: pre-training on ImageNet.

## 2.2 Related Work

### 2.2.1 Transfer Learning

The goal of transfer learning is to pre-train a model on a source task and reuse the information the model has learned to improve performance on a target task [115]. Transfer learning was first shown to work in neural networks by [117] and was subsequently applied to problems in computer vision [133] and medical image analysis [105].



A commonly used approach to apply transfer learning is to pre-train the CNN on a task and domain and to (partially) fine-tune the CNN on a target task and domain. In computer vision, the CNN is often pre-trained on the ILSVRC'12 (ImageNet) dataset [121]. Work on medical image analysis has used CNNs pre-trained on ImageNet classification [111] as well. Recently, the use of CNNs pre-trained on ImageNet for medical image analysis has been questioned. Recent work has found that the transfer learning benefits gained from pre-training on ImageNet classification were inconsistent on diabetic retinopathy grade classification on fundus photographs and thoracic pathology classification on chest X-Ray scans [118]. As a result, other data sets and tasks have been investigated as alternatives to ImageNet classification for transfer learning in medical image analysis [80; 140].

Other research has developed alternative methodologies to the pre-training and fine-tuning procedure that is widely used [90; 134]. Spot Tune is a method that adaptively decides to freeze or fine-tune specific layers in the CNN for each input image [90]. Co-Tuning is a method that fully re-uses the pre-trained CNN by learning a mapping from the target classes to the source classes and uses these labels as an additional supervision signal during fine-tuning.

### 2.2.2 Domain Adaptation

Domain adaptation (or *transductive transfer learning*) is a special case of transfer learning in which the source and target task are the same but the data distribution of the source and target domains differ[114]. The goal of domain adaptation is to build domain invariant models that learn similar features from the source and target domains. Techniques are based on minimizing

the difference between the feature distributions acquired from the source and target domain [86; 106; 122; 128; 129; 138]. For example, prior work has proposed a method by which statistical dependence was preserved by using a reproducing kernel Hilbert space [129]. Other work has proposed a manifold criterion to create an intermediate domain, which is related to the target domain, using source data [138].



Research in medical image analysis has also made use of domain adaptation [87; 89; 97]. Prior research on MR has applied domain adaptation to generalize automated segmentation of white matter hyper intensities to follow-up scans using fine-tuning [87]. Another method has used adversarial learning to generalize segmentation of abnormalities on brain MR scans after traumatic brain injury [97].

### 2.2.3 Task Transfer Learning

Task transfer learning (or *inductive transfer learning*) is a special case of transfer learning in which the source and target task differ but the data distribution of the source and target domains is the same [114]. In computer vision, several studies have investigated the relationship between different tasks [75; 126; 135]. One study investigated the relationship between individual source and target tasks to create a taxonomy of the degree to which tasks transfer to each other by fine-tuning each target task on each source task [135]. This study also investigated the performance gain achieved when features from models pre-trained on different source tasks were combined to learn a target task. Developing a practical method to decide which source tasks are the most important in decision making support when computational resources are limited has consequently been investigated [79]. Other research has focused avoiding having to fine-tune networks by finding the affinity between various classification tasks [75].

### 2.2.4 Few Shot Learning

Few-shot learning is a machine learning sub-field that aims to learn from a few training examples (for example five cases) per individual class [130]. In computer vision, much research has been dedicated to developing few-shot

learning methods [84; 100; 137]. For medical image analysis, few-shot learning has also been adopted for organ segmentation [113; 120].

## 2.3 Materials



This section describes the used datasets, the pre-processing, and the CNN architectures.

### 2.3.1 Datasets

Two natural image datasets were used to pre-train the CNNs: the **taskonomy** dataset [135], consisting of 4.6 million images of indoor scenes with multiple annotations per image, and the **ImageNet** dataset [121] consisting of 1.2 million images of 1000 different objects.

Four medical datasets that consist of T1 brain MRIs were used in our study. Firstly, we have used the Brain Age Healthy Cohort (BAHC) [82] to pre-train the CNNs. The BAHC is a dataset compiled of 2001 scans from 14 different data sources. Ground truth annotations were created using a combination of Nipype [88], FSL [125] and ITK [132]. Additional information about participants and the scan acquisition parameters can be found in table 2.3, in appendix 2.9.1. Secondly, we have used the Anatomical Tracings of Lesions After Stroke (ATLAS) R1.2 **(Stroke lesion dataset)** [104], which is a manually annotated T1 MRI dataset of lesions after ischemic stroke and consists of scans from 304 patients. Each scan contains at least one lesion. The annotations consist of a primary lesion and secondary non-contiguous lesions. Additional information about the scans is included in table 2.4, appendix 2.9.1. Thirdly, we have included 30 scans of the **Multiple Sclerosis (MS) lesion** dataset [102]. The scans were acquired using a 3T Siemens Magnetom Trio. The resulting scans had a resolution of $0.57 \times 0.57 \times 3.003.30$ mm. Fourthly, we included the **Brain Anatomy (BA) dataset** which consists of 35 scans from the OASIS project [108]. Manual annotations were combined into six classes [110]. The images were acquired on a Siemens Vision 1.5T scanner and had a resolution of $1.0 \times 1.0 \times 1.25$ mm.

### 2.3.2 Pre-processing

All scans were reoriented and resampled to the MNI-ICBM 152 template [85], axially zero padded to the taskonomy dataset dimensions. Empty axial slices were discarded, being 61, 29, 11 and 11 slices starting from the top of the scan volume for the BAHC, stroke lesion, MS lesion, and BA dataset respectively. For the MS lesion and BA dataset also 50 slices from the bottom were discarded. Voxel intensities above the $99^{th}$ and below the $1^{st}$ percentile were clipped, and intensities were then normalized using min-max normalization.

### 2.3.3 CNN Architectures

This study uses CNNs developed in earlier work [135]. The CNN architectures consist of an encoder (figure 2.1A) along with only a decoder (figure 2.1B), an encoder along with a decoder and a discriminator, (figure 2.1C) or an encoder along with a fully connected layer (figure 2.1D). The CNNs were pre-trained on the source tasks.

#### *The Encoder*

The encoder is based on the ResNet-50 [92] architecture. To ensure that the encoder has a latent space of $16 \times 16 \times 4$, the fully connected layers of the ResNet-50 are replaced by a transposed convolution with a stride of two. The encoder uses the ReLU activation function [107].

#### *The Decoder*

The decoder consists of convolutions and transposed convolutions. It up-samples the features from $16 \times 16 \times 4$ to an image of $256 \times 256$ with the number of channels required for the task. Each convolution layer had a stride of one and each transposed convolution had a stride of two. The kernel size was $3 \times 3$. The first two layers were convolutional layers, the eight subsequent layers alternated between a transposed convolutional layer and regular convolutional layer, ending with the latter. The decoder used a leaky ReLU activation function [107] with alpha set to 0.2.

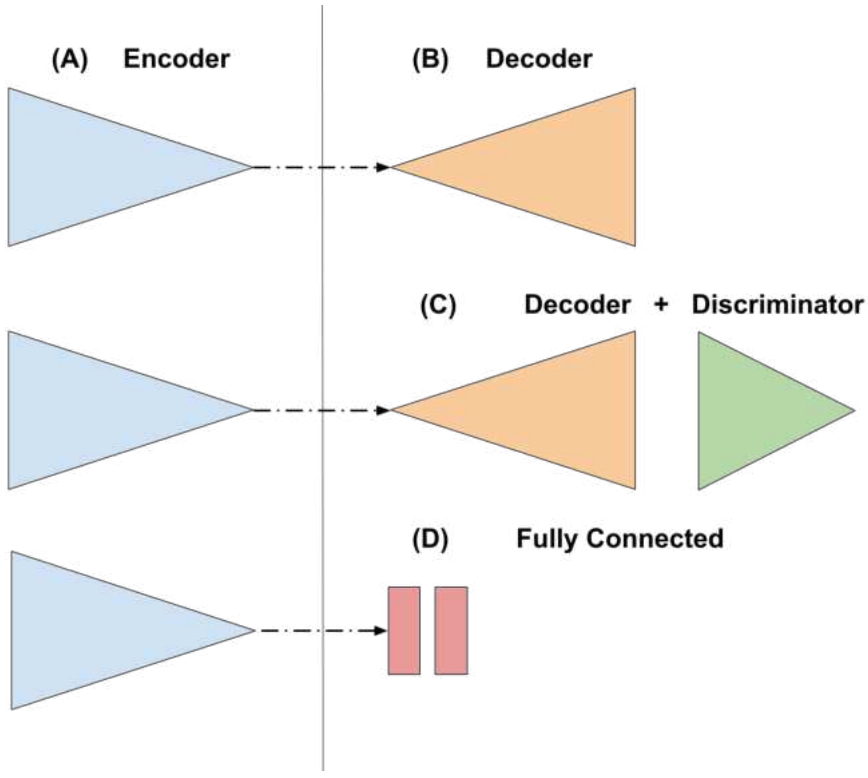



**Figure 2.1:** Components of the CNN architectures. A CNN architecture consists of an encoder (A) followed by either a decoder (B), a decoder and a discriminator (C), or a fully connected layer (D).

### *The Discriminator*

The image and the decoder output were used as input of the discriminator[96]. The first convolutions had a stride of one and a kernel size of five. Next, two convolutional layers followed with a stride of four. The final two layers had a kernel size of four and a stride of one. The discriminator used a leaky ReLU activation function [107] with alpha set to 0.2.

### *The Fully Connected Block*

The fully connected block consisted of two fully connected layers. The first layer had a hidden size of 2048, the second layer a size of 16. The first fully connected layer [95] uses a ReLU activation function[107] and the second fully connected layer uses a softmax activation function.

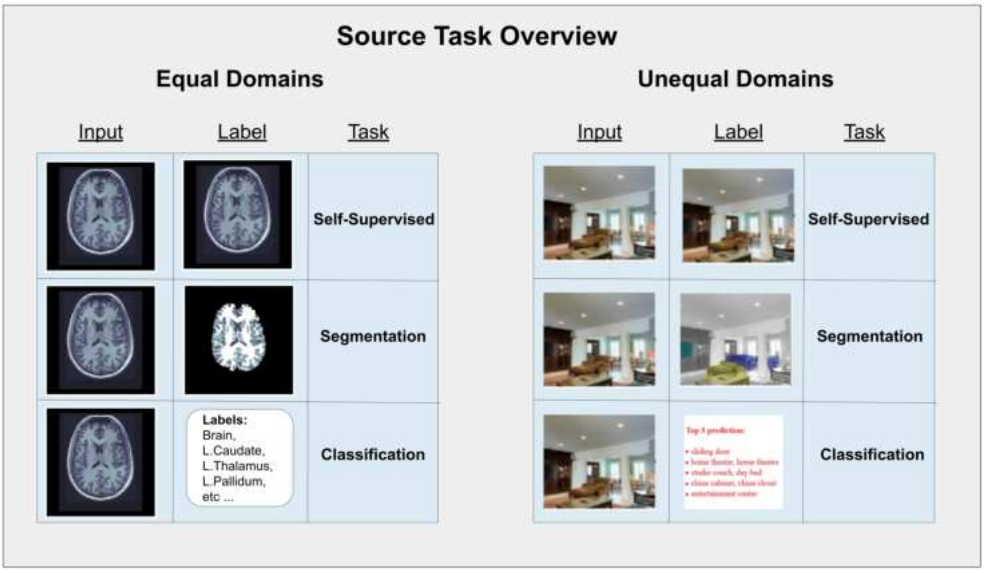



**Figure 2.2:** Schematic representation of the source tasks for the equal (T1 MRI) (left) and unequal (natural images) (right) domain. For each task, an example is given of the input (left column), the label (middle column), and the task name (right column). The source tasks are; self-supervised, segmentation, and classification.

## 2.4 Methods: Source and Target Tasks

The source and target domains have been addressed in section 2.3. In the current section, the source and target tasks are addressed.

### 2.4.1 Source Tasks

A schematic representation of the source tasks of the equal (T1 MRI) and unequal (natural image) domain is shown in figure 2.2. The hyper-parameters used to pre-train the CNNs on the equal domain as the target tasks are described in appendix 2.9.2. The CNNs were pre-trained on axial slices from the scans. The hyperparameters used to pre-train the CNNs on the unequal domain are described in appendix 2.9.3.

#### *Segmentation Source Tasks*

The segmentation source tasks were brain tissue and indoor object segmentation for the equal and unequal domain, respectively. The brain tissue ground truth for three classes was generated using the FAST algorithm [139] from the FSL toolkit. The indoor object ground truth was available for 17 classes [103] & [135]. The segmentation CNN architecture consisted of the encoder, followed by the decoder. The loss function used was the weighted cross-entropy and used class weights calculated by [78].

#### *Self-Supervised Source Task*

For both source domains, auto-encoding was the self-supervised source task [94]. The auto-encoding CNN architecture consisted of an encoder, a decoder and a discriminator. The loss consisted of the weighed sum of the $L_1$ norm and the GAN loss. The weights used were 0.996 for the $L_1$ norm and 0.004 for the Generative Adversarial Network (GAN) loss.

#### *Classification Source Tasks*

The equal domain classification source task was brain and sub-cortical structure classification, which was a multi-label classification task. Each axial slice contained annotations indicating whether the brain and specific sub-cortical structures were present. Ground truth segmentations of these structures were generated using the FIRST [116] and BET [124] algorithm from the FLS toolkit. The classification of a subset of 100 ImageNet classes of indoor scenes was used as the unequal domain object classification task[135]. The binary cross entropy was used as the loss function class-wise. The used CNN architecture was the encoder followed by a fully connected block.

### 2.4.2 Comparison to pre-training on the full-extent of ImageNet

The previously described transfer learning experiments include source tasks with a similar amount of data to pre-train for a fair comparison. The most commonly used source task, ImageNet classification [121], uses at least ten

times more data than the models described above. To compare the other approaches to the the most commonly used benchmark, a CNN pre-trained on the full-extent of the ImageNet classification dataset was included as a source task.



### 2.4.3 Target Tasks

For all experiments, the encoder was initialized using one of the source tasks, and the decoder was initialized randomly. The CNNS were fine-tuned with multiple sub-sample sizes, which will be referred to as the **fine-tuning set size**. The CNNs were fine-tuned on axial slices obtained from the included scans.

All CNNs used a batch size of 32, a learning rate of $10^{-4}$ and a weight decay of $2 \cdot 10^{-4}$. The number of epochs varied per task and are discussed per experiment. After half the training epochs were completed, the learning rate was decayed by 10. For the segmentation tasks, the weighted cross entropy was calculated voxel wise.

#### *Stroke Lesion Segmentation*

The stroke lesion segmentation target task consisted of segmenting stroke lesions from non affected tissue and background.

The data was split randomly into a training and testing set of 200 and 104 scans, respectively. The fine-tuning set size was incremented from 10 to 100 scans with steps of 10. For each fine-tuning set size, ten fine-tuning sets were randomly sampled from the training set. CNNs were fine-tuned for 30 epochs.

#### *Multiple Sclerosis Lesion Segmentation*

The MS lesion segmentation task consisted of segmenting MS lesions from non affected tissue and background. The MS lesion data was split randomly into a fine-tuning set of 20 scans and a testing set of 10 scans, respectively. CNNs were fine-tuned for 60 epochs.

#### *Brain Anatomy Segmentation*

The brain anatomy segmentation task consisted of segmenting seven anatomical regions. The BA dataset was divided into 15 scans for fine-tuning and 20 for testing [110]. CNNs were fine-tuned for 300 epochs.



### 2.4.4 Evaluation Metrics

The Dice coefficient and the mean Intersection Over Union (mIOU) were used to assess the spatial accuracy for single and multi-class segmentation tasks, respectively. Both measures were calculated per axial slice and averaged over the entire test set.

For the MS and stroke lesion segmentation tasks, the lesion detection accuracy was assessed. A connected component analysis was performed on the ground truth segmentation mask to separate all individual lesions in the mask. Voxels were considered to be part of a common lesion if they were 8-connected in the mask. A lesion was considered detected if the percentage of accurately automatically quantified voxels exceeded a pre-set threshold. This threshold is referred to as the *voxel overlap cutoff*. We used voxel overlap cutoff values of 0% and 20%. The higher the voxel overlap cutoff, the more difficult it is for a lesion to be detected by the model. In addition, we wanted to assess the lesion detection accuracy for different lesion volumes. To this end, we progressively excluded lesions below a pre-set volume. We refer to this parameter as the *lesion volume cutoff*, and we used pre-set values of 0 mL (including all lesions), 1 mL and 2 mL. For stroke lesion detection, this analysis was conducted on one of the CNNs fine-tuned with a fine-tuning set size of 100.

## 2.5 Experiments & Results

### 2.5.1 Stroke Lesion Segmentation

The Dice coefficient and lesion detection accuracy for the stroke lesion segmentation models are shown in figure 2.3. For the equal domain experiments (figure 2.3a), the segmentation source task transfer learning model resulted

in the largest Dice coefficient. If the voxel overlap cutoff was set to zero, the lesion detection accuracy was similar for models pre-trained on each source task. When the voxel overlap cutoff was set to 20%, the lesion detection accuracy dropped overall. However, the segmentation source task resulted in models with a higher lesion detection accuracy than the other models.

For unequal source domains, shown in figure 2.3b, the self-supervised source task resulted models with the highest Dice coefficient. No single source task resulted in a model with a higher lesion detection accuracy than models pre-trained on the other source tasks.

Comparing equal and unequal domains, the best equal domain source task (segmentation) resulted in models with a higher Dice coefficient than the other source tasks. Models pre-trained on the classification source tasks consistently yielded a low Dice coefficient. Results for the lesion detection accuracy were more ambiguous. With a voxel overlap cutoff greater than 20%, the model pre-trained on the best equal domain source task resulted in a higher lesions detection accuracy. However, a voxel overlap cutoff greater than 0% resulted in the model pre-trained on the unequal domain source tasks achieving a higher lesion detection accuracy.

In comparison to the full-extent ImageNet pre-trained model, the equal source domain and task model obtained the highest Dice coefficient (figure 2.3c). The lesion detection accuracy is similar for both approaches (figure 2.3c). However, when the voxel overlap cutoff was set to 20%, the ImageNet pre-trained model detected more lesions.

### 2.5.2 Brain Anatomy Segmentation

Spatial agreement results for the BA segmentation target task are shown in table 2.1. For the equal domain, the results show that the model pre-trained on the segmentation source task resulted in the highest mIOU. For the unequal domain, all models had lower accuracy than the model pre-trained on the best equal domain source task. Even the full-extent ImageNet pre-trained model was outperformed by the best performing equal domain transfer learning model.

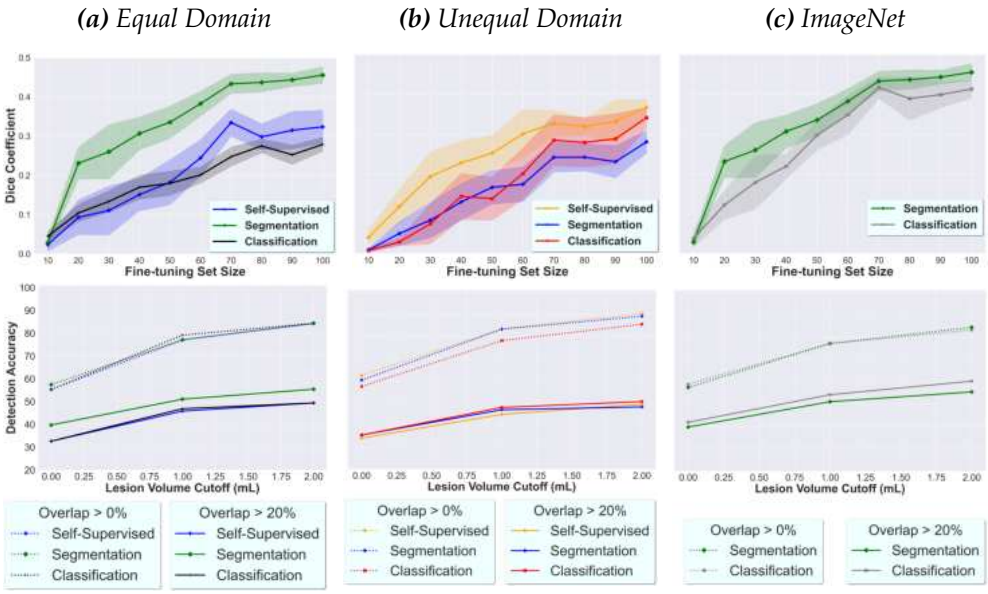



**Figure 2.3:** Spatial agreement (top) and lesion detection accuracy (bottom) for the stroke lesion segmentation task. The spatial agreement is assessed using the Dice coefficient as a function of the fine-tuning set size. The lesion detection accuracy is given as a function of the lesion volume cutoff, which is the minimal volume for which lesions are considered detected. The lesion detection accuracy is calculated with 0% and 20% overlap cutoffs (bottom). (a) Equal domain (T1-MRI) pre-training. Included source tasks are self-supervised, segmentation, and classification. (b) Unequal domain (natural images) pre-training. Included source tasks are self-supervised, classification, and segmentation. (c) Comparison of full-extent ImageNet classification pre-training and pre-training on equal domain and task: brain tissue segmentation.

### 2.5.3 Multiple Sclerosis Lesion Segmentation

The results for the MS lesion segmentation show that the Dice coefficient was generally low, with segmentation and the self-supervised source task resulting in the highest Dice coefficient for equal and unequal domain, respectively. In addition, the classification source task on the equal and unequal domain consistently results in a low Dice coefficient. The results are shown in table 2.2.

The MS lesion detection results are shown in figure 2.4. For the equal domain, the self-supervised source task resulted in models with the highest lesion detection accuracy, regardless of the voxel overlap or lesion volume

| | Brain Anatomy Segmentation |
|---|---|
| **Equal Domain: T1 MR** | |
| Segmentation | <u>0.62</u> |
| Self-Supervised | 0.58 |
| Classification | 0.59 |
| **Unequal Domain: Natural Images** | |
| Segmentation | 0.56 |
| Self-Supervised | 0.57 |
| Classification | 0.55 |
| ImageNet Classification | 0.52 |

**Table 2.1:** mIOU of the brain anatomy multi-class segmentation task for various source domains (T1 MRI vs natural images) and source tasks. The self-supervised task is autoencoding. The highest mIOU is underlined.

| | MS Lesion Segmentation |
|---|---|
| **Equal Domain: T1 MR** | |
| Segmentation | 0.16 |
| Self-Supervised | 0.12 |
| Classification | 0.14 |
| **Unequal Domain: Natural Images** | |
| Segmentation | 0.14 |
| Self-Supervised | 0.16 |
| Classification | 0.13 |
| ImageNet Classification | <u>0.17</u> |

**Table 2.2:** Dice coefficient for the MS lesion segmentation for various source domains. The equal domain is T1 MR, and the unequal domain is natural images. The highest Dice coefficient is underlined. The self-supervised tasks are autoencoding.

cutoff.

For the unequal domain, the classification and segmentation source tasks resulted in models with a higher lesion detection accuracy regardless of the lesion volume cutoff. The full-extent ImageNet pre-trained model resulted in a slightly higher Dice coefficient and lesion detection accuracy relative to the best performing equal domain source task.

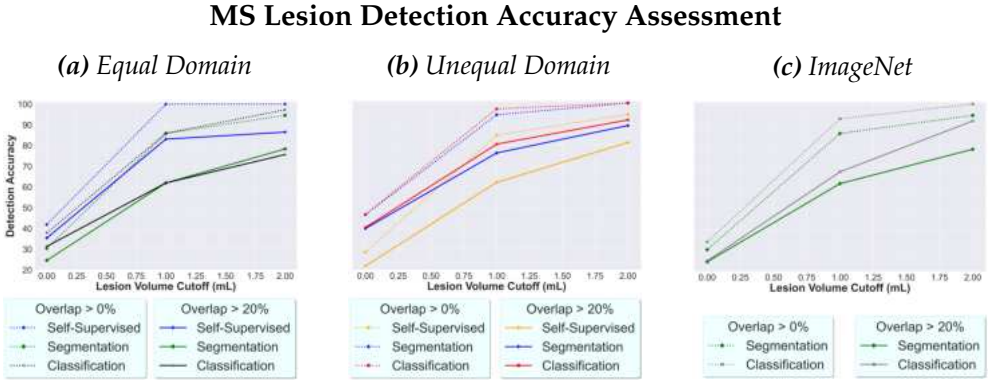



**Figure 2.4:** MS lesion detection accuracy as a function of the lesion volume cutoff with 0% and 20% volume overlap cutoffs. (a) Equal domain pre-training (T1 MRI) for the self-supervised, segmentation, and classification source tasks. (b) Unequal domain (natural images) pre-training for the source tasks classification, and segmentation. (c) Pre-training with the full-extent ImageNet classification source task vs medical image segmentation source task.

### 2.5.4 Qualitative Analysis

We performed a qualitative analysis by visual comparison of the automatically generated segmentation masks. Examples of these masks are shown in figure 2.5.

Visual inspection resulted in two observations. Firstly, pre-training on the segmentation task on an equal domain resulted in the largest segmentations. Secondly, the self-supervised source task pre-trained on the unequal domain resulted in implausible segmentations. For example, it falsely predicted lesions in both hemispheres.

For the BA segmentation, the ImageNet pre-trained model resulted in a larger number of false positives for white matter segmentation. In addition, the ImageNet pre-trained model resulted in a larger number of false positives and negatives of the cerebellum.

## 2.6 Discussion

Our study provides the first empirical comparison between the most frequently chosen source tasks for medical segmentation target tasks. These source tasks are self-supervised, classification and segmentation on both nat-

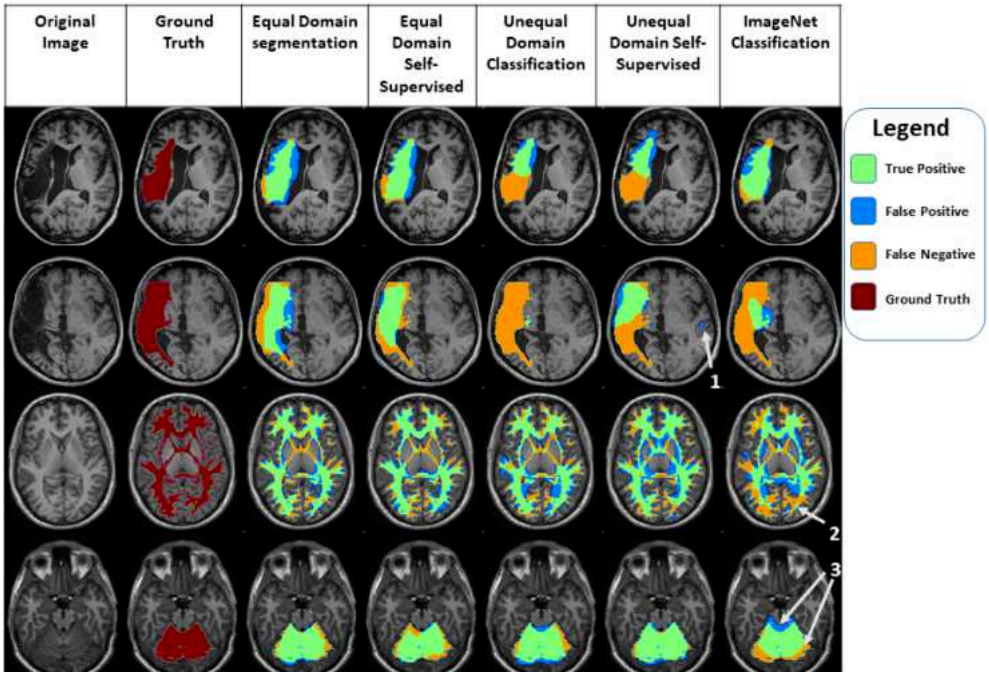



**Figure 2.5:** Examples of segmentation results obtained by the different source tasks on the target tasks stroke lesion (top two rows) and BA (bottom two rows) segmentation. From left to right, the original T1 MR scan, the ground truth segmentation and different segmentation results are shown after transfer learning using the source tasks: segmentation, the self-supervised source task, and classification on the same domain, the self-supervised source task on an unequal domain and full-extent ImageNet pre-training.

ural images and medical images for medical segmentation transfer learning performance. Previous work gave evidence of the advantage of transfer learning for medical segmentation target tasks using various source tasks and domains [80; 87; 136; 140]. We build on top of this work by studying how transfer learning performance was influenced by the choice of source task and domain. Our findings corroborate those found in a single source domain [135]; source tasks that are more similar to the target task result in higher transfer learning performance.

A possible explanation of our finding, is that higher layers are more specialized to perform the source task [133]. As such, these layers have a better initialization to perform tasks that are similar to the source task, which may result in finding a better optimum after fine-tuning.

Research comparing the transfer learning performance of various source tasks in medical image analysis has focused mostly on target classification tasks. Previous work showed that supervised and self-supervised pre-training on CT scans yielded a lower transfer learning performance than self-supervised pre-training on natural images [123]. Another study provided evidence that pre-training on a large natural image dataset resulted in an equivalent classification performance using less data, faster convergence, greater robustness against domain shift, and little influence on the calibration of uncertainty estimation [112]. However, other work provided evidence that pre-training on natural images only resulted in faster convergence of the networks but did not result in an improvement of classification performance [118].

Studies done on natural images have similar findings to our own; similar domains and tasks result in optimal transfer learning performance. This was shown for various classification target tasks [75] and a segmentation target task [83]. However, a domain shift can cause the most similar source task to the target segmentation task to result in sub-optimal transfer learning performance [83]. This result is corroborated by another study, which showed that increasing the amount of data used for pre-training could adversely influence transfer learning performance on classification target tasks if the additional data was not from a similar domain [127]. Our study observes the same phenomenon in transfer learning for medical image segmentation.

To promote comparability, our work firstly focused on a single target domain, allowing us to exploit large amounts of data available to pre-train CNNs and the existence of open source tools to automatically create segmentation annotations. Secondly, we only used the ResNet-50 architecture, because pre-trained weights were available for natural image analysis tasks. The specialized U-Net [119] architecture for medical image analysis tasks does not have weights from pre-training tasks on natural images available. Furthermore, U-Net does not generalize well to classification tasks in a straightforward manner because of skip connections. Thirdly, we chose fine-tuning and not feature extraction by pre-trained weights. Hereby, we limited the influence of otherwise confounding variables on the accuracy assessments of transfer learning.

A first limitation of our approach is that we have only tested two source

domains, natural images and T1 MR brain scans, and one target domain, T1 MR brain scans. However, there is evidence from other studies that pre-training a model on an equal domain to the target task results in similar better results [80; 123]. Hence, it is reasonable to assume this would apply to other medical image domains as well.



A second limitation of our research is the focus on 2D segmentation. Methods that use 3D self-supervised source tasks result in better performance than 2D ImageNet pre-trained models [140]. There is prior evidence showing that transfer learning strategies work well across different architectures [80; 81; 101] . Therefore, we expect our findings in 2D to generalize to 3D as well.

A third limitation is that the Dice coefficient was low for all MS lesion segmentation target task regardless of the pre-training approach. In T1-weighed MRI, MS lesions in white matter appear as slightly hypo-intense, with intensities similar to gray matter. This makes segmenting MS lesions a challenging task. We found that transfer learning is of limited additional benefit for this challenging task.

A fourth limitation in our study is that we have used a single CNN architecture, i.e. ResNet-50. This architecture was used because pre-trained weights were available for all necessary natural image tasks. Several studies have shown that transfer learning strategies work well across different architectures [80; 81; 101]. Considering the significant amount of additional computational resources needed to pre-train additional models and fine-tune them, in our study we focus on an archetypal CNN architecture to derive our insights.

In this work, we have thoroughly compared the medical segmentation performance for various target segmentation tasks on brain MR imaging using transfer learning with various source domains and tasks used for pre-training. Our results suggest that medical segmentation tasks benefit from transfer learning with pre-training on segmentation source tasks on the same domain.

## 2.7 Conclusion

Our transfer learning experiments targeting brain MRI segmentation tasks suggest that selecting a similar (segmentation) source task and domain results in equal or better spatial agreement than other choices of source task and domain combinations. Even with a source dataset 10 times as large, pre-training on ImageNet classification did not outperform the equal source and target task and domain combination in two out of three target tasks: stroke lesion and brain anatomy segmentation. However, source task and domain selection have an inconsistent effect on the lesion detection accuracy.



## 2.8 Acknowledgements

This publication is based on the STAIRS project under the TKI-PPP program. The collaboration project is co-funded by the PPP Allowance made available by Health-Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships. Dr. H.A. Marquering is a cofounder and shareholder of Nico.Lab. This work has been supported by the donation of a Quadro P6000 GPU by NVIDIA.

# 2.9 Supplementary Material

## 2.9.1 Participant and Scan Characteristics



| Study | N | Age Mean(SD) | Sex Male/Female | Scanner (Field Strength) | Scan Sequence | Voxel Dimensions |
|---|---|---|---|---|---|---|
| **OASIS (Open Access Series of imaging studies)** | 288 | 44.1 (23) | 106/188 | Siemens Vision (1.5T) | MPRAGE | $1.0 \times 1.0 \times 1.25$ mm |
| **IXI (Information eXtraction from Images)** | 561 | 48.6 (16.5) | 250/311 | Philips Intera(3T); Phillips Gyeroscan Intera (1.5T;GE Signa (1.5T); | T1-FFE; MPRAGE | $0.94 \times 0.94 \times 1.2$ mm |
| **ICBM (International Consortium for Brain Mapping)** | 322 | 24.8 (5.1) | 177/145 | Siemens Magnetom (1.5T) | MPRAGE | $1.0 \times 1.0 \times 1.0$ mm |
| **ABIDE (Autism Brain Imaging Data Exchange)** | 184 | 16.9 (6.7) | 161/23 | Various (all 3T) | MPRAGE | Various |
| **Beijing Normal University** | 179 | 21.3 (1.9) | 72/107 | Siemens(3T) | MPRAGE | $1.33 \times 1.0 \times 1.0$mm |
| **Nathan Kline Institute (NKI)/ Rockland** | 160 | 41.5 (18.1) | 96/64 | Siemens Tim Trio (3T) | MPRAGE | $1.0 \times 1.0 \times 1.0$mm |
| **MCIC (Mind Clinical Imaging Consortium)** | 93 | 32.5 (12) | 64/29 | Siemens Sonata/Trio (1.5/3T); GE Signa (1.5T) | MPRAGE; SPGR | $0.63 \times 0.63 \times 1.5$mm |
| **Berlin School of Brain & Mind** | 49 | 31 (7.1) | 24/25 | Siemens Tim Trio (3T) | MPRAGE | $1.0 \times 1.0 \times 1.0$mm |
| **NEO2012** | 39 | 29.6 (8.4) | 18/21 | Siemens Allegra(3T) | MPRAGE | $1.0 \times 1.0 \times 1.0$mm |
| **TRAIN-39** | 36 | 22.7 (2.6) | 11/25 | Siemens Allegra (3T) | MPRAGE | $1.33 \times 1.33 \times 1.3$mm |
| **Cleveland Clinic** | 31 | 43.6 (11.1) | 11/20 | Siemens Tim Trio (3T) | MPRAGE | $1.0 \times 1.0 \times 1.2$mm |
| **WUSL** | 24 | 23 (1.4) | 4/20 | Siemens Tim Trio (3T) | MPRAGE | $1.0 \times 1.0 \times 1.0$mm |
| **MIRIAD (Minimal Interval Resonance Imaging in Alzheimer's Disease)** | 23 | 69.7 (7.2) | 12/11 | GE Signa (1.5T) | 3D IR-FSPGR | $0.94 \times 0.94 \times 1.5$mm |
| **CADDementia** | 12 | 62.3 (6.3) | 9/3 | GE Signa (3T) | 3D IR-FSPGR | $0.9 \times 0.9 \times 1.0$mm |
| **Dataset Total** | 2001 | 37 (18.1) | 1016/985 | - | - | - |

**Table 2.3:** Additional scan and participant information of the Brain Age Healthy Cohort dataset [82].

| | | | | Vascular Territory | | | |
|---|---|---|---|---|---|---|---|
| **Scanner Brand** | **Field Strength** | **Resolution** | **N** | *MCA* | *ACA* | *Lacunar* | Other |
| GE 750 Discovery | 3T | 1.0 x 1.0 x 1.0 mm | 76 | 12 | 0 | 46 | 18 |
| GE Signa | 1.5 T | 0.9 x 0.9 x 0.9 mm | 31 | 14 | 2 | 11 | 4 |
| GE Signa Excite | 3T | 1.0 x 1.0 x 1.0 mm | 34 | 10 | 0 | 18 | 6 |
| GE Signa HD-X | 3T | 1.0 x 1.0 x 1.0 mm | 26 | 1 | 0 | 24 | 1 |
| Phillips Achieva | 3T | 1.0 x 1.0 x 1.0 mm | 36 | 15 | 0 | 18 | 3 |
| Siemens Trio | 3T | 1.0 x 1.0 x 1.0 mm | 101 | 56 | 1 | 32 | 12 |

**Table 2.4:** Additional information about the patients scans of the stroke lesion Data-set (ATLAS R1.2).

### 2.9.2 Equal Domain Source Task Hyper-parameters

The CNNs were pre-trained for 30 epochs using a batch size of 32 and batch normalization [95] before the activation function. The encoder and decoder used a learning rate of $10^{-4}$, a weight decay rate of $2 \cdot 10^{-6}$ and an Adam optimizer [99]. If a discriminator was used, it was turned on after 26000 iterations. The discriminator used a learning rate of $10^{-5}$ and an SGD optimizer. The encoder and a fully connected block used the same hyper-parameters as the encoder and decoder.

To ensure that the CNNs pre-trained equal and unequal source domain used a similar amount of data, 870 scans from the BAHC dataset were randomly selected. These scans were randomly split into 670 training scans, 100 validation scans and 100 testing scans.

### 2.9.3 Unequal domain Source Task Hyper-parameters

The CNNs were pre-trained as described in the Taskonomy study [135]. The CNNs were pre-trained for 30 epochs using a batch size of 32 and batch normalization [95] before the activation function. The encoder and decoder used a learning rate of $10^{-4}$, a weight decay rate of $2 \cdot 10^{-6}$ and an Adam optimizer [99]. If a discriminator was used, it was turned on after 25000 iterations. The discriminator used a learning rate of $10^{-5}$ and an SGD optimizer. The learning rate was annealed by a factor of 10 after 80000 iterations.

The CNNs were trained on 120000, validated on 16000 and tested on 17000 images.

# Bibliography

[75] A. Achille, M. Lam, R. Tewari, A. Ravichandran, S. Maji, C. C. Fowlkes, S. Soatto, and P. Perona. Task2vec: Task embedding for meta-learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6430–6439, 2019.

[76] M. A. Akhloufi and M. Chetoui. Chest XR COVID-19 detection. `https://cxr-covid19.grand-challenge.org/`, August 2021. Online; accessed September 2021.

[77] J. Amin, M. Sharif, M. Yasmin, T. Saba, M. A. Anjum, and S. L. Fernandes. A new approach for brain tumor segmentation and classification based on score level fusion using transfer learning. *Journal of medical systems*, 43(11):1–16, 2019.

[78] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

[79] Q. Chen, Z. Zheng, C. Hu, D. Wang, and F. Liu. On-edge multi-task transfer learning: Model and practice with data-driven task allocation. *IEEE Transactions on Parallel and Distributed Systems*, 31(6):1357–1371, 2019.

[80] S. Chen, K. Ma, and Y. Zheng. Med3d: Transfer learning for 3d medical image analysis. *arXiv preprint arXiv:1904.00625*, 2019.

[81] M. Christopher, A. Belghith, C. Bowd, J. A. Proudfoot, M. H. Goldbaum, R. N. Weinreb, C. A. Girkin, J. M. Liebmann, and L. M. Zangwill. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. *Scientific reports*, 8(1):1–13, 2018.

[82] J. H. Cole, R. P. Poudel, D. Tsagkrasoulis, M. W. Caan, C. Steves, T. D. Spector, and G. Montana. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*, 163:115–124, 2017.

[83] K. Dwivedi and G. Roig. Representation similarity analysis for efficient task taxonomy & transfer learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12387–12396, 2019.

[84] C. Finn, P. Abbeel, and S. Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pages 1126–1135. PMLR, 2017.

[85] V. S. Fonov, A. C. Evans, R. C. McKinstry, C. Almli, and D. Collins. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, (47):S102, 2009.

[86] Y. Ganin and V. Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pages 1180–1189. PMLR, 2015.

[87] M. Ghafoorian, A. Mehrtash, T. Kapur, N. Karssemeijer, E. Marchiori, M. Pesteie, C. R. Guttmann, F.-E. de Leeuw, C. M. Tempany, B. Van Ginneken, et al. Transfer learning for domain adaptation in mri: Application in brain lesion segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 516–524. Springer, 2017.

[88] K. Gorgolewski, C. D. Burns, C. Madison, D. Clark, Y. O. Halchenko, M. L. Waskom, and S. S. Ghosh. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front Neuroinform*, 5, 08 2011.

[89] H. Guan and M. Liu. Domain adaptation for medical image analysis: a survey. *arXiv preprint arXiv:2102.09508*, 2021.

[90] Y. Guo, H. Shi, A. Kumar, K. Grauman, T. Rosing, and R. Feris. Spottune: transfer learning through adaptive fine-tuning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4805–4814, 2019.

[91] T. Han, V. X. Nunes, L. F. D. F. Souza, A. G. Marques, I. C. L. Silva, M. A. A. F. Junior, J. Sun, and P. P. Reboucas Filho. Internet of medical

thingsbased on deep learning techniques for segmentation of lung and stroke regions in ct scans. *IEEE Access*, 8:71117–71135, 2020.

[92] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.

[93] M. H. Hesamian, W. Jia, X. He, and P. Kennedy. Deep learning techniques for medical image segmentation: achievements and challenges. *Journal of digital imaging*, 32(4):582–596, 2019.

[94] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[95] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[96] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[97] K. Kamnitsas, C. Baumgartner, C. Ledig, V. Newcombe, J. Simpson, A. Kane, D. Menon, A. Nori, A. Criminisi, D. Rueckert, et al. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In *International conference on information processing in medical imaging*, pages 597–609. Springer, 2017.

[98] J. Kang and J. Gwak. Ensemble of instance segmentation models for polyp segmentation in colonoscopy images. *IEEE Access*, 7:26440–26447, 2019.

[99] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[100] G. Koch, R. Zemel, R. Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2. Lille, 2015.

[101] S. Kornblith, J. Shlens, and Q. V. Le. Do better imagenet models transfer better? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2661–2671, 2019.

[102] Ž. Lesjak, A. Galimzianova, A. Koren, M. Lukin, F. Pernuš, B. Likar, and Ž. Špiclin. A novel public mr image dataset of multiple sclerosis patients with lesion segmentations based on multi-rater consensus. *Neuroinformatics*, 16(1):51–63, Jan 2018.

[103] Y. Li, H. Qi, J. Dai, X. Ji, and Y. Wei. Fully convolutional instance-aware semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2359–2367, 2017.

[104] S.-L. Liew, J. M. Anglin, N. W. Banks, M. Sondag, K. L. Ito, H. Kim, J. Chan, J. Ito, C. Jung, S. Lefebvre, et al. The anatomical tracings of lesions after stroke (atlas) dataset-release 1.1. *bioRxiv*, page 179614, 2017.

[105] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[106] M. Long, Y. Cao, J. Wang, and M. Jordan. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, pages 97–105. PMLR, 2015.

[107] A. L. Maas, A. Y. Hannun, and A. Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3, 2013.

[108] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner. Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*, 19(9):1498–1507, 2007.

[109] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014.

[110] P. Moeskops, J. M. Wolterink, B. H. van der Velden, K. G. Gilhuijs, T. Leiner, M. A. Viergever, and I. Išgum. Deep learning for multi-task medical image segmentation in multiple modalities. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 478–486. Springer, 2016.



[111] M. A. Morid, A. Borjali, and G. Del Fiol. A scoping review of transfer learning research on medical image analysis using imagenet. *Computers in biology and medicine*, page 104115, 2020.

[112] B. Mustafa, A. Loh, J. Freyberg, P. MacWilliams, A. Karthikesalingam, N. Houlsby, and V. Natarajan. Supervised transfer learning at scale for medical imaging. *arXiv preprint arXiv:2101.05913*, 2021.

[113] C. Ouyang, C. Biffi, C. Chen, T. Kart, H. Qiu, and D. Rueckert. Self-supervision with superpixels: Training few-shot medical image segmentation without annotation. In *European Conference on Computer Vision*, pages 762–780. Springer, 2020.

[114] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[115] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2010.

[116] B. Patenaude, S. M. Smith, D. N. Kennedy, and M. Jenkinson. A bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage*, 56(3):907–922, 2011.

[117] L. Y. Pratt et al. Discriminability-based transfer between neural networks. *Advances in neural information processing systems*, pages 204–204, 1993.

[118] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio. Transfusion: Understanding transfer learning with applications to medical imaging. *arXiv preprint arXiv:1902.07208*, 2019.

[119] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on*

*Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[120] A. G. Roy, S. Siddiqui, S. Pölsterl, N. Navab, and C. Wachinger. squeeze & exciteguided few-shot segmentation of volumetric images. *Medical image analysis*, 59:101587, 2020.

[121] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

[122] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3723–3732, 2018.

[123] T. Schlegl, J. Ofner, and G. Langs. Unsupervised pre-training across image domains improves lung tissue classification. In *International MICCAI Workshop on Medical Computer Vision*, pages 82–93. Springer, 2014.

[124] S. M. Smith. Fast robust automated brain extraction. *Human brain mapping*, 17(3):143–155, 2002.

[125] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, et al. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, 23:S208–S219, 2004.

[126] T. Standley, A. Zamir, D. Chen, L. Guibas, J. Malik, and S. Savarese. Which tasks should be learned together in multi-task learning? In *International Conference on Machine Learning*, pages 9120–9132. PMLR, 2020.

[127] J.-C. Su, S. Maji, and B. Hariharan. When does self-supervision improve few-shot learning? In *European Conference on Computer Vision*, pages 645–666. Springer, 2020.

[128] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–7176, 2017.

[129] S. Wang, L. Zhang, W. Zuo, and B. Zhang. Class-specific reconstruction transfer learning for visual recognition across domains. *IEEE Transactions on Image Processing*, 29:2424–2438, 2019.

[130] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys (CSUR)*, 53(3):1–34, 2020.

[131] Y. Xie and D. Richmond. Pre-training on grayscale imagenet improves medical image classification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018.

[132] T. S. Yoo, M. J. Ackerman, W. E. Lorensen, W. Schroeder, V. Chalana, S. Aylward, D. Metaxas, and R. Whitaker. Engineering and algorithm design for an image processing api: a technical report on itk-the insight toolkit. *Studies in health technology and informatics*, pages 586–592, 2002.

[133] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural networks? In *Advances in neural information processing systems*, pages 3320–3328, 2014.

[134] K. You, Z. Kou, M. Long, and J. Wang. Co-tuning for transfer learning. *Advances in Neural Information Processing Systems*, 33, 2020.

[135] A. R. Zamir, A. Sax, W. Shen, L. J. Guibas, J. Malik, and S. Savarese. Taskonomy: Disentangling task transfer learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3712–3722, 2018.

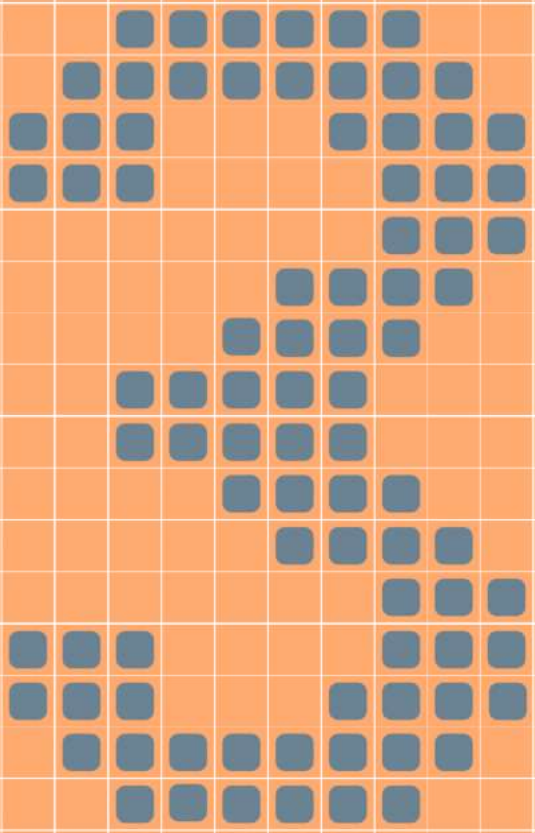
[136] G. Zeng and G. Zheng. Multi-stream 3d fcn with multi-scale deep supervision for multi-modality isointense infant brain mr image segmentation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 136–140. IEEE, 2018.

[137] L. Zhang, J. Liu, B. Zhang, D. Zhang, and C. Zhu. Deep cascade model-based face recognition: When deep-layered learning meets small data. *IEEE Transactions on Image Processing*, 29:1016–1029, 2019.

[138] L. Zhang, S. Wang, G.-B. Huang, W. Zuo, J. Yang, and D. Zhang. Manifold criterion guided transfer learning via intermediate domain generation. *IEEE transactions on neural networks and learning systems*, 30(12):3759–3773, 2019.

[139] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging*, 20(1):45–57, 2001.

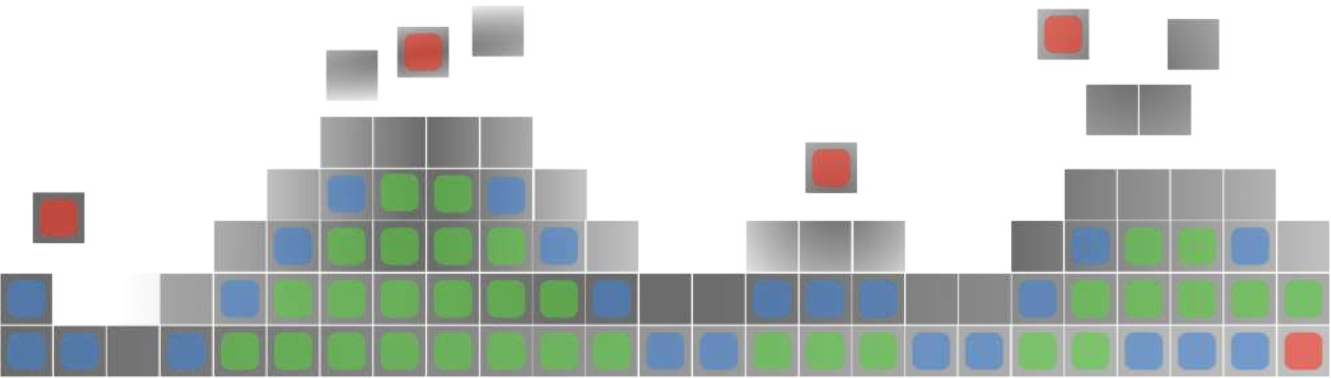
[140] Z. Zhou, V. Sodha, M. M. R. Siddiquee, R. Feng, N. Tajbakhsh, M. B. Gotway, and J. Liang. Models genesis: Generic autodidactic models for 3d medical image analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 384–393. Springer, 2019.



# Chapter 3

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

**Based on:** R. Zoetmulder, P.R. Konduri, I.V. Obdeijn, E. Gavves, I. Išgum, C.B.L.M. Majoie, D.W.J Dippel, Y.B.W.E.M. Roos, M. Goyal, P.J. Mitchell, B.C.V. Campbell, D.K. Lopes, G. Reimann, T.G. Jovin, J.L. Saver, K.W. Muir, P. White, S. Bracard, B. Chen, S. Brown, W.J. Schonewille, E. van der Hoeven, V.Puetz and H.A. Marquering (2022). "Automated Final Lesion Segmentation in Posterior Circulation Acute Ischemic Stroke using Deep Learning". Diagnostics, 11(9), 1621.



# 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. Applicability of deep learning approaches to PCS is limited due to its lower incidence compared to ACS. We evaluated strategies to develop a CNN for PCS lesion segmentation by using image data from both ACS and PCS patients. We included follow-up NCCT 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 performance of only including PCS patients, a CNN was trained on PCS data. Third, to evaluate performance when combining the datasets, a CNN was trained on both datasets. Finally, to evaluate 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 ICC of 0.88 (95%CI: 0.83-0.92) versus 0.55 to 0.83 and a lesion detection rate of 87% versus 41-77 for the other strategies. Hence, transfer learning improved the FLV quantification and detection rate of PCS lesions compared to the other strategies.

## 3.1 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 Ranking Scale (mRS) after 90 days) [144]. Various trials have shown a strong association of infarct volume with functional outcome in patients suffering from a stroke due to a large vessel occlusion in the anterior circulation [144; 172].



However, in patients with a stroke due to a Posterior Circulation Stroke (PCS), the relation between infarct volume and outcome is understudied [167]. 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 [163] could not show a beneficial effect of endovascular treatment with functional outcome used as outcome measure. Alternatively, secondary outcome measures such as infarct volume 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 [141]. Developing methods that automatically segment lesions due to a PCS, would help investigate Final Lesion Volume (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 [143; 151]. However, these studies have only considered FLV of patients with an AIS due to an occlusion of the anterior circulation [143; 144]. To achieve good performance, CNNs typically require large amounts of labeled training data. However, PCS constitutes only 26 % of AIS cases [145; 156] and training of CNNs for automatic PCS lesion segmentation is thus hindered by 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 [158]. 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 is available. Transfer learning has been successfully applied to solve various medical image analysis problems [153].

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.



## 3.2 Materials & Methods

### 3.2.1 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 [149], 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 inclusion criteria are shown in Figure 3.4a, in the appendix. Patients were excluded if no Follow-Up Non-Contrast Computed Tomography (FU-NCCT) was made in the time window of 12 hours 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 3.1.

The BASICS dataset consists of 168 patients who suffered from a PCS and was obtained from the BASICS trial [152; 166], which investigated the effectiveness of endovascular therapy for treating patients with a PCS. This trial

| 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 | | |
| IV Thrombolysis, No. [%] | 92/107 [86] | 872/1018 [85.7] |
| Time | | |
| Stroke onset to IVT, minutes, Mean [Standard Deviation](N) | 176.9 [176.102] (90) | 112.2 [57.2] (871) |

**Table 3.1:** Baseline characteristics, treatment and time data for patients with posterior circulation stroke and anterior circulation stroke. Prior Posterior stroke, Transient Ischemic Attack, posterior circulation TIA and Atrial Fibrillation (history or 12 lead ECG) were not available (NAV) for the HERMES dataset.



included patients from 23 centers, collected between 2011 and 2019. Inclusion criteria for our study are shown in Figure 3.4b, in the appendix. 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. Baseline characteristics of the included BASICS patients are shown in Table 3.1. The infarcted regions are shown in Table 3.2, in the appendix.

### 3.2.2 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 [144]. In summary, a window width of 30 Hounsfield Units (HU) and a center level of 35 HU was set in ITK-Snap [171]. All hypodense 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 segmented by a single trained observer (IVO) and were checked by an experienced radiologist (CBLMM) who had more than 5 years of experience. Lesions were segmented by using the aforementioned window width and center level on ITK-Snap software [171]. Posterior Circulation Alberta Stroke Program Early Computed Tomography Score (PC-ASPECTS) [159] were used when available to identify the infarcted territory.



### 3.2.3 Preprocessing

The intracranial region as a volume of interest was obtained automatically using a combination of preprocessing steps [143]. 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 [155] 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 x 256 x 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 [154; 168] and Python 2.7.

### 3.2.4 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 3.5a in the appendix), was inspired by U-Net [161] and ResNet [150]. 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, followed by a max pooling layer with a pooling size of two and a stride of two. The features generated by the input block consisted of eight channels.

Subsequent to the input block, three downsampling blocks were added, consisting of three 3D ResNet layers, shown in Figure 3.5b, in the appendix. 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, followed by a transposed convolution with a stride of two and a kernel size of three. Each upsampling block took the features 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.

### 3.2.5 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 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 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, $5 \cdot 10^{-4}$, $2 \cdot 10^{-4}$, $10^{-4}$ , 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, $5 \cdot 10^{-6}$, $2 \cdot 10^{-6}$, 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 five-fold cross-validation. Given the number of available PCS patients, the first four testing splits consisted of 20 % and the fifth of 22 % of the data. The training splits were of equal size and consisted of 78 % of the PCS data.

### 3.2.6 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 [147]. Following their guidelines, an ICC $< 0.4$ is defined as poor, an ICC between 0.4 and 0.6 as fair, an ICC between 0.6 and 0.75 as good and an ICC greater than 0.75 as excellent. In addition, a Bland-Altman analysis was performed to assess the bias and spread in volume measurements. Statistical significance between ICCs was evaluated by using Fisher's r-to-z transformation.

To determine whether our model accurately detected lesions independent of size, the number of detected lesions was determined. For this measure,

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 threshold. 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 percentage of overlapping voxels was set to either greater than zero percent or 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. 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 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, version 0.3.1 [165].

## 3.3 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 ($p<0.01$). The bias and spread of the volume measurements are shown in Figure 3.1 and Table 3.3.

The lesion detection rate of the TL-CNN was higher than for the other learning strategies (Figure 3.2). Figure 3.2 shows that the lesion detection rate increases with increasing lesion volume and that the lesion detection rate decreases with increasing thresholds of overlapping voxels.

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 3.3). 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 3.4. For the anterior circulation stroke lesions, the ACS-CNN achieved an average Dice coefficient of 0.60±0.07.



## 3.4 Discussion

Our study found that transfer learning results in a high level of agreement between manually delineated and automatically quantified lesion volumes on 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 one of the few that addressed automated segmentation of posterior circulation stroke lesions on follow-up NCCT. Previous work include anterior stroke lesion segmentation in a variety of imaging modalities, such as baseline CTP [148] baseline CTA [157], follow-up NCCT [143] and baseline and follow-up DWI [170] 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 [143] 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

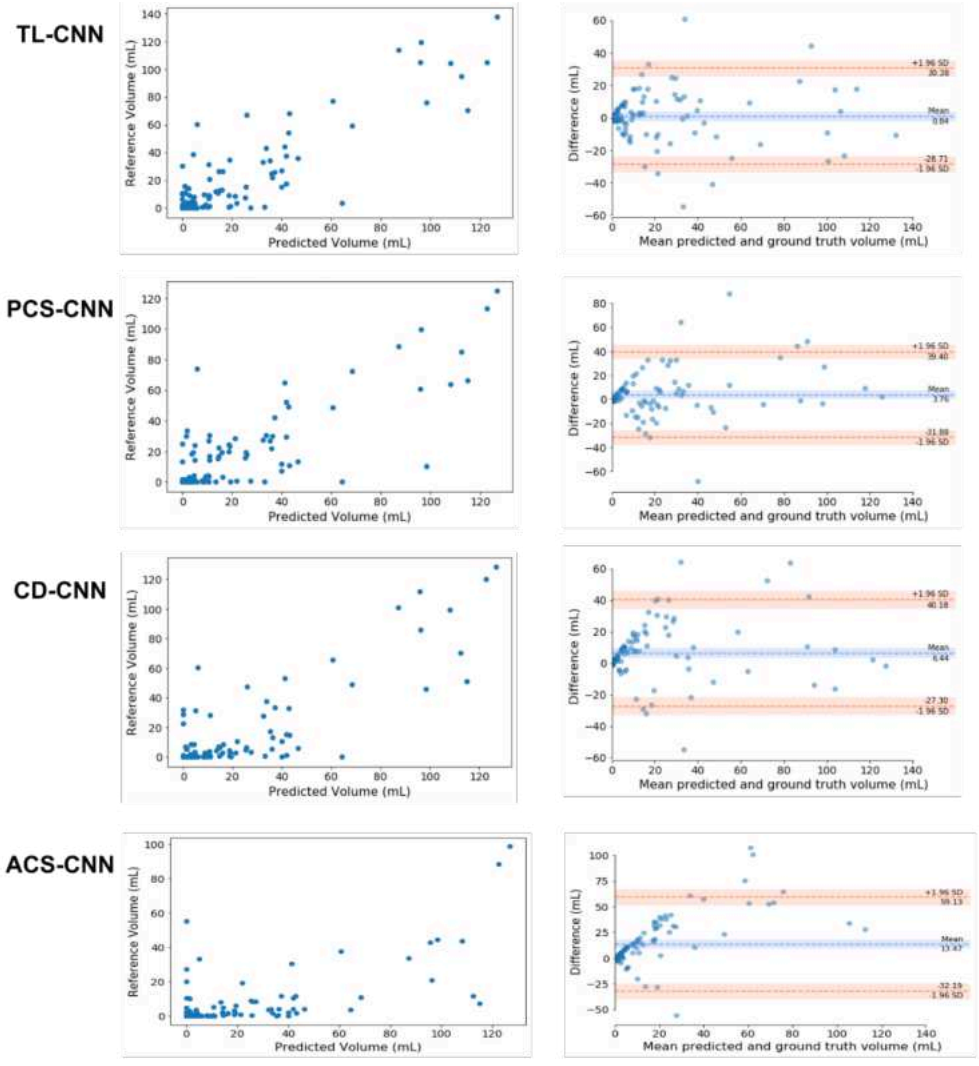



**Figure 3.1:** Comparison of the automated and reference segmentation volume for the Transfer Learned CNN (TL-CNN), Posterior Circulation Stroke CNN (PCS-CNN), Combined Datasets CNN (CD-CNN) and Anterior Circulation Stroke CNN (ACS-CNN). Left column: Scatter plots comparing lesion volumes derived from 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 y-axis.

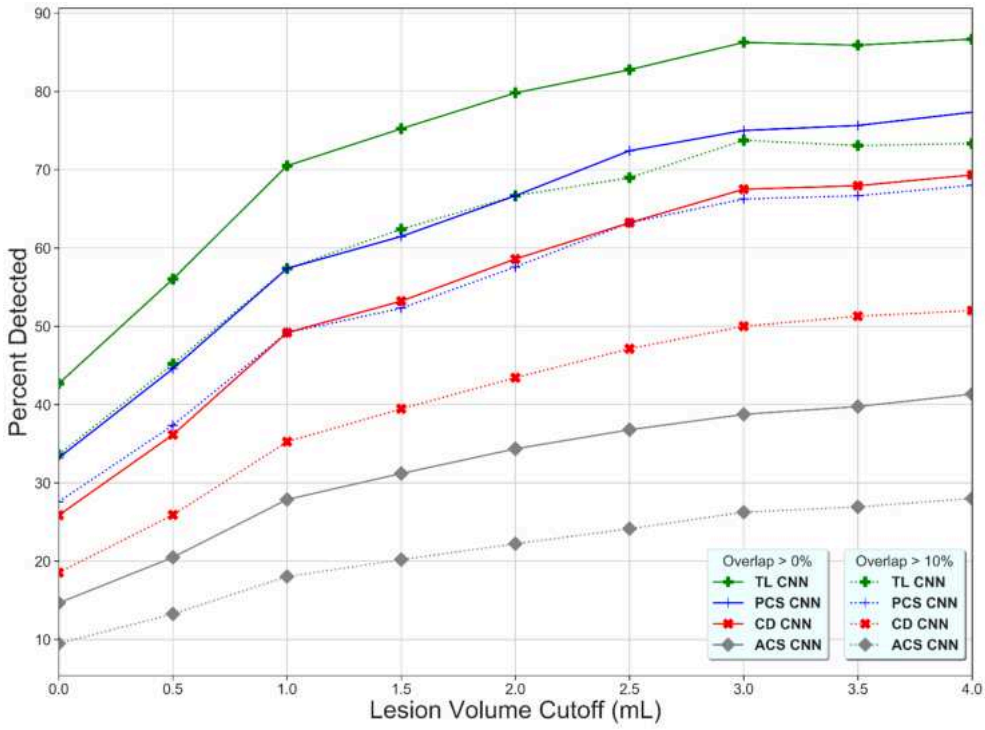



**Figure 3.2:** Percentage of detected lesions (y-axis) as a function of the minimum volume requirement (x-axis) and the minimum percentage of overlapping voxels for the Transfer Learned CNN (TL-CNN), Posterior Circulation Stroke CNN (PCS-CNN), Combined Datasets CNN (CD-CNN) and Anterior Circulation 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).

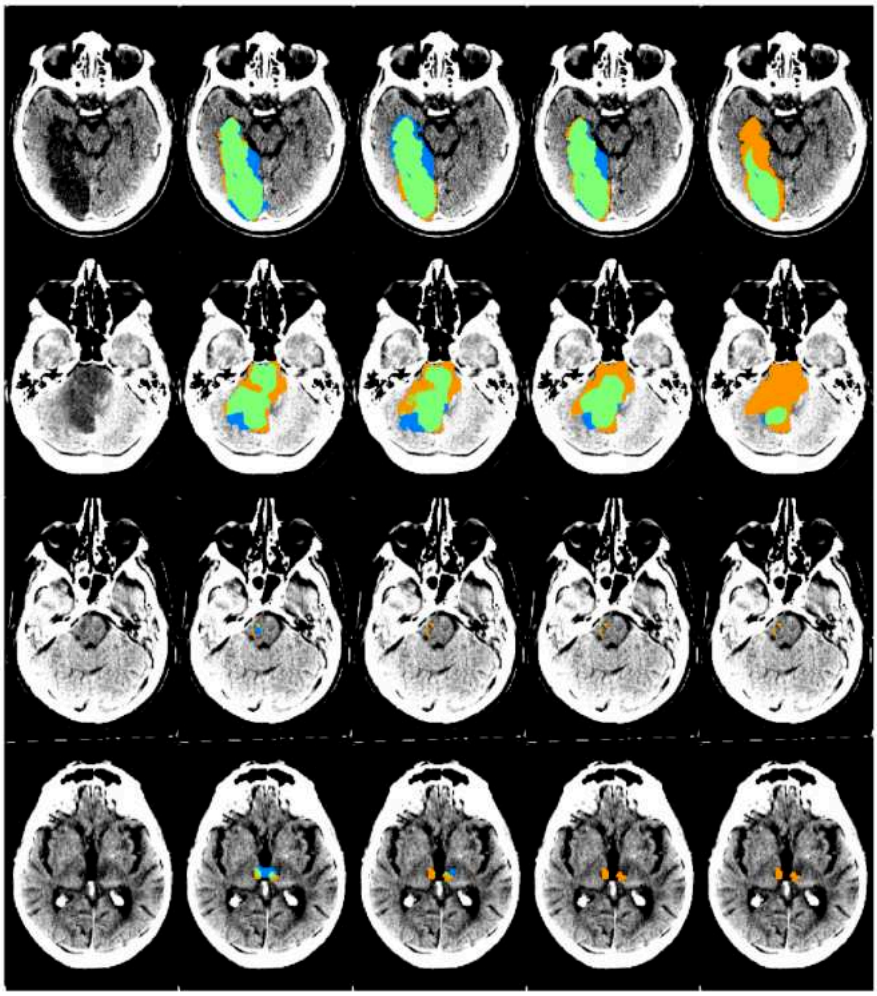


**Figure 3.3:** An example of automatic segmentation results obtained by the 4 CNNs on the PCS test set. From the left to the right column: the original scan, the automatic segmentation results from the TL-CNN, PCS-CNN, CD-CNN and ACS-CNN are shown. The segmentation maps show true positives (green), false positives (blue) and false negatives (orange). The scans were plotted using a window center around 35, with a window width of 30.

the larger dataset available for training. In addition, the Dice coefficient is a global overlap metric which works well for large lesions. If lesions are smaller the Dice coefficient may be too sensitive to small errors and detection metrics are more important.

Other methods for automatic ACS lesion segmentation used information from the contralateral hemisphere to improve segmentation accuracy [142; 157]. 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 [160]. 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 [164]. 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 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 [146]. 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 [173]. 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 [162] as a benchmark. These studies used transfer learning to improve 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 [173].

This study has several limitations. First, most of the FU-NCCT scans of PCS patients included in this study were mostly obtained 24 hours after 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 hours. In addition, FLV segmentations for patients with PCS were performed by only one trained observer, hence inter-observer agreement could not be assessed. However, the reference segmentations were verified by an experienced radiologist. 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 was 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 2mL. 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 [169]. 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.

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 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 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 towards clinical application.

## 3.5 Supplementary materials

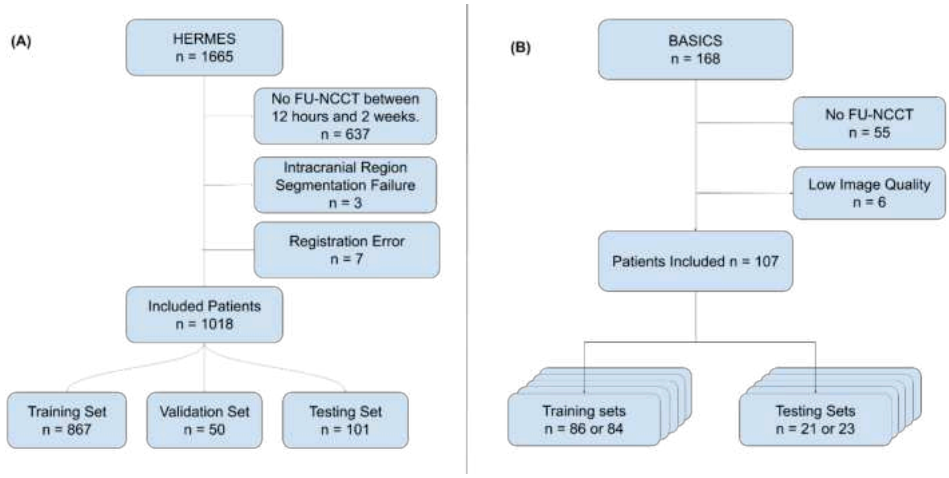



**Figure 3.4:** (A) Flowchart showing the exclusion criteria used for the HERMES dataset. Exclusion criteria were: No FU-NCCT available which was acquired between 12 hours and 2 weeks (n = 637), an intracranial region segmentation error (n = 3) and a registration error (n = 7). In total, 1018 patients were included. These were split into a training set ( n = 876 ), a val-idation 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 5 training and testing sets for the five-fold cross-validation. Patients could only belong to one test set.

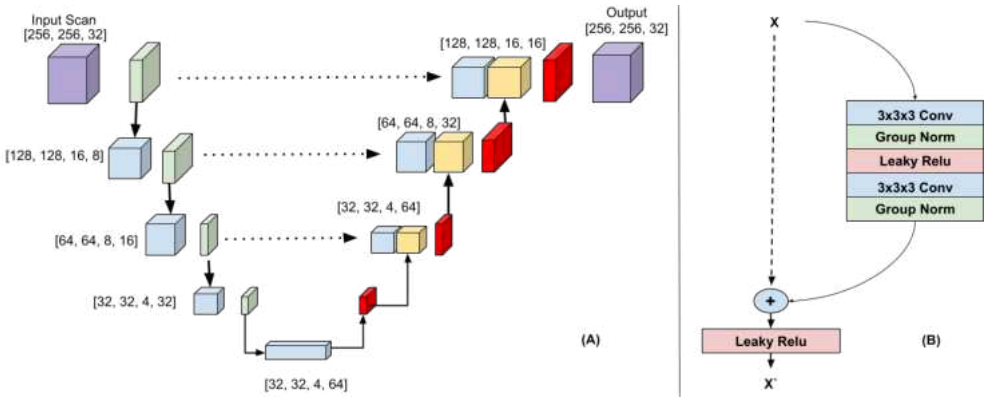



**Figure 3.5:** A) 3D-Unet Architecture. The downsampling path (left) consisted of 3DResNet blocks with max pooling (green). The upsampling path (right) consisted of ResNet blocks followed by transposed convolutions (red). The features created in the downsampling path are colored blue and the features 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.

| **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 3.2:** Lesion location in the posterior fossa, scored manually by using the PC-ASPECTS.

| 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-CN** | 0.55 (95% CI: 0.4-0.67) | 0.07±0.03 | 13.5 mL | -32.2 to 59.1 mL |

**Table 3.3:** 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.



| | | Dice Coefficient | | Bias | |
|---|---|---|---|---|---|
| **Method 1** | **Method 2** | **W** | **P-Value** | **W** | **P-Value** |
| **TL-CNN** | **PCS-CNN** | 766 | <.05 | 2205 | .28 |
| **TL-CNN** | **CD-CNN** | 216 | <.01 | 1018 | <.01 |
| **TL-CNN** | **ACS-CNN** | 62 | <.01 | 938 | <.01 |
| **PCS-CNN** | **CD-CNN** | 535 | <.01 | 1958 | <.05 |
| **PCS-CNN** | **ACS-CNN** | 114 | <.01 | 1350 | <.01 |
| **CD-CNN** | **ACS-CNN** | 88 | <.01 | 1443 | <.01 |

**Table 3.4:** 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.

# Bibliography

[141] J. L. Banks and C. A. Marotta. Outcomes validity and reliability of the modified rankin scale: Implications for stroke clinical trials - A literature review and synthesis, 3 2007.

[142] A. Barman, M. E. Inam, S. Lee, S. Savitz, S. Sheth, and L. Giancardo. Determining ischemic stroke from ct-angiography imaging using symmetry-sensitive convolutional networks. In *Proceedings - International Symposium on Biomedical Imaging*, volume 2019-April, pages 1873–1877. IEEE Computer Society, 4 2019.

[143] R. S. Barros, M. L. Tolhuisen, A. M. Boers, I. Jansen, E. Ponomareva, D. W. Dippel, A. Van Der Lugt, R. J. Van Oostenbrugge, W. H. Van Zwam, O. A. Berkhemer, M. Goyal, A. M. Demchuk, B. K. Menon, P. Mitchell, M. D. Hill, T. G. Jovin, A. Davalos, B. C. Campbell, J. L. Saver, Y. B. Roos, K. W. Muir, P. White, S. Bracard, F. Guillemin, S. D. Olabarriaga, C. B. Majoie, and H. A. Marquering. Automatic segmentation of cerebral infarcts in follow-up computed tomography images with convolutional neural networks. *Journal of NeuroInterventional Surgery*, 12(9):848–852, 9 2020.

[144] A. M. Boers, I. Jansen, l. F. M Beenen, T. Devlin, L. san roman, J. hoe heo, M. Ribó, S. Brown, M. Almekhlafi, D. Liebeskind, J. Teitelbaum, h. F. Lingsma, W. van Zwam, P. Cuadras, M. Beaumont, M. M. Brown, a. J. Yoo, r. J. van Oostenbrugge, B. K. Menon, G. Donnan, J. louis Mas, Y. B. W M roos, C. Oppenheim, A. van der lugt, r. J. Dowling, M. D. Hill, A. Davalos, T. Moulin, N. Agrinier, a. M. Demchuk, D. K. Lopes, L. aja rodríguez, D. W. J Dippel, B. V. Campbell, P. J. Mitchell, F. Al-ajlan, T. Jovin, J. Madigan, g. W. Albers, S. Soize, F. Guillemin, V. K. Reddy, S. Bracard, J. Blasco, K. W. Muir, R. Nogueira, P. M. White, M. Goyal, s. M. Davis, H. Marquering, and c. B. M Majoie. Association of follow-up infarct volume with functional outcome in acute ischemic stroke: a pooled analysis of seven randomized trials. *J NeuroIntervent Surg*, 10:1137–1142, 2018.

[145] J. Bogousslavsky, G. V. Melle, and F. Regli. The lausanne stroke registry: Analysis of 1,000 consecutive patients with first stroke. *Stroke*, 19(9):1083–1092, 1988.

[146] S. Chen, K. Ma, and Y. Zheng. MED3D: Transfer learning for 3D medical image analysis, 2019.

[147] D. V. Cicchetti. Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychological Assessment*, 6(4):284–290, 1994.

[148] A. Clèrigues, S. Valverde, J. Bernal, J. Freixenet, A. Oliver, and X. Lladó. Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks. *Computers in Biology and Medicine*, 115, 12 2019.

[149] M. Goyal, B. K. Menon, W. H. Van Zwam, D. W. Dippel, P. J. Mitchell, A. M. Demchuk, A. Dávalos, C. B. Majoie, A. Van Der Lugt, M. A. De Miquel, G. A. Donnan, Y. B. Roos, A. Bonafe, R. Jahan, H. C. Diener, L. A. Van Den Berg, E. I. Levy, O. A. Berkhemer, V. M. Pereira, J. Rempel, M. Millán, S. M. Davis, D. Roy, J. Thornton, L. S. Román, M. Ribó, D. Beumer, B. Stouch, S. Brown, B. C. Campbell, R. J. Van Oostenbrugge, J. L. Saver, M. D. Hill, and T. G. Jovin. Endovascular thrombectomy after large-vessel ischaemic stroke: A meta-analysis of individual patient data from five randomised trials. *The Lancet*, 387(10029):1723–1731, 4 2016.

[150] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2016-Decem, pages 770–778, 12 2016.

[151] Y. C. Kim, J. E. Lee, I. Yu, H. N. Song, I. Y. Baek, J. K. Seong, H. G. Jeong, B. J. Kim, H. S. Nam, J. W. Chung, O. Y. Bang, G. M. Kim, and W. K. Seo. Evaluation of Diffusion Lesion Volume Measurements in Acute Ischemic Stroke Using Encoder-Decoder Convolutional Network. *Stroke*, 50(6):1444–1451, 2019.

[152] L. C. Langezaal, E. J. van der Hoeven, F. J. MontAlverne, J. J. de Carvalho, F. O. Lima, D. W. Dippel, A. van der Lugt, R. T. Lo, J. Boiten, G. J.

Lycklama à Nijeholt, J. Staals, W. H. van Zwam, P. J. Nederkoorn, C. B. Majoie, J. C. Gerber, M. Mazighi, M. Piotin, A. Zini, S. Vallone, J. Hofmeijer, S. O. Martins, C. H. Nolte, K. Szabo, F. A. Dias, D. G. Abud, M. J. Wermer, M. J. Remmers, H. Schneider, C. M. Rueckert, K. F. de Laat, A. J. Yoo, P.-J. van Doormaal, A. C. van Es, B. J. Emmer, P. Michel, V. Puetz, H. J. Audebert, O. M. Pontes-Neto, J.-A. Vos, L. J. Kappelle, A. Algra, and W. J. Schonewille. Endovascular Therapy for Stroke Due to Basilar-Artery Occlusion. *New England Journal of Medicine*, 384(20):1910–1920, 5 2021.



[153] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez. A survey on deep learning in medical image analysis, 12 2017.

[154] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek. The design of simpleITK. *Frontiers in Neuroinformatics*, 7(DEC), 12 2013.

[155] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellyn, and W. Eubank. Nonrigid multimodality image registration. In *Medical Imaging 2001: Image Processing*, volume 4322, pages 1609–1620. SPIE, 7 2001.

[156] T. Moulin, L. Tatu, F. Vuillier, E. Berger, D. Chavot, and L. Rumbach. Role of a stroke data bank in evaluating cerebral infarction subtypes: Patterns and outcome of 1,776 consecutive patients from the Besancon Stroke Registry. *Cerebrovascular Diseases*, 10(4):261–271, 2000.

[157] O. Öman, T. Mäkelä, E. Salli, S. Savolainen, and M. Kangasniemi. 3D convolutional neural networks applied to CT angiography in the detection of acute ischemic stroke. *European radiology experimental*, 3(1):8, 2019.

[158] S. J. Pan and Q. Yang. A survey on transfer learning, 2010.

[159] V. Puetz, P. N. Sylaja, S. B. Coutts, M. D. Hill, I. Dzialowski, P. Mueller, U. Becker, G. Urban, C. O'Reilly, P. A. Barber, P. Sharma, M. Goyal, G. Gahn, R. Von Kummer, and A. M. Demchuk. Extent of hypoattenuation on CT angiography source images predicts functional outcome in patients with basilar artery occlusion. *Stroke*, 39(9):2485–2490, 9 2008.

[160] D. Pustina, H. Branch Coslett, P. E. Turkeltaub, N. Tustison, M. F. Schwartz, and B. Avants. Automated Segmentation of Chronic Stroke Lesions Using LINDA: Lesion Identification With Neighborhood Data Analysis. 2016.

[161] O. Ronneberger, P. Fischer, and T. Brox. 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)*, volume 9351, pages 234–241, Munich, Germany, 2015. Springer Verlag.

[162] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252, 9 2014.

[163] W. J. Schonewille, C. A. Wijman, P. Michel, C. M. Rueckert, C. Weimar, H. P. Mattle, S. T. Engelter, D. Tanne, K. W. Muir, C. A. Molina, V. Thijs, H. Audebert, T. Pfefferkorn, K. Szabo, P. J. Lindsberg, G. de Freitas, L. Jaap Kappelle, A. Algra, A. M. Weber, G. A. Donnan, V. Thijs, A. Peeters, G. de Freitas, A. B. Conforto, M. Miranda-Alves, A. Massaro, P. Ijäs, T. Bogoslovsky, P. J. Lindsberg, C. Weimar, J. Benemann, K. Kraywinkel, C. Haverkamp, D. Michalski, K. Weissenborn, M. Goertler, A. Kloth, A. Bitsch, T. Mieck, J. Machetanz, P. Möller, R. Huber, S. Kaendler, C. Rueckert, H. Audebert, R. Müller, B. Vatankhah, T. Pfefferkorn, T. E. Mayer, K. Szabo, C. Disque, O. Busse, C. Berger, W. Hacke, Y. Schwammenthal, D. Orion, D. Tanne, M. Bergui, E. Pozzati, W. J. Schonewille, A. Algra, L. J. Kappelle, G. J. Luijckx, P. Vroomen, M. D. Vergouwen, Y. Roos, J. Stam, P. Bienfait, F. E. de Leeuw, P. de Kort, D. Dippel, J. Pagola, M. Ribo, C. Molina, A. Gonzales, A. Gil-Peralta, B. Norrving, M. Arnold, U. Fischer, J. Gralla, H. Mattle, G. Schroth, P. Michel, S. T. Engelter, S. Wetzel, P. Lyrer, J. Gandjour, N. Michael, R. Baumgartner, B. Tettenborn, H. Hungerbuehler, T. Baird, K. Muir, C. A. Wijman, A. Finley Caulfield, M. Lansberg, N. Schwartz, C. Venkatasubramanian, Z. Garami, S. Bogaard, F. Yatzu, and J. Grotta. Treatment and outcomes of acute basilar artery occlusion in the Basilar Artery International Cooper-

ation Study (BASICS): a prospective registry study. *The Lancet Neurology*, 8(8):724–730, 8 2009.

[164] N. Tomita, S. Jiang, M. E. Maeder, and S. Hassanpour. Automatic post-stroke lesion segmentation on MR images using 3D residual convolutional neural network. *NeuroImage: Clinical*, 27, 1 2020.

[165] R. Vallat. Pingouin: statistics in Python. *Journal of Open Source Software*, 3(31):1026, 11 2018.

[166] E. J. Van Der Hoeven, W. J. Schonewille, J. A. Vos, A. Algra, H. J. Audebert, E. Berge, A. Ciccone, M. Mazighi, P. Michel, K. W. Muir, V. Obach, and V. Puetz. The Basilar Artery International Cooperation Study (BASICS): study protocol for a randomised controlled trial. Technical report, 2013.

[167] K. Villringer, M. Florczak-Rzepka, U. Grittner, P. Brunecker, H. Tepe, C. H. Nolte, and J. B. Fiebach. Characteristics associated with outcome in patients with first-ever posterior fossa stroke. *European Journal of Neurology*, 25(6):818–824, 6 2018.

[168] Z. Yaniv, B. C. Lowekamp, H. J. Johnson, and R. Beare. SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research. *Journal of Digital Imaging*, 31(3):290–303, 6 2018.

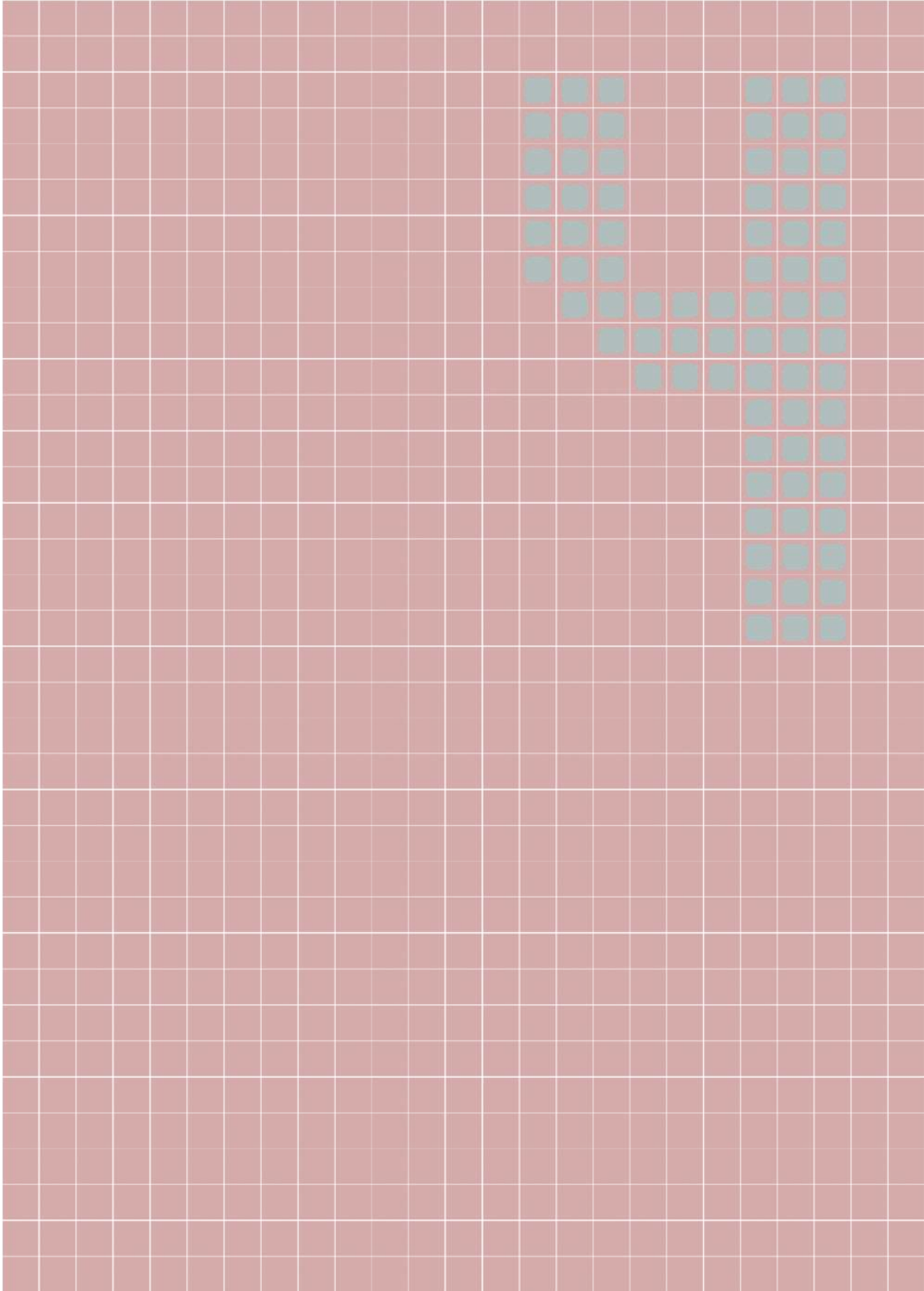
[169] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems*, volume 4, pages 3320–3328, 11 2014.

[170] Y. Yu, Y. Xie, T. Thamm, E. Gong, J. Ouyang, C. Huang, S. Christensen, M. P. Marks, M. G. Lansberg, G. W. Albers, and G. Zaharchuk. Use of Deep Learning to Predict Final Ischemic Stroke Lesions From Initial Magnetic Resonance Imaging. *JAMA network open*, 3(3):e200772, 3 2020.

[171] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, 31(3):1116–1128, 7 2006.

[172] S. F. Zaidi, A. Aghaebrahim, X. Urra, M. A. Jumaa, B. Jankowitz, M. Hammer, R. Nogueira, M. Horowitz, V. Reddy, and T. G. Jovin. Final infarct volume is a stronger predictor of outcome than recanalization in patients with proximal middle cerebral artery occlusion treated with endovascular therapy. *Stroke*, 43(12):3238–3244, 12 2012.

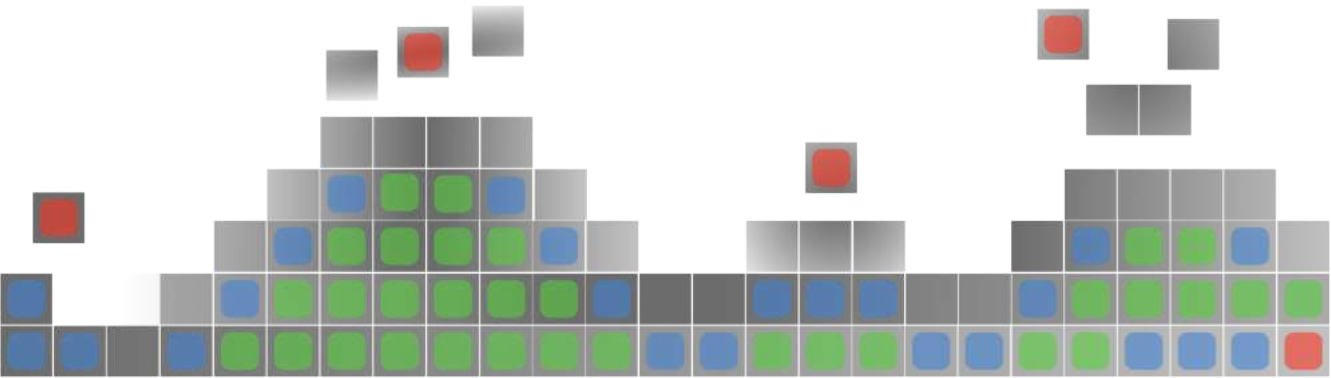
[173] Z. Zhou, V. Sodha, M. M. Rahman Siddiquee, R. Feng, N. Tajbakhsh, M. B. Gotway, and J. Liang. Models genesis: generic autodidactic models for 3d medical image analysis. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11767 LNCS, pages 384–393, 2019.



# Chapter 4

# Deep-Learning-Based Thrombus Localization and Segmentation in Patients with Posterior Circulation Stroke

**Based on:** R. Zoetmulder, A.A.E. Bruggeman, I.Išgum, E.Gavves, C.B.L.M. Majoie, L.F.M. Beenen, D.W.J. Dippel, N. Boodt, S.J. den Hartog, P.J. van Doormaal, S.A.P. Cornelissen, Y.B.W.E.M. Roos, J. Brouwer, W.J. Schonewille, A.F.V. Pirson, W.H. van Zwam, C. van der Leij, R.J.B. Brans, A.C.G.M van Es and H.A. Marquering (2022). "Deep-Learning-Based Thrombus Localization and Segmentation in Patients with Posterior Circulation Stroke." Diagnostics, 12(6), 1400.



# Abstract

Thrombus volume in posterior circulation stroke (PCS) has been associated with out-come, through recanalization. Manual thrombus segmentation is impractical for large scale analysis of image characteristics. Hence, in this study we develop the first automatic method for thrombus localization and segmentation on CT in patients with PCS. In this multicenter retrospective study, 187 patients with PCS from the MR CLEAN Registry were included. We developed a convolutional neural network (CNN) that segments thrombi and restricts the volume-of-interest (VOI) to the brainstem (Polar-UNet). Furthermore, we reduced false positive localization by removing small-volume objects, referred to as volume-based removal (VBR). Polar-UNet is benchmarked against a CNN that does not restrict the VOI (BL-UNet). Performance metrics included the intra-class correlation coefficient (ICC) between automated and manually segmented thrombus volumes, the thrombus localization precision and recall, and the Dice coefficient. The majority of the thrombi were localized. Without VBR, Polar-UNet achieved a thrombus localization recall of 0.82, versus 0.78 achieved by BL-UNet. This high recall was accompanied by a low precision of 0.14 and 0.09. VBR improved precision to 0.65 and 0.56 for Polar-UNet and BL-UNet, respectively, with a small reduction in recall to 0.75 and 0.69. The Dice coefficient achieved by Polar-UNet was 0.44, versus 0.38 achieved by BL-UNet with VBR. Both methods achieved ICCs of 0.41 (95% CI: 0.27-0.54). Restricting the VOI to the brainstem improved the thrombus localization precision, recall, and segmentation overlap compared to the benchmark. VBR improved thrombus localization precision but lowered recall.

## 4.1 Introduction

Stroke due to a large vessel occlusion in the posterior circulation accounts for approximately 1% [186; 191] of all cases of acute ischemic stroke and is associated with poor outcome [199]. Posterior circulation stroke (PCS) may have an acute onset or a progressive and stuttering onset, and may produce symptoms that are not typically associated with anterior stroke, such as vertigo and nausea [186]. Therefore, PCS is associated with a higher chance of misinterpretation and under-diagnosis in clinical practice than anterior circulation stroke (ACS) [175; 182], which has resulted in long delays in door-to-needle times of patients suffering from a PCS compared to patients suffering from an ACS [193; 196]. Reducing time from symptom onset to treatment may improve the outcome of patients suffering from a PCS who are treated with intravenous alteplase treatment and/or endovascular treatment [177; 181; 189].



Localizing an occluding thrombus causing PCS on radiological imaging is not a problem for expert neuro-radiologists and can be done quickly [197]. However, timely access to the services of an expert neuro-radiologist may not be possible in primary stroke centers because of the limited number of available neuro-radiologists. Automated localization of PCS may help to avoid misinterpretation in the case of a PCS. In addition to localization, segmentation of the thrombus would allow for automatic quantification of thrombus volume. Thrombus volume was previously reported to be negatively associated with recanalization and a higher likelihood of poor functional outcome [176; 194]. Recently proposed methods for the automated localization and segmentation of thrombi in stroke patients make use of convolutional neural networks (CNNs) [184; 198; 201]. Training CNNs requires imaging data from a large number of patients to reach an acceptable localization accuracy. For anterior circulation stroke, data from large numbers of patients are available even in single medical centers. However, data from patients suffering from a PCS are scarce.

As opposed to thrombi in patients with an ACS, thrombi in patients with a PCS occur in a more limited area around the brainstem. This characteristic can be used to improve the performance of a CNN-based thrombus localization and segmentation method. A CNN can learn how to center a moving Volume-

of-Interest (VOI) on the brainstem and evaluate only locations with a high likelihood of a thrombus. We hypothesize that tracking the brainstem with a small VOI improves the localization and segmentation of thrombi in the posterior circulation. In addition, we hypothesize that by removing small, segmented thrombi, the number of false positive thrombus localizations can be reduced.

In this study, we aim to develop and evaluate the first automatic CNN-based method for thrombus localization and segmentation on baseline on non-contrast CT (NCCT) and CT angiography (CTA) in patients suffering from a PCS.

4

## 4.2 Materials & Methods

### 4.2.1 Materials

The data used in this study were obtained from a prospective and nationwide observational study on endovascular treatment in the Netherlands (the MR CLEAN Registry [180]. The data were collected in 16 centers that performed endovascular treatment in the Netherlands between March 2014 and January 2019. The dataset used in our study included all patients from the MR CLEAN Registry that suffered from a PCS and consisted of NCCT and CTA scans from 268 patients. Hence, patients were also included if they suffered from an anterior circulation stroke (three patients) or a subarachnoid hemorrhage (one patient) in addition to a PCS. Patients were excluded if no baseline CTA was available, the scan quality was too low, registration of the NCCT to the CTA was unsuccessful, or if the patient had a non-occlusive thrombus or dissection. The Medical Ethics Committee of the Erasmus University Medical Center in Rotterdam, the Netherlands, approved the MR CLEAN Registry (MEC-2014-235). In addition, the institutional review board of each participating center approved the MR CLEAN research protocol.

### 4.2.2 Reference Annotations

Two types of annotations were made: reference thrombus segmentations, and annotations of the reference path, which runs through the structures of the brainstem. The reference segmentations of the thrombi were obtained by manual annotation. One trained observer (RZ) manually segmented the thrombi, and the second observer (AAEB) corrected the segmentations if necessary. A window width of 30 Hounsfield units (HU) and a center level of 35 HU was used for the NCCT scan, and a window width of 600 HU and a center level of 300 HU was used for the CTA scan. Annotations were made using ITK-Snap. For each patient, a case record form was available from an imaging core lab, which indicated the location of the occlusion and whether a Hyper-dense Artery Sign (HAS) was visible on the NCCT. The information about the location of the occlusion was used to guide the manual annotation. If the patient had a HAS, the NCCT was used to create the reference segmentation. Otherwise, the absence of arterial filling on the CTA was used to create the reference segmentation.

The reference path through the brainstem was annotated by one trained observer (RZ). The path started in the spinal cord, continued through the medulla oblongata, pons and midbrain. The reference path was marked by annotating a single voxel per axial slice in ITK Snap.

### 4.2.3 Preprocessing

To limit the size of the scan, all slices 25 cm below the top of the skull were excluded. Next, to align the anatomical structures in the CTA to those in the NCCT, the CTAs were registered to the NCCT by using rigid transformations. Finally, the voxel intensities were clipped between 100 and 200 HU and normalized between minus one and one. The pre-processing was done using SimpleITK and python 3.8.5.

### 4.2.4 CNNs for Automatic Posterior Circulation Thrombus Localization and Segmentation

Existing deep-learning methods result in false positive localization and segmentation if they are used on small objects. We developed a method to reduce the number of false positive localizations by restricting the VOI to a region that included the brain stem, which was most likely to contain a thrombus in patients suffering from a PCS. To restrict the VOI to the brainstem, the method had to learn how to move the VOI towards the brainstem on an axial slice. The movement from the center of the VOI at a given location was parame-terized as a circle with a radius and an angle (polar coordinates). The method learned to regress these polar coordinates and was named polar-UNet. In addition to regressing the polar coordinates, polar-UNet classified what action should be taken at each step: to move the VOI within the current axial slice, segment and move up one axial slice, or to stop the inference procedure.



The architecture of the regression and classification heads of the polar-UNet is shown in Figure 4.1A. To create the polar-UNet, the regression and classification head were added to the down-sampling path of the baseline UNet (BL-UNet), which is shown in Figure 4.1B. Features were extracted at four levels of the down-sampling path. These features were input into four separate convolutional layers and global average-pooled. Next, the extracted features were concatenated with the relative coordinates of the VOI in the scan volume and input into two shared fully connected layers, which consisted of 256 neurons each. The output of the shared layer was passed to the individual network heads, which regressed either the angle or the radius, or classified the action. The angle regression head used a TanH activation, the radius regression head used a sigmoid activation, and the action classification head used the softmax activation function in the output layer. Finally, polar-UNet also had the same up-sampling path as BL-UNet.

The BL-UNet was inspired by U-Net [192] and consisted of 3D ResNet [179] blocks. The down-sampling path started with two convolutional layers, which consisted of a kernel size of three and a stride of one. It was followed by a max-pooling operation with a kernel size and stride of two. Subsequently, three down-sampling blocks, which consisted of a 3D ResNet layer each,

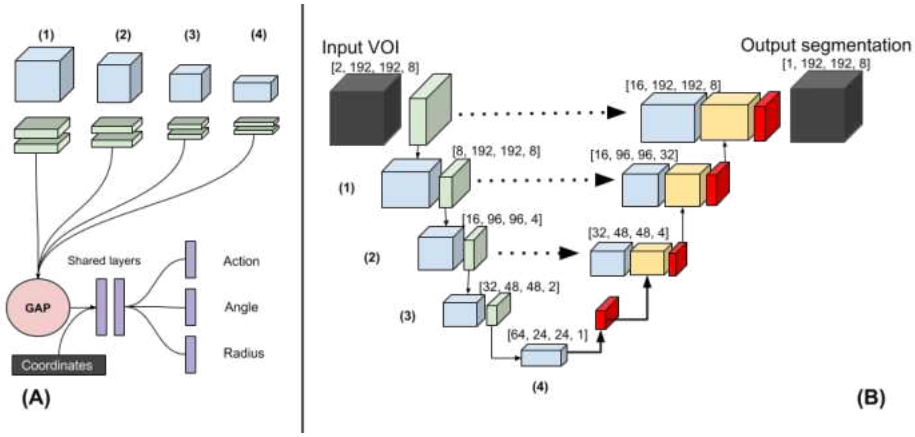



**Figure 4.1:** **(A)** The polar-UNet is created by attaching the following operations to the BL-UNet. The features from four levels in the down-sampling path are input to four blocks consisting of two convolutions. The output is global average-pooled and concatenated before being passed to the fully connected shared layers. Finally, individual fully connected layers are used to classify the action and regress the angle and the radius. **(B)** Three-dimensional Baseline UNet (BL-UNet). Three-dimensional ResNet blocks followed by a max-pooling operation (green) were used to construct the down-sampling path (left). Three-dimensional ResNet blocks followed by a transposed convolution were used to construct the up-sampling path (right). The features generated in the down- and up-sampling paths are blue and yellow, respectively. Skip connections were added between the up-sampling and down-sampling paths.

were added. The first two blocks were followed by a max-pooling layer with a stride and pooling size of two. The up-sampling path started with a transposed convolution. Next, two up-sampling blocks, which consisted of a ResNet block followed by a transposed convolution, were added. Each transposed convolution had a stride of two and a kernel size of three. The features of the down-sampling path were concatenated with the features in the up-sampling path. The up-sampling path was followed by a three-by-three convolution and a one-by-one convolution. During inference, BL-UNet was applied to a non-overlapping grid of VOIs extracted from the entirety of each scan. The CNNs were implemented using Pytorch 1.5.1 [190].



### 4.2.5 Experimental Setup

The BL-UNet was randomly initialized and trained on VOIs sampled from the scan volumes. To ensure the data contained sufficient positive examples, 40 percent of the sampled VOIs contained a thrombus. The up-sampling and down-sampling paths of the polar-UNet were initialized by reusing the weights obtained by the pre-trained BL-UNet and were not updated during training. The BL-UNet used group normalization with 4 groups, the ReLU activation, a batch size of 32, a cyclical learning rate schedule, the Adam optimizer and a weight decay of $20^{-6}$. The BL-UNet was trained for 300 epochs with 148 iterations per epoch.

The polar-UNet applied batch-normalization to the layers that were updated during training and used a batch size of 128. The focal loss function [183] was used for the segmentation and classification tasks, and the $L_2$ loss for the regression tasks. The polar-UNet was trained for 100 epochs with 148 iterations per epoch. Both CNNs used a maximum learning rate of $10^{-3}$, a minimum learning rate of $20^{-5}$, and the weight decay was set to $20^{-5}$. The learning rate was linearly increased to the maximum value in 300 iterations and decreased to the minimum value in 300 iterations.

The polar coordinate reference values that were needed to move the VOI to the area around the brainstem were calculated using the reference path. The angle reference value was calculated as the angle between the coordinates of the center of the VOI and the coordinates of the reference path on the same axial slice. To allow for easier optimization, the angle reference values

were normalized between 0 and 2. The radius reference value indicated the step size that the VOI had to be moved to minimize the Euclidean distance between the center of the VOI and the reference path on the same axial slice. The radius was limited to the width of the VOI and normalized between 0 and 1 to allow for easier optimization.

The inference procedure of the polar-UNet is shown in the supplementary materials Algorithm 1. We initialized the VOI at the center of the most caudal axial slice in the scan. Per axial slice, the VOI could only be moved within the axial slice 10 times before the VOI location was moved up by one axial slice. If the VOI moved out of the bounds of the volume, the VOI location was moved up by one slice and reset to the center of the axial slice.

Data augmentation was applied during training. The axial slices in the scan volumes were rotated at an angle between minus ten and ten degrees and, with a probability of 0.5, were flipped. Furthermore, the scan volumes were magnified by a factor between 0.9 and 1.1 and translated in the axial plane by a maximum of 80 voxels. Polar coordinates reference values were updated accordingly. The sampled VOIs had dimensions of $192 \times 192 \times 8$.

We used stratified five-fold cross-validation to evaluate the performance of the CNN architectures. The hyper-parameters used in our study were found by evaluating multiple sets of hyper-parameters on the first fold and selecting the ones with the highest testing performance. To prevent inflation of the results due to the first fold being used to find the optimal hyper-parameters, this fold was excluded from the evaluation. To improve the localization precision of the CNNs, only connected components larger than 0.065 mL were included in further analyses. We refer to this step as Volume-based Removal (VBR).

### 4.2.6 Evaluation

To evaluate the localization performance of our models, we calculated the thrombus localization precision and recall. Hence, the number of True Positive (TP), False Positive (FP), and False Negative (FN) localizations had to be calculated. To calculate the TP, FP, and FN localizations, the segmentation maps had to be divided into individual connected regions. Thus, a connected-components analysis was run on the automatically created and

manually annotated segmentations. If two connected components from the manually annotated segmentation and the corresponding automatically created segmentation had ten percent or more of their voxels overlapping, this was counted as a TP localization. If a connected component from the automatically created segmentation had less than ten percent overlap with a connected component from the manually annotated segmentation it was counted as a FP localization. If a connected component from the manually annotated segmentation did not have more than ten percent overlap with any of the connected components from the automatically created segmentation, it was counted as a FN.



The intra-class correlation coefficient (ICC) was used to evaluate the agreement be-tween the volume derived from the automatically and manually segmented thrombi. We included the 95% confidence interval (95% CI) in our analysis. Furthermore, we performed a Bland-Altman analysis to evaluate the bias and limits of agreement of the volume measurements. We tested whether the ICC of the automated segmentations created by the two CNN architectures with and without VBR differed significantly by using Fishers r to z transformation.

The segmentation overlap of the CNNs was evaluated by calculating the Dice coefficient between the reference and the automatic segmentation. If no thrombus was visible in the scan, the Dice coefficient would always equal zero and would always deflate the results artificially. Hence, patients with no visible thrombus in the scan were excluded from the analysis of the Dice coefficient.

The obtained metrics were tested for normality with Shapiro-Wilk tests. If the obtained metrics followed a normal distribution, paired t-tests were used for comparison, otherwise Wilcoxon rank-sum tests were used. A Bonferroni correction was applied to P-values to correct for family-wise error. Finally, thrombi in the posterior circulation most commonly occur in the basilar artery. Therefore, we evaluated whether thrombus location influenced true positive thrombus localization, volume agreement, and segmentation overlap. Thrombus location was obtained from the case report file of each patient. The Python library Pingouin, version 0.3.1 was used for all statistical testing.

## 4.3 Results

The patient flow chart is shown in the supplementary materials, Figure 4.4. In total, 187 patients were included in our study. The baseline characteristics of the included patients are shown in the supplementary materials, Table 4.2.

In 185 out of 187 scans (99%) a thrombus was visible and was manually annotated. After exclusion of the patients in the test set of the first fold, due to this fold being used for hyper-parameter optimization, 149 scans were left to evaluate the results on.

The thrombus localization recall varied between 129 (69%) and 146 (82%). The thrombus localization recall before and after VBR for the BL-UNet and polar-UNet is shown in Table 1. Overall, the Polar-UNet improved the thrombus localization recall over BL-UNet, and VBR reduced the thrombus localization recall of both CNNs.



| Localization Recall | | Overall | VA | BA | VA + BA | BA + PCA | VA + BA + PCA | PCA | NAV |
|---|---|---|---|---|---|---|---|---|---|
| | | N = 149 | N = 9 | N = 37 | N = 22 | N = 49 | N = 8 | N = 19 | N = 5 |
| **No VBR** | **BL-UNet** | 0.78 | 0.58 | 0.86 | 0.76 | 0.89 | 1 | 0.46 | 0 |
| | **Polar-Unet** | 0.82 | 0.75 | 0.94 | 0.83 | 0.87 | 0.9 | 0.46 | 0.33 |
| **VBR** | **BL-UNet** | 0.69 | 0.42 | 0.61 | 0.69 | 0.73 | 0.8 | 0.15 | 0 |
| | **Polar-UNet** | 0.75 | 0.67 | 0.81 | 0.83 | 0.8 | 0.9 | 0.31 | 0.33 |
| **Localization Precision** | | | | | | | | | |
| **No VBR** | **BL-UNet** | 0.09 | 0.10 | 0.09 | 0.12 | 0.11 | 0.15 | 0.04 | 0 |
| | **Polar-UNet** | 0.14 | 0.27 | 0.14 | 0.17 | 0.13 | 0.23 | 0.05 | 0.04 |
| **VBR** | **BL-UNet** | 0.56 | 0.56 | 0.55 | 0.65 | 0.53 | 0.73 | 0.29 | 0 |
| | **Polar-UNet** | 0.62 | 0.57 | 0.66 | 0.77 | 0.55 | 1 | 0.4 | 0.2 |
| **Dice Coefficient** | | N = 147 * | | | | | | | |
| **No VBR** | **BL-UNet** | 0.38 | 0.30 | 0.42 | 0.42 | 0.45 | 0.5 | 0.15 | 0.09 |
| | **Polar-UNet** | 0.44 | 0.34 | 0.49 | 0.50 | 0.48 | 0.56 | 0.2 | 0.26 |
| **VBR** | **BL-UNet** | 0.35 | 0.21 | 0.37 | 0.42 | 0.44 | 0.5 | 0.07 | 0 |
| | **Polar-UNet** | 0.44 | 0.33 | 0.50 | 0.52 | 0.48 | 0.62 | 0.12 | 0.17 |

**Table 4.1:** Thrombus localization recall and precision in the posterior circulation and the Dice coefficient of the BL-UNet and Polar-UNet, with and without volume-based removal (VBR). Results are shown for the entire dataset (overall), the vertebral artery (VA), basilar artery (BA), posterior cerebral artery (PCA), thrombi which crossed multiple segments (VA + BA, BA + PCA, VA + BA + PCA), and patients for whom the thrombus location was not indicated (NAV). ∗ Two scans did not have a visible thrombus. Hence, the Dice coefficient was calculated for 8 patients with a thrombus in the VA and 4 with an unspecified occlusion location (NAV).

The thrombus localization precision before and after VBR for the BL-UNet and polar-UNet is shown in Table 4.1. In all cases, Polar-UNet improved the thrombus localization precision over the BL-UNet, and VBR improved the thrombus localization precision of both CNNs (Figure 4.2). However, for all

cases the thrombus localization precision is low with an overall maximum of 62%.

The median thrombus volume was 0.15 (IQR: 0.070.34) mL. The ICC for the volume agreement of the BL-UNet and Polar-UNet without VBR were 0.38 (95% CI: 0.230.51) and 0.4 (95% CI: 0.250.53). With VBR, the ICCs for the volume agreement of the BL-UNet and Polar-UNet were both 0.41 (95% CI: 0.270.54). There was no statistically significant difference between the volumetric agreement of the various methods. The Bland-Altman analysis for each of the CNNs resulted in biases ranging from 0.06 mL for the Polar-UNet without VBR to 0.06 mL for the BL-UNet with VBR. The LoAs were the smallest for the BL-UNet with VBR, ranging from 0.64 to 0.77 mL. The LoAs were largest for the Polar-UNet without VBR, ranging from 0.83 to 0.71 mL. The bias and LoAs for the volume analysis are shown in the supplementary materials, Table 4.4, and the Bland-Altman results are shown in Figure 4.3.

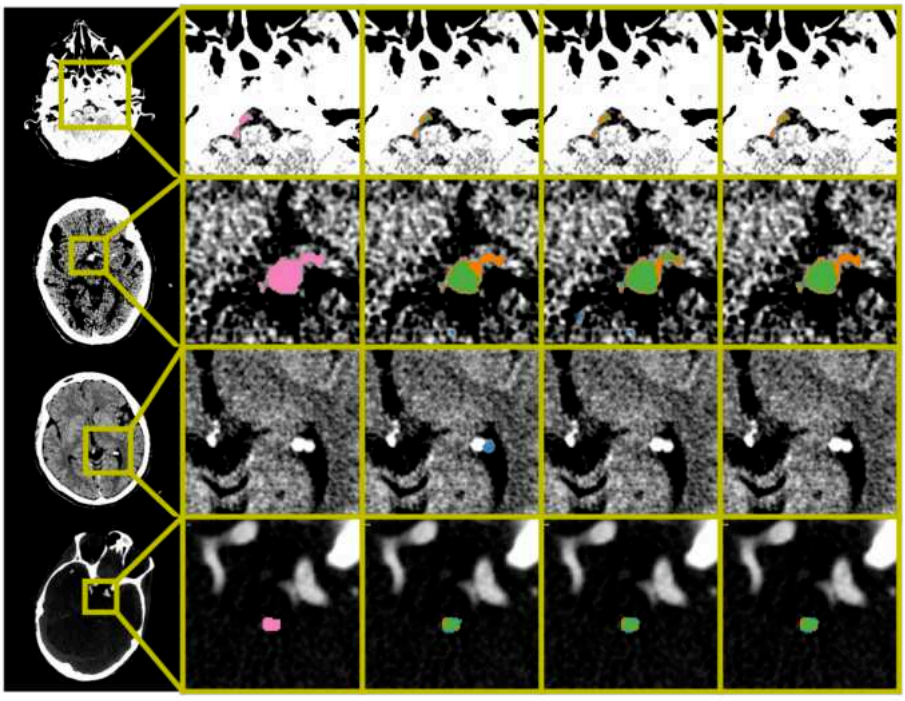


**Figure 4.2:** Examples of automatic segmentation results obtained by BL-UNet and Polar-UNet. From the left to right column: The original scan with a bounding box indicating the zoom location, the ground truth segmentation map, the results obtained from the BL-UNet without volume-based removal (VBR), the results obtained from the Polar-UNet without VBR, and the results obtained from the Polar-UNet with VBR. The top three rows display NCCT scans; the bottom row shows a CTA scan. The top row shows the difficulty all CNN methods have with segmenting a thrombus in the vertebral arteries. The second row from the top shows an example of small false positives removed by the VBR step. The third row from the top row shows false positives that are re-moved by restricting the volume-of-interest to the posterior circulation with Polar-UNet. The bottom row shows an example of a scan without a hyperdense artery sign. The segmentation maps show the ground truth (pink), true positive (green), false negative (orange) and false positive (blue). The NCCT scans were plotted using a window center level of 35, with a window width of 30. The CTA scan was plotted using a window center level of 300, with a window width of 600.

The Dice coefficients for the automated thrombus segmentation are shown in Table 4.1. The results of the Wilcoxon rank-sum test are shown in the supplementary materials, Table 4.3. The results show that the Polar-UNet with VBR results in a statistically significantly larger Dice coefficient than the other methods. Furthermore, the Polar-UNet without VBR results in a

significantly greater Dice coefficient than the BL-UNet with or without VBR.



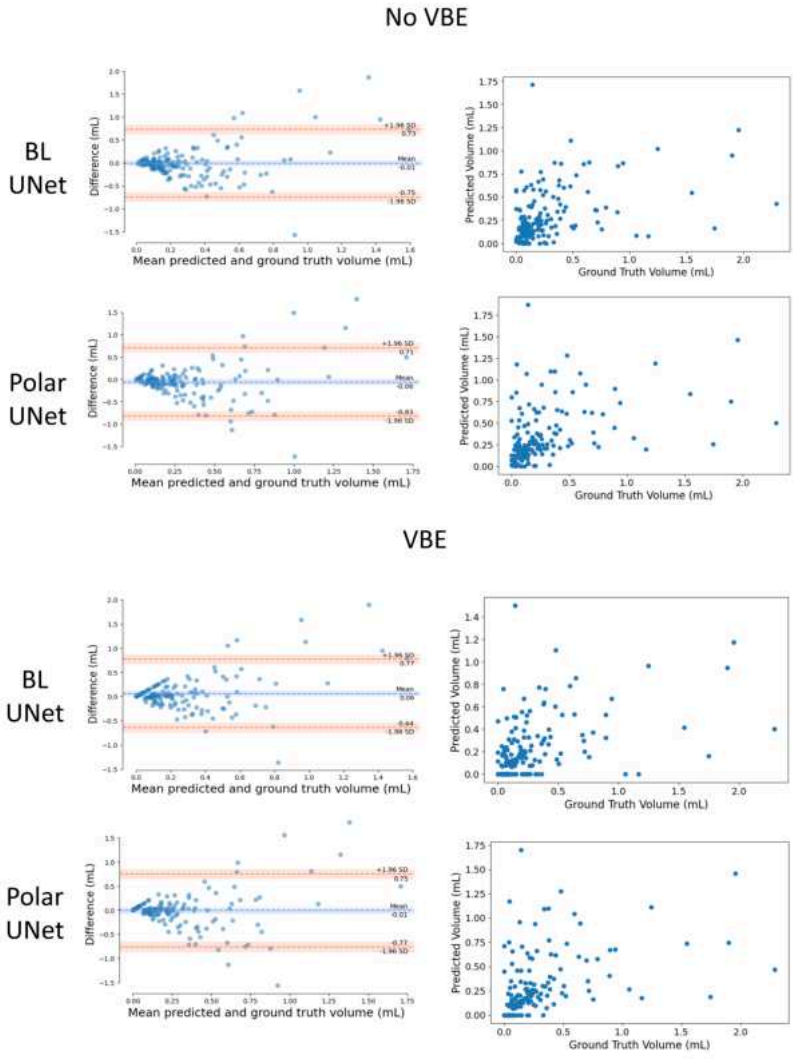



**Figure 4.3:** Comparison of the automated and manually segmented volume for the BL-UNet and Polar-UNet with and without volume-based removal (VBR). Left column: BlandAltman plots of the lesion volumes. The volumes corresponding to the reference and automatic segmentations are shown on the x-axis, and the volume difference is shown on the y-axis. Right column: scatter plots comparing lesion volumes derived from the reference segmentations (y-axis) and from the automatic segmentations determined by the CNN (x-axis).

## 4.4 Discussion

In this study we have presented the first CNN-based method to segment and localize thrombi in patients suffering from PCS. The restriction of the VOI to areas in the vicinity of the brainstem improved the segmentation overlap and thrombus localization precision and recall over the baseline method. The VBR of small objects improved the thrombus localization precision of our method but reduced the recall. Regardless of whether VBR of small objects is applied, the bias of the automated volume quantifications is small, but the limits of agreement are large. Finally, there is room to improve the methods volume agreement and segmentation overlap.



Our study is the first to address the automated localization and segmentation of thrombi in the posterior circulation on CTA and NCCT. Prior work has focused on identifying scans with a Large Vessel Occlusion (LVO) in the anterior circulation on a single-phase [174; 195] and multi-phase [187; 198] CTA. Two studies included patients suffering from both types of stroke, either in the posterior or anterior circulation [188; 198]. However, a limitation is that not all studies report results specific to the posterior stroke patient group [198]. Another limitation of prior work is that the thrombus is not localized, nor is it segmented. Only the presence or absence of an LVO in a scan is indicated.

Other work aimed to localize or segment thrombi on NCCT scans of patients suffering from either anterior or posterior circulation stroke [184; 185; 201] . One study used feature extraction and a random forest classifier [185]. This study did not mention the location of the thrombi in their dataset. Other work has used deep learning to localize or segment thrombi on NCCT scans of patients in the anterior circulation [184; 201]. A key feature of these methods is that a comparison between hemispheres was made to improve localization or segmentation performance. A comparison between hemispheres has no added value to PCS thrombus localization and segmentation, because of the absence of lateral symmetry of the basilar artery.

Our study has several limitations. First, our method is not accurate at localizing and segmenting thrombi in the vertebral and posterior cerebral artery. This is likely because of the rareness of thrombi in these sections. Hence, data from few patients with occlusions in these regions were available to train our

method on. In contrast, our method does achieve a good localization precision and a reasonable segmentation accuracy for thrombi in the basilar artery, which is the most common location for PCS. Second, our data was annotated by one trained observer. Hence, the inter-observer agreement could not be assessed. However, the manual segmentations were verified and corrected by another experienced observer. Third, the method achieved a low volumetric agreement and segmentation overlap between the manually and automatically segmented thrombi. Still, the volumetric agreement is similar to those reported in a study that focuses on thrombus segmentation and localization in the anterior circulation [201]. This is despite our method not being able to use lateral symmetry information, as was done by methods for anterior stroke localization and segmentation. Furthermore, the low Dice coefficient can be explained by the small size of the thrombi. For small object segmentation, the Dice coefficient is over-sensitive to small errors [200]. Fourth, manual segmentation of thrombi in patients without a HAS and with poor collateral status on single phase CTA leads to an overestimation of thrombus length [178]. The overestimation of thrombus length is caused by a delayed contrast arrival. Hence, our method may, just like manual segmentations, overestimate the true length of the occluding thrombus. Fifth, since there were no additional large datasets with patients with posterior stroke available, our method lacks validation on an external dataset. Thus, results may vary when our algorithm is trained on or applied to other datasets. However, the data in our study were collected from sixteen different centers. Our algorithm was nonetheless evaluated on a heterogeneous set of scans. As more image data of patients with posterior stroke become available, our method may be improved and validated by including this image data.

A valuable application of deep learning is to automate time-critical and labor-intensive tasks such as the localization and segmentation of small objects in radiological imaging. Our study has shown that for small object segmentation and localization, a standard deep-learning-based segmentation method would result in large amounts of false positives. In this study we have presented an approach that combines restricting the segmentation and localization to the region in which the occlusion occurs, and excluding small objects. We have shown that this combination reduces the number of false

positives.

The automated posterior thrombus localization and segmentation method is the first step towards a method that allows associations between thrombus characteristics, such as thrombus volume and length, and PCS outcome to be easily studied in clinical trials.

## 4.5 Supplementary Materials

### 4.5.1 Appendix A

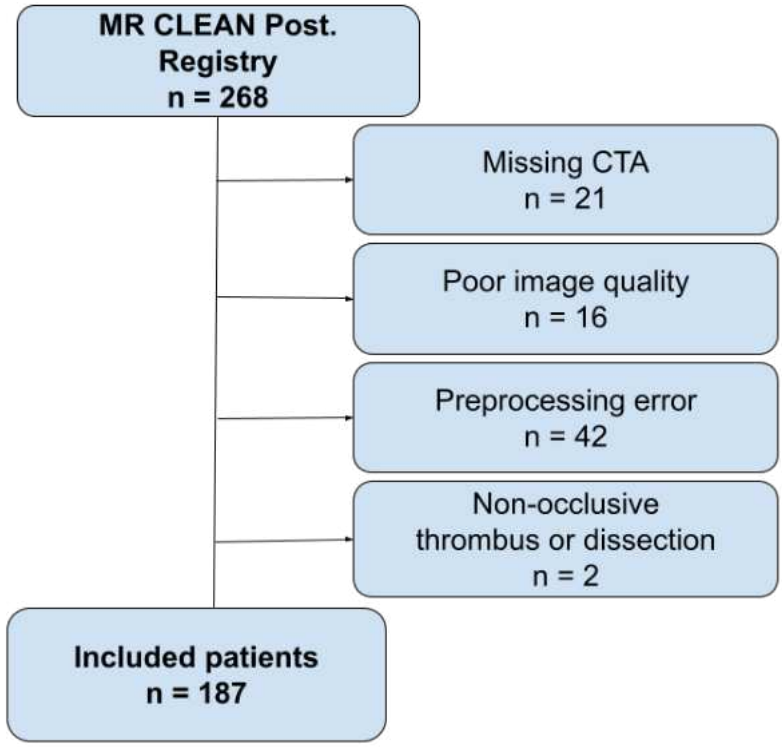



**Figure 4.4:** Flow chart showing the inclusion of the patients with a posterior circulation stroke in the MR CLEAN Registry. Exclusion criteria were: no CTA available (n = 21), poor image quality (n = 16), preprocessing error (n = 42), or a non-occlusive thrombus or dissection (n = 2). In total, 187 patients were included.

### 4.5.2 Appendix B

| Parameter | Posterior Stroke |
|---|---|
| Clinical | |
| Age, years, median (IQR) | 65.5 (20.8) |
| Sex, F, no. [%] | 72/186 [38.7] |
| NIHSS at baseline, median(IQR) [N] | 15 [22] (183) |
| Prior Conditions | |
| Diabetes mellitus, no. [%] | 32/184 [17.4] |
| Hypertension, no. [%] | 97/181 [53.6] |
| Stroke, no. [%] | 30/183 [16.4] |
| Atrial fibrillation, no. [%] | 25/184 [13.6] |
| Pre-Stroke mRS | |
| 0, no. [%] | 126/180 [70] |
| 1, no. [%] | 19/180 [10.6] |
| 2, no. [%] | 17/180 [8.4] |
| 3, no. [%] | 7/180 [3.9] |
| 4, no. [%] | 8/180 [4.4] |
| 5, no. [%] | 3/180 [1.7] |
| Treatment | |
| IV Thrombolysis, no. [%] | 102/186 [54.8] |
| Time | |
| Stroke onset to IVT, minutes, median [IQR](N) | 98 [73] (93) |

**Table 4.2:** Baseline characteristics, treatment, and time data for included patients with posterior circulation stroke.

| Method 1 | Method 2 | Dice Coefficient | | Bias | |
|---|---|---|---|---|---|
| | | W | p-Value | W | p-Value |
| BL-UNet | BL-UNet (with VBR) | 3039 | >0.05 | 3698 | >0.05 |
| BL-UNet | Polar-Unet | 2415 | <0.01 | 5392 | >0.05 |
| BL-UNet | Polar-Unet (with VBR) | 2481 | <0.01 | 4361 | >0.05 |
| BL-UNet (with VBR) | Polar-Unet | 2551 | <0.01 | 4832 | >0.05 |
| BL-UNet (with VBR) | Polar-Unet (with VBR) | 1367 | <0.01 | 3703 | >0.05 |
| Polar-Unet | Polar-Unet (with VBR) | 2431 | <0.05 | 3153 | <0.01 |

**Table 4.3:** 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. The Polar UNet (with VBR) produced a significantly greater Dice coefficient relative to the other methods. The Polar- UNet produced a significantly greater Dice coefficient relative to the BL-UNet (with VBR) and the BL-UNet. No method significantly improved the bias of the volume estimation other than the Polar-UNet (with VBR) over the Polar-UNet.



| | | Bias | LoA |
|---|---|---|---|
| **No VBR** | **BL-UNet** | 0.01 mL | 0.75 to 0.73 mL |
| | **Polar-UNet** | 0.06 mL | 0.83 to 0.71 mL |
| **VBR** | **BL-UNet** | 0.06 mL | 0.64 to 0.77 mL |
| | **Polar-Unet** | 0.01 mL | 0.77 to 0.75 mL |

**Table 4.4:** Bias and limits of agreement between the automatically and manually segmented thrombus volumes, respectively, for the BL-UNet and Polar-UNet with and without volume-based removal (VBR).

### 4.5.3 Appendix C

**Algorithm 1:** Inference procedure for the Polar-UNet

**Data:** scan : $S_{H\times W\times D}$ ,Max Iterations per Slice : $M$ = 10 ,Starting Coordinates : $C$ = $(256, 256)$ ,Predicted Segmentation : $P = O_{H\times W\times D}$

**for** $i \leftarrow 0$ **to** $D$ **do**
  **for** $j \leftarrow 0$ **to** $M$ **do**
    $VOI \leftarrow \text{extract}(S, C)$ ;
    $angle \leftarrow \text{predict}(VOI)$ ;
    $radius \leftarrow \text{predict}(VOI)$ ;
    $action \leftarrow \text{predict}(VOI)$ ;
    $VOI_{segmentation} \leftarrow \text{predict}(VOI)$ ;
    **if** $action == move$ **then**
      $C \leftarrow \text{updateCoordinates}(angle, radius)$;
    **else if** $action == segment$ **then**
      $P \leftarrow \text{updateSegmentation}(VOI_{segmentation}, C)$ ;
      $C \leftarrow \text{updateCoordinatesUpward}()$ ;
    **else if** $action == stop$ **then**
      stopInferenceProcedure() ;
  **end**
**end**
**return** $P$

# Bibliography

[174] S. A. Amukotuwa, M. Straka, H. Smith, R. V. Chandra, S. Dehkharghani, N. J. Fischbein, and R. Bammer. Automated detection of intracranial large vessel occlusions on computed tomography angiography a single center experience. *Stroke*, 50(10):2790–2798, 10 2019.

[175] A. E. Arch, D. C. Weisman, S. Coca, K. V. Nystrom, C. R. Wira, and J. L. Schindler. Missed Ischemic Stroke Diagnosis in the Emergency Department by Emergency Medicine and Neurology Services. *Stroke*, 47(3):668–673, 3 2016.

[176] J. H. Baek, J. Yoo, D. Song, Y. D. Kim, H. S. Nam, B. M. Kim, D. J. Kim, H. S. Lee, and J. H. Heo. Predictive value of thrombus volume for recanalization in stent retriever thrombectomy. *Scientific Reports 2017 7:1*, 7(1):1–8, 11 2017.

[177] T. Dorák, R. Herzig, M. Kuliha, R. Havlíček, D. Školoudík, D. Šaák, M. Köcher, V. Procházka, J. Lacman, F. Charvát, A. Krajina, D. Krajíčková, M. Král, T. Veverka, M. Roubec, L. Hajduková, and J. Zapletalová. Endovascular treatment of acute basilar artery occlusion: time to treatment is crucial. *Clinical Radiology*, 70(5):e20–e27, 5 2015.

[178] A. M. Frölich, D. Schrader, E. Klotz, R. Schramm, K. Wasser, M. Knauth, and P. Schramm. 4D CT Angiography More Closely Defines Intracranial Thrombus Burden Than Single-Phase CT Angiography. *American Journal of Neuroradiology*, 34(10):1908–1913, 10 2013.

[179] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2016-Decem, pages 770–778, 12 2016.

[180] I. G. Jansen, M. J. Mulder, and R. J. B. Goldhoorn. Endovascular treatment for acute ischaemic stroke in routine clinical practice: Prospective, observational cohort study (MR CLEAN Registry). *BMJ (Online)*, 360, 3 2018.

[181] W. Kong, J. Yuan, J. Huang, J. Song, C. Zhao, H. Sang, W. Luo, D. Xie, F. Gao, H. Li, J. Luo, S. Liu, D. Xue, Y. Yu, F. Li, Z. Qiu, W. Zi, and Q. Yang. Outcomes of Endovascular Therapy in Acute Basilar Artery Occlusion With Severe Symptoms. *JAMA Network Open*, 4(12):e2139550–e2139550, 12 2021.

[182] N. M. Lever, K. V. Nyström, J. L. Schindler, J. Halliday, C. Wira, and M. Funk. Missed opportunities for recognition of ischemic stroke in the emergency department. *Journal of Emergency Nursing*, 39(5):434–439, 9 2013.

[183] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):318–327, 8 2017.

[184] A. Lisowska, E. Beveridge, K. Muir, and I. Poole. Thrombus detection in CT brain scans using a convolutional neural network. In *BIOIMAGING 2017 - 4th International Conference on Bioimaging, Proceedings; Part of 10th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2017*, volume 2017-Janua, pages 24–33, 2017.

[185] P. Löber, B. Stimpel, C. Syben, A. Maier, H. Ditt, P. Schramm, B. Raczkowski, and A. Kemmling. Automatic thrombus detection in non-enhanced computed tomography images in patients with acute ischemic stroke. In *VCBM 2017 - Eurographics Workshop on Visual Computing for Biology and Medicine*, pages 125–129, 2017.

[186] H. P. Mattle, M. Arnold, P. J. Lindsberg, W. J. Schonewille, and G. Schroth. Basilar artery occlusion, 11 2011.

[187] M. Meijs, F. J. Meijer, M. Prokop, B. v. Ginneken, and R. Manniesing. Image-level detection of arterial occlusions in 4D-CTA of acute stroke patients using deep learning. *Medical Image Analysis*, 66:101810, 12 2020.

[188] M. Olive-Gadea, C. Crespo, C. Granes, M. Hernandez-Perez, N. Pérez De La Ossa, C. Laredo, X. Urra, J. Carlos Soler, A. Soler, P. Puyalto, P. Cuadras, C. Marti, and M. Ribo. Deep Learning Based Software to

Identify Large Vessel Occlusion on Noncontrast Computed Tomography. *Stroke*, pages 3133–3137, 2020.

[189] J. Pagola, M. Ribo, J. Alvarez-Sabin, M. Rubiera, E. Santamarina, O. Maisterra, R. Delgado-Mederos, G. Ortega, M. Quintana, and C. A. Molina. Thrombolysis in Anterior Versus Posterior Circulation Strokes: Timing of Recanalization, Ischemic Tolerance, and Other Differences. *Journal of Neuroimaging*, 21(2):108–112, 4 2011.

[190] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, 32, 2019.

[191] F. A. V. Pirson, N. Boodt, J. Brouwer, A. A. Bruggeman, S. J. den Hartog, R.-J. B. Goldhoorn, L. C. Langezaal, J. Staals, W. H. van Zwam, C. van der Leij, R. J. Brans, C. B. Majoie, J. M. Coutinho, B. J. Emmer, D. W. Dippel, A. van der Lugt, J.-A. Vos, R. J. van Oostenbrugge, W. J. Schonewille, Y. B. Roos, J. Boiten, J. A. Vos, I. G. Jansen, M. J. Mulder, K. C. Compagne, M. Kappelhof, W. H. Hinsenveld, B. Roozenbeek, A. v. d. Lugt, Y. B. Roos, J. A. Vos, M. J. Wermer, M. A. van Walderveen, A. C. van Es, J. Hofmeijer, J. M. Martens, G. J. Lycklama à Nijeholt, J. Boiten, S. F. de Bruijn, L. C. van Dijk, H. B. van der Worp, R. H. Lo, E. J. van Dijk, H. D. Boogaarts, J. de Vries, P. L. de Kort, J. van Tuijl, J. P. Peluso, P. Fransen, J. S. van den Berg, B. A. van Hasselt, L. A. Aerden, R. J. Dallinga, M. Uyttenboogaart, O. Eschgi, R. P. Bokkers, T. H. Schreuder, R. J. Heijboer, K. Keizer, L. S. Yo, H. M. den Hertog, E. J. Sturm, P. J. Brouwers, G. J. Lycklama à Nijeholt, M. A. van Walderveen, M. E. Sprengers, S. F. Jenniskens, R. van den Berg, A. J. Yoo, L. F. Beenen, A. A. Postma, S. D. Roosendaal, B. F. van der Kallen, I. R. van den Wijngaard, A. C. van Es, J. M. Martens, L. S. Yo, J. A. Vos, J. Bot, P.-J. van Doormaal, A. Meijer, E. Ghariq, R. P. Bokkers, M. P. van Proosdij, G. Menno Krietemeijer, J. P. Peluso, H. D. Boogaarts, R. Lo, D. Gerrits, W. Dinkelaar, A. P. Appelman, B. Hammer, S. Pegge, A. van der Hoorn, S. Vinke, S. Cornelissen,

Y. B. Roos, G. J. L. à Nijeholt, J. Boiten, J. A. Vos, J. Hofmeijer, J. M. Martens, H. B. van der Worp, R. H. Lo, J. Hofmeijer, H. Z. Flach, H. F. Lingsma, N. e. Ghannouti, M. Sterrenberg, W. Pellikaan, R. Sprengers, M. Elfrink, M. Simons, M. Vossers, J. de Meris, T. Vermeulen, A. Geerlings, G. van Vemde, T. Simons, G. Messchendorp, N. Nicolaij, H. Bongenaar, K. Bodde, S. Kleijn, J. Lodico, H. Droste, M. Wollaert, S. Verheesen, D. Jeurrissen, E. Bos, Y. Drabbe, M. Sandiman, N. Aaldering, B. Zweedijk, J. Vervoort, E. Ponjee, S. Romviel, K. Kanselaar, D. Barning, E. Venema, V. Chalos, R. R. Geuskens, T. van Straaten, S. Ergezen, R. R. Harmsma, D. Muijres, A. de Jong, O. A. Berkhemer, A. M. Boers, J. Huguet, P. Groot, M. A. Mens, K. R. van Kranendonk, K. M. Treurniet, M. L. Tolhuisen, H. Alves, A. J. Weterings, E. L. Kirkels, E. J. Voogd, L. M. Schupp, S. L. Collette, A. E. Groot, N. E. LeCouffe, P. R. Konduri, H. Prasetya, N. Arrarte-Terreros, and L. A. Ramos. Endovascular Treatment for Posterior Circulation Stroke in Routine Clinical Practice: Results of the Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands Registry. *Stroke*, page STROKEAHA121034786, 11 2021.

[192] O. Ronneberger, P. Fischer, and T. Brox. 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)*, volume 9351, pages 234–241, Munich, Germany, 2015. Springer Verlag.

[193] A. Sarraj, S. Medrek, K. Albright, S. Martin-Schild, W. Bibars, F. Vahidy, J. C. Grotta, and S. I. Savitz. Posterior circulation stroke is associated with prolonged door-to-needle time. *International Journal of Stroke*, 10(5):672–678, 7 2015.

[194] G. Schulte-Altedorneburg, G. F. Hamann, M. Mull, D. Kühne, M. Liebetrau, W. Weber, H. Brückmann, and T. E. Mayer. Outcome of acute vertebrobasilar occlusions treated with intra-arterial fibrinolysis in 180 patients. *American Journal of Neuroradiology*, 27(10):2042–2047, 2006.

[195] S. A. Sheth, V. Lopez-Rivera, A. Barman, J. C. Grotta, A. J. Yoo, S. Lee, M. E. Inam, S. I. Savitz, and L. Giancardo. Machine Learning-Enabled

Automated Determination of Acute Ischemic Core From Computed Tomography Angiography. *Stroke*, 50(11):3093–3100, 11 2019.

[196] P. Sommer, L. Seyfang, A. Posekany, J. Ferrari, W. Lang, E. Fertl, W. Serles, T. Töll, S. Kiechl, and S. Greisenegger. Prehospital and intra-hospital time delays in posterior circulation stroke: results from the Austrian Stroke Unit Registry. *Journal of Neurology*, 264(1):131–138, 1 2017.

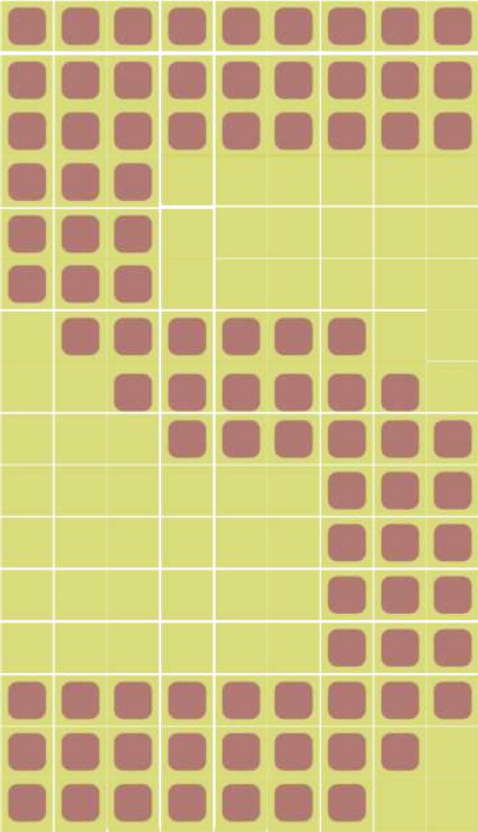
[197] A. Srinivasan, M. Goyal, C. Lum, T. Nguyen, and W. Miller. Processing and interpretation times of CT angiogram and CT perfusion in stroke. *Canadian Journal of Neurological Sciences*, 32(4):483–486, 2005.

[198] M. T. Stib, J. Vasquez, M. P. Dong, Y. H. Kim, S. S. Subzwari, H. J. Triedman, A. Wang, H.-L. C. Wang, A. D. Yao, M. Jayaraman, J. L. Boxerman, C. Eickhoff, U. Cetintemel, G. L. Baird, and R. A. McTaggart. Detecting Large Vessel Occlusion at Multiphase CT Angiography by Using a Deep Convolutional Neural Network. *Radiology*, 297(3):640–649, 12 2020.

[199] M. M. Sumer, I. Ozdemir, and N. Tascilar. Predictors of outcome after acute ischemic stroke. *Acta Neurologica Scandinavica*, 107(4):276–280, 4 2003.

[200] A. A. Taha and A. Hanbury. An Efficient Algorithm for Calculating the Exact Hausdorff Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(11):2153–2163, 11 2015.

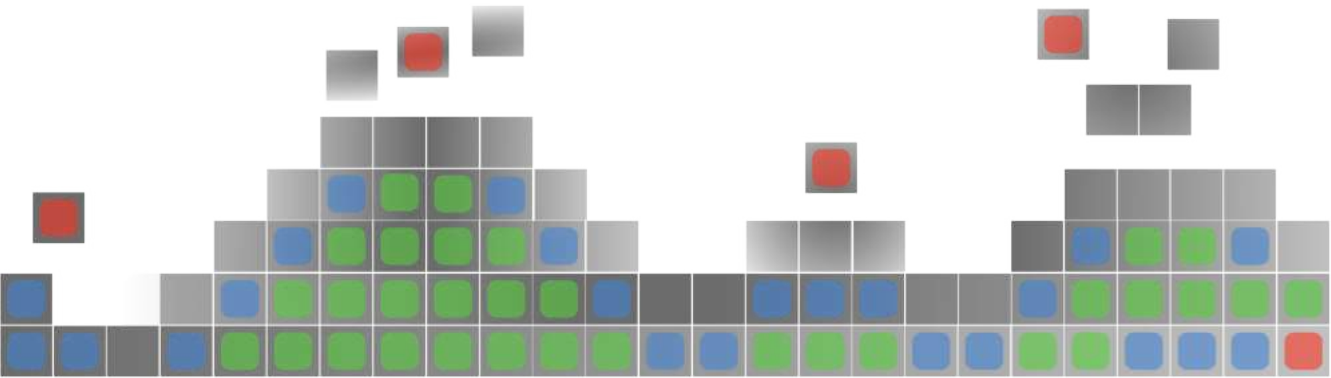
[201] M. L. Tolhuisen, E. Ponomareva, A. M. M. Boers, I. G. H. Jansen, M. S. Koopman, R. Sales Barros, O. A. Berkhemer, W. H. van Zwam, A. van der Lugt, C. B. L. M. Majoie, and H. A. Marquering. A Convolutional Neural Network for Anterior Intra-Arterial Thrombus Detection and Segmentation on Non-Contrast Computed Tomography of Patients with Acute Ischemic Stroke. *Applied Sciences*, 10(14):4861, 7 2020.



# Chapter 5

# Brain Segmentation in Patients with Perinatal Arterial Ischemic Stroke

**Based on:** R. Zoetmulder, L. Baak, N. Khalili, H. Marquering, N. Wagenaar, N. van der Aa, M. Benders, and I. Išgum (2022). "Brain- and -Hemisphere Tissue and Ischemic Lesion Segmentation for Perinatal Arterial Ischemic Stroke", submitted



# Abstract

**Background:** Perinatal arterial ischemic stroke (PAIS) is associated with adverse neurological outcomes. Quantification of brain development in these infants relies on labor-intensive manual assessment of brain tissues and ischemic lesions. Hence, we propose an automatic method exploiting convolutional neural networks (CNNs) to segment brain tissue classes and ischemic lesions in MRI scans of infants suffering from PAIS.

**Materials and Methods:** This single-center retrospective study included 115 patients with PAIS that underwent MRI after the stroke onset (baseline) and after three months (follow-up). Nine baseline and 12 follow-up MRI scans were manually annotated to provide reference segmentations (white matter, gray matter, basal ganglia and thalami, brainstem, ventricles, extra-ventricular cerebrospinal fluid, and cerebellum, and additionally on the baseline scans the ischemic lesions). Two CNNs were trained to perform automatic segmentation in the baseline and follow-up MRIs. Automatic segmentations were quantitatively evaluated using the Dice coefficient (DC) and the mean surface distance (MSD). Agreement between volumes obtained by automatic and reference segmentations was computed. Moreover, the scan quality and automatic segmentations were qualitatively evaluated in a larger set of MRIs without manual annotation by two experts. Agreement between the method and the raters and between the raters was established.

**Results:** Brain tissues segmentation led to the average DC and MSD between 0.78-0.92 and 0.18-1.1 mm for baseline, and between 0.88-0.95 and 0.10-0.58 mm for follow-up scans, respectively. For the ischemic lesions at baseline these measurements were on average 1.23 and 2.2 mm. Volumetric measurements indicate limited oversegmentation for the extra-ventricular cerebrospinal fluid and the ischemic lesions, and undersegmentation of the ischemic lesions of the automatic method in the baseline and follow-up scans. In scans without imaging artifacts, brain tissue segmentation was graded as excellent in more than 85% and 91% of cases, respectively for the baseline and follow-up scans. For the ischemic lesions at baseline, this was in 61% of cases.

**Conclusions:** The results show that automatic segmentation of brain tissue classes and ischemic lesions in MRI scans of patients with PAIS is feasible. The method may allow evaluation of the brain development and efficacy of treatment in large datasets.

## 5.1 Introduction

PAIS has an incidence rate of 1±5000 live births [205; 208; 215; 221; 229] and is associated with adverse motor and cognitive outcomes [216; 226]. Infants suffering from PAIS often present with hemi-convulsions and subsequently undergo neonatal brain MRI to diagnose PAIS (baseline). A follow-up scan, acquired weeks or months later, allows evaluation of residual damage and may improve prediction of outcome (follow-up). Currently, stroke size and location on the neonatal MRI scan and the effect of the stroke on brain development are assessed by qualitative visual evaluation of the MRI. However, qualitative evaluation is subjective and prone to intra- and inter-observer variability. Hence, quantitative evaluation would be preferred for the assessment of the ischemic lesion and the brain tissue classes affected and unaffected by stroke. Moreover, quantitative analysis would allow evaluation of the effects of neuro-regenerative interventions such as recombinant human erythropoietin (rhEPO) and mesenchymal stromal cells and might improve long-term outcome prediction [202–204]. For this quantitative analysis, accurate segmentation of the brain tissue classes and ischemic lesions in each hemisphere in baseline and follow-up MRI acquisitions is needed.



Given the complexity of the task, manual segmentation of brain tissue classes is practically infeasible in the clinical routine as well as in large studies. Hence, automatic segmentation would be required. In prior research, methods for automated segmentation of brain tissue classes in MRI scans of infants without large pathology have been developed [212; 213; 219]. However, stroke impacts appearance and the shape of the brain tissues and therefore, methods developed for analysis of the brain without substantial pathology are not directly applicable for segmentation of MRIs presenting stroke. In addition, after the stroke onset, ischemic areas are well visible on DWI and poorly visible on T2-weighted MRI that best shows different brain tissue classes. Therefore, several methods have been developed to segment ischemic lesions in neonatal MRI scans that analyzed one or both of these MRI sequences. For example, Murphy et al. developed a method to segment hypoxic-ischemic brain tissue in new-born infants with hypoxic-ischemic encephalopathy on

DWI [220]. This method classified ischemic voxels based on their spatial and intensity features with a random forest classifier. Ghosh et al. developed method to segment the ischemic lesion using DWI and apparent diffusion coefficient (ADC) maps [209]. Their method used hierarchical region splitting and symmetry based region growing to segment the ischemic lesion. Išgum et al. segmented the ischemic lesions using DWI and ADC maps. Initial segmentations were made by comparing the ADC maps of the patients to those of control subjects. Subsequently, these segmentations were refined by using spatial and texture features along with a linear discriminant classifier [211].

To the best of our knowledge, deep learning methods for ischemic stroke segmentation have been applied only to brain MRI scans in adults. Zhang et al proposed a 3D convolutional neural network (CNN) to segment ischemic stroke in adult DWI [231]. Thereafter, Praveen et al. used an auto-encoder in conjunction with a support vector machine to discriminate between normal and ischemic tissue on co-registered T1-weighted, T2-weighted, DWI and FLAIR MR scans [223]. All of the aforementioned methods segment brain tissue affected by ischemia but do not segment unaffected brain tissue classes. However, to quantify the treatment effect or improve outcome prediction after stroke, quantitative analysis of these tissues may be important as well. Given that PAIS typically affects one hemisphere, analysis of brain tissue classes per hemisphere would allow comparison and quantification of the stroke damage.

Hence, in this study, we propose an automatic method that uses a CNN to segment the brain tissue classes and the ischemic lesions per hemisphere in infants suffering from PAIS on baseline and follow-up MRI scans. Specifically, we segment the white matter (WM), gray matter (GM), basal ganglia and thalami (BGT), brainstem (BST), ventricular cerebro-spinal fluid (vCSF), the extra-ventricular cerebro-spinal fluid (eCSF), and cerebellum (CB), and ischemic tissue. Given the differences in the visibility of the stroke and differences in brain tissues caused by brain maturation in newborns, we train two age-specific instances of the network architecture: one for the segmentation of baseline (baseline network) and another for the follow-up (follow-up network) brain MRI scans. Brain tissues are best visible on T2-weighted MRI

scans while the early ischemic changes result in diffusion restriction, best visualized with DWI. Therefore, the baseline network analyses both DWI and T2-weighted MRI scans. On follow-up scans, the diffusion restriction on DWI is no longer present and tissue damaged by the diffusion restriction is often filled with eCSF. Hence, only the T2-weighted scan is analyzed by our follow-up network.

## 5.2 Materials & Methods

### 5.2.1 Materials

The study includes 115 newborn infants admitted to the Neonatal Intensive Care Unit, University Medical Center Utrecht, the Netherlands that suffered from PAIS, which was confirmed on MRI. The patient characteristics are shown in Table 5.1. Informed consent was obtained from the parents verbally to perform an MRI.



Patients underwent baseline MRI of the brain usually within one week after birth. Three scanners were used, a 1.5T Phillips Achieva, a 3T Phillips Achieva, and a 3T Phillips Ingenia Elition X scanner. All scans were acquired in the axial plane. The image acquisition parameters are listed in Table 5.2. Two to three months after the baseline scan, a follow-up MRI scan was made. Patients were scanned with a 1.5T or a 3T Phillips Achieva scanner. The image acquisition parameters are shown in Table 5.2.

### 5.2.2 Reference annotations

To train and evaluate the automatic segmentation, manual expert annotations of brain tissue classes and ischemic lesions were performed in 9 baseline and in 12 follow-up MRI scans. These scans contained no image artifacts. Manual annotations of brain tissue classes (WM, GM, BGT, BST, vCSF, eCSF, CB) were made in T2-weighted MRI at both baseline and follow-up following the definition described in [210]. Manual annotation of the ischemic lesions was made using the T2-weighted and DWI MRI available at baseline. Given the

| Parameter | Baseline (N=115) | Follow-up (N=97) |
|---|---|---|
| Clinical | | |
| Birthweight, grams, Median (IQR) | 3360 (596) | 3390(619) |
| Sex, F, No. [%] | 50 [43.5] | 46[42.2] |
| Seizures, No. [%] | 85[74] | 78[72] |
| Postnatal day at first seizure, Median(IQR)[N] | 1(2)[60] | 1(2)[54] |
| Postnatal day at MRI, Median(IQR) | 4(2) | 91(25) |
| Side of Ischemic Lesion | | |
| Right, No. [%] | 44/115[38] | 40/109[37] |
| Left, No. [%] | 65/115[56] | 62/109[57] |
| Bilateral, No. [%] | 6/115[5] | 7/109[6] |
| Vascular Territory | | |
| Main MCA, No. [%] | 21/115[18] | 21/109[19] |
| Anterior MCA branch, No. [%] | 12/115[10] | 9/109[9] |
| Middle MCA branch, No. [%] | 16/115[14] | 16/109[15] |
| Posterior MCA branch, No. [%] | 24/115[21] | 22/109[20] |
| Cortical MCA branch, No. [%] | 18/115[16] | 16/109[15] |
| Perforator branch, No. [%] | 17/115[15] | 17/109[16] |
| PCA, No. [%] | 5/115[4] | 6/109[6] |
| ACA, No. [%] | 2/115[2] | 2/109[2] |

**Table 5.1:** Patient characteristics, ischemic lesion hemisphere, and vascular territory of the scans acquired at baseline and follow-up.

| Scan Time | Sequence | Matrix Size | Voxel Dimensions | Field Strength | TE/TR | N |
|---|---|---|---|---|---|---|
| Baseline | T2w | 512 x 512 | 0.35 x 0.35 x 2.0 mm | 3T | 120/6629 | 89 |
| | | 512 x 512 | 0.45 x 0.45 x 2.0 mm | 3T | 120/6277 | 1 |
| | | 256 x 256 | 0.7 x 0.7 x 2.0 mm | 1.5T | 50/7670 | 24 |
| | | 384 x 384 | 0.34 x0.34 x 2.0 mm | 3T | 120/6276 | 1 |
| | DWI | 224 x 224 | 0.8 x 0.8 x 3.0 mm | 3T | 114/5453 | 69 |
| | | 256 x 256 | 0.7 x 0.7 x 4.0 mm | 1.5T | 89/4000 | 26 |
| | | 256 x 256 | 0.9 x 0.9 x 3.0 mm | 3T | 70/2452 | 19 |
| | | 224 x 224 | 0.8 x 0.8 x 3.3 mm | 3T | 118/5695 | 1* |
| Follow-Up | T2w | 512 x 512 | 0.35 x 0.35 x 2.0 mm | 1.5T | 50/7670 | 83 |
| | | 256 x 256 | 0.7 x 0.7 x 2.0 mm | 3T | 120/6629 | 26 |

**Table 5.2:** Image acquisition parameters for the scans. *scan made on Phillips Ingenia Elition X, rather than Phillips Achieva scanner.

very extensive workload of creating the manual annotations, brain tissue segmentations were pre-segmented automatically using the algorithm described in [218] and subsequently manually corrected by one of two trained experts (LB and NW). The corrections were made using in-house developed software [212]. Subsequently, to divide the brain into two hemispheres a reference midline of the brain was annotated for both the baseline and follow-up T2-weighted scans in ImageJ [225] that was subsequently used to separate the segmentation of the brain tissue classes into tissue classes per hemisphere resulting in a total of 14 brain tissue classes. At baseline the ischemic lesion was segmented in addition to the brain tissue. For this, the DWI (B800 or B1000) was registered to the T2-weighted MRI. Hence, the baseline and follow-up segmentation consisted of 16 and 14 classes, respectively. Examples of reference segmentations of the baseline and follow-up scans are shown in Figure 5.1.



### 5.2.3 Method

Our method consisted of two main stages. In the preprocessing stage, T2-weighted and DWI scans were normalized, resampled and aligned to allow joint analysis. Thereafter, convolutional neural networks were used to segment brain tissue classes per hemisphere on baseline and follow-up T2-weighted MRI scans, and the ischemic lesions in the baseline using T2-weighted and DWI scans.

### 5.2.4 Preprocessing

To focus the analysis on the brain only, prior to segmentation, the T2-weighted scans were skull stripped [228]. Given that ischemic region is best visible on DWI scans and brain tissue classes on the T2-weighted scans, the DWI and T2-weighted scan were aligned by rigid and thereafter by deformable registration. For this, the DWI was resampled to the spatial resolution of the T2-weighted scan. Registration was performed using SimpleITK. The Mattes mutual information with 64 bins was used as the loss function for both types of registration. For the rigid registration the gradient descent optimizer was used with a learning rate of 1. Optimization was performed for 150 itera-

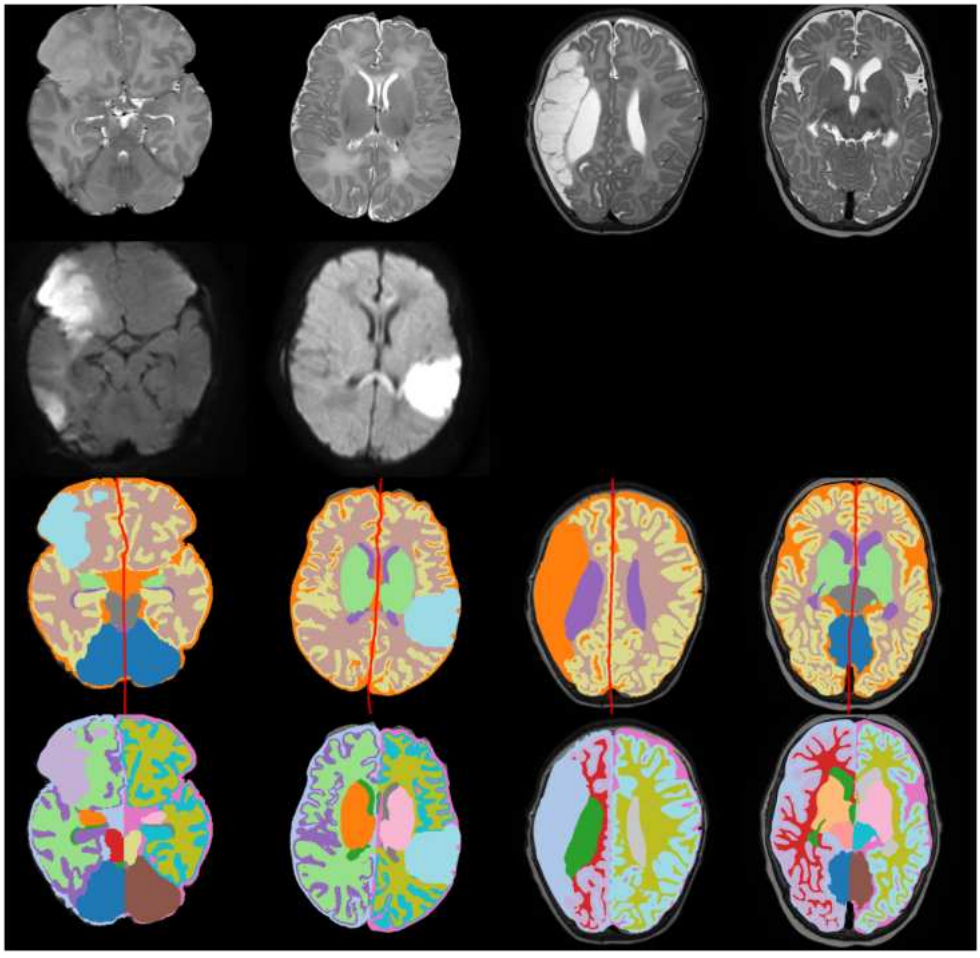


**Figure 5.1:** Examples of reference segmentations that are used during training and testing. From left to right: the first two columns show baseline scans and the third and fourth columns show follow-up scans. From top to bottom: The first row shows a axial slices from the T2-weighted scan of 4 different patients. The second row shows the DWI (only for the baseline) of the same patients. The third row shows the T2-weighted scan overlaid with the brain tissue segmentation and the midline, and the fourth row shows the scan overlaid with the reference segmentation for each hemisphere.

tions. For the deformable registration, the space between the control points was set to 25 mm and an L-BFGS optimizer[217] was applied for 100 iterations.

The intensities of MRI voxels vary per acquisition protocol and scanner. Hence, all voxel intensities were clamped between zero and the 99th percentile of the voxel intensity values in each scan. The voxel values were subsequently normalized between -1 and 1, and each slice was resampled to 512 x 512 voxels.

### 5.2.5 Automatic Segmentation

Currently, UNet-like architectures have been shown to be most successful at segmentation of medical images [224]. Their success is attributed to the combination of coarse features from the down-sampling path that are combined with fine-grained features from the up-sampling path, via the use of skip-connections. For our research, we have chosen UNet++ (Figure 5.2) architecture because it uses deep supervision to co-train an ensemble of sub-networks of varying depth within the architecture. Prior literature has shown that this outperforms the standard UNet architecture [232]. Each node in UNet++ consisted of two 3x3 convolutions, followed by batch normalization and the ReLU activation function. The features were down-sampled by max-pooling with a stride of two, up-sampled by a transposed convolution with a stride of two or passed to the next layer. Each node was connected to all previous nodes in the same row via skip-connections. To successfully co-train the nested ensembles that UNet++ consists of, deep supervision was applied to the three nodes before the output node. Before the deep supervision loss functions were applied, a one by one convolution was used to reduce the number of feature maps to one. The final loss was calculated by taking a weighted average between the deep supervision losses and output loss. The output loss had a weight that was seven times greater than that of each individual deep supervision losses. In total, the baseline and follow-up network respectively segmented 14 and 16 classes. During training, the weighted cross entropy loss function was used. To remedy the class imbalance that was caused by the differing number of voxels per class, the tissue classes were assigned a weight ten times greater than the background class. Multiscale analysis has been shown to improve segmentation performance [232]. To allow multiscale

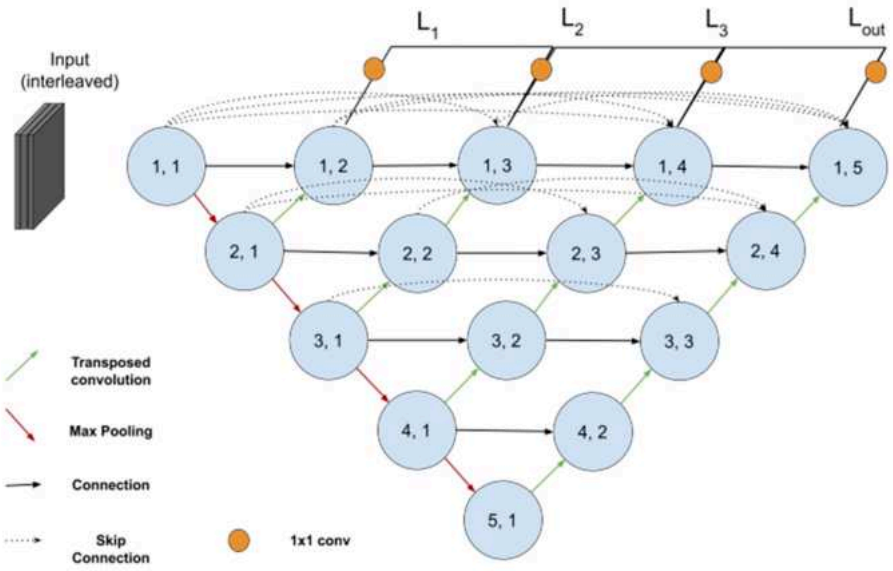





**Figure 5.2:** The UNet++ architecture used to segment the brain tissues and ischemic lesion for each brain hemisphere. The input consisted of three consecutive axial slices from each MRI sequence. The baseline DWI and T2-weighted MRI scans were interleaved. For the baseline scans, each T2-weighted slice was concatenated to the anatomically corresponding DWI slice. Each node outside of the down-sampling path was connected to all the nodes that preceded it horizontally via skip-connections and to the node that was one horizontal and diagonal index lower. Deep supervision was applied to all the nodes in the top horizontal layer, other than the input node.

analysis during inference, the output of each deep supervision layer was averaged to create the final prediction. The UNet++ was implemented using Pytorch 1.5.1 [222].

The UNet++ was randomly initialized and trained on stacks of three consecutive axial slices. The segmentation label was predicted for the axial slice at the center of the stack. The slices adjacent to the slice of interest were provided as additional spatial context. To allow analysis using the information from the baseline T2-weighted and DWI images, each axial slice of the T2-weighted scan was followed by its anatomically corresponding axial slice of the DWI scan.

Given that voxel-wise segmentation is applied, the segmentation result may contain small isolated clusters of voxels which would not be physiologically possible. Therefore, the tissue classes within the largest connected component of the binarized segmentation map were retained. Similarly, to ensure that no gaps exist due to an incorrect hemisphere segmentation, morphological closing with a 2D diamond shaped structuring element with a connectivity of 1 and a rank of 2 was applied to each hemisphere segmentation.

### 5.2.6 Evaluation

Quantitative evaluation of the brain tissue classes and the ischemic lesion segmentation per hemisphere was performed in scans with manual reference annotations. To evaluate the overlap between the reference and automatic segmentation the Dice coefficient was computed. To evaluate segmentation agreement along the tissue boundary, the mean surface distance (MSD) between the automatic and reference segmentation was calculated. Moreover, the bias and limits of agreement were calculated for the brain tissue and ischemic volumes obtained from automatic and reference segmentations.



In scans without manual reference annotations, automatic segmentation was qualitatively evaluated. In addition, to assess whether image quality impacts automatic segmentation, presence of image corruption caused by imaging artifacts was rated (None, Mild, Moderate, Severe) as listed in Table 5.3. Images graded as severely affected by artefacts (Severe) were excluded from the qualitative evaluation of the automatic segmentation. The automatic segmentation was evaluated qualitatively by two expert observers, NW (Observer 1) and LB (Observer 2), on a three-point scale (Excellent, Moderate, Poor) as listed in Table 5.3. Additionally, in the follow-up images the quality of the segmentation of the brain tissues surrounding the location of the former ischemic lesion was also rated on the same three-point scale. However, if no tissue damage was visible, surrounding tissue was rated as Invisible.

The presented automatic method segments brain tissues and ischemic lesions per hemisphere. However, rating automatic segmentations in 16 and 14 classes in each baseline and follow-up MRIs respectively is an extremely

| Scale | |
|---|---|
| *Excellent* | Segmentation clinically useable after no or minor manual corrections |
| *Moderate* | Segmentation clinically useable after major manual corrections |
| *Poor* | Segmentation clinically unuseable |

**Table 5.3:** Grading the quality of brain tissue segmentation, ischemic lesion segmentation, hemisphere separation, and quality of the segmentation surrounding the former location of the ischemic lesion.

| Rating | |
|---|---|
| *None* | No artefacts visible in the scan |
| *Mild* | Differentiation between tissue classes is clear despite mild artefacts being present |
| *Moderate* | Differentiation between tissue classes is visible, but with limited uncertainty |
| *Severe* | Differentiation between tissue classes is not visible due to artefacts |

**Table 5.4:** Grading of the degree to which an image is corrupted by image artefacts.



time consuming task, especially in a larger set of images. Therefore, to make the qualitative evaluation feasible, rating was performed per tissue class as described above and hemisphere separation was graded separately for the whole brain. The intra- and inter-rater agreement of the qualitative evaluation of the scan quality and automatic segmentation were evaluated by the accuracy between the raters.

## 5.3 Experiments and Results

### 5.3.1 Experimental Setup

Given that scans of 9 patients made at baseline were manually annotated and available for training and quantitative evaluation, automatic segmentation was performed in nine-fold cross-validation experiments. This means that in each experiment, 8 scans were available for training and 1 scan for testing. In case of the follow-up data, scans of 12 patients were manually annotated. Hence, automatic segmentation was performed in six-fold cross-validation experiments, where every time 10 training scans and 2 test scans were utilized.

| | Baseline | | | | Follow-up | | | |
|---|---|---|---|---|---|---|---|---|
| | Left | | Right | | Left | | Right | |
| | *Dice* | *MSD* | *Dice* | *MSD* | *Dice* | *MSD* | *Dice* | *MSD* |
| **WM** | 0.90 | 0.22 | 0.91 | 0.18 | 0.90 | 0.14 | 0.90 | 0.16 |
| **GM** | 0.81 | 0.18 | 0.82 | 0.15 | 0.92 | 0.10 | 0.91 | 0.12 |
| **BGT** | 0.90 | 0.63 | 0.84 | 1.08 | 0.94 | 0.34 | 0.93 | 0.31 |
| **BST** | 0.88 | 0.50 | 0.85 | 0.45 | 0.89 | 0.33 | 0.88 | 0.34 |
| **vCSF** | 0.83 | 0.30 | 0.78 | 0.38 | 0.90 | 0.56 | 0.90 | 0.25 |
| **eCSF** | 0.80 | 0.29 | 0.80 | 0.45 | 0.88 | 0.20 | 0.88 | 0.23 |
| **CB** | 0.91 | 0.41 | 0.92 | 0.44 | 0.95 | 0.29 | 0.95 | 0.58 |
| **IL** | 0.72 | 2.18 | 0.86 | 1.23 | - | - | - | - |
| **Mean** | 0.84 | 0.59 | 0.85 | 0.35 | 0.91 | 0.28 | 0.91 | 0.28 |

**Table 5.5:** Median Dice coefficient (Dice) and mean surface distance (MSD) in mm between the reference and automatic segmentation. Results are displayed for each brain tissue type in the left and right hemisphere: white matter (WM), gray matter (GM), basal ganglia and thalamus (BGT), brainstem (BST), the ventricular cerebrospinal fluid(vCSF), extra-ventricular cerebrospinal fluid (eCSF), and the cerebellum (CB). In addition, the ischemic lesion is reported for the baseline scans. The mean Dice coefficient over all brain tissue types and the IL is listed. Among the baseline scans four had an ischemic lesion in the left hemisphere and five in the right hemisphere.



To optimize the networks, an Adam optimizer [214], a weight decay of 5e-5 and a batch size of 16 were used. The UNet++ was optimized for 300 epochs using a cyclical learning rate schedule [227] with a cycle length of 1,000 iterations. The minimum learning rate was 1e-4 and the maximum learning rate was 1.1e-3.

Automatic segmentation including preprocessing and segmentation took under 4 minutes per scan.

### 5.3.2 Quantitative Evaluation

Quantitative results for the baseline and follow-up scans are listed in Table 5.5. The results show that segmentation performance for the brain tissue classes on the baseline scans was similar for the left and the right hemisphere, except for the right BGT. Visual analysis of the segmentations revealed that in several cases ischemic lesions were erroneously segmented as BGT likely causing to the asymmetry in the segmentation of BGT. On the follow-up scans, the per-

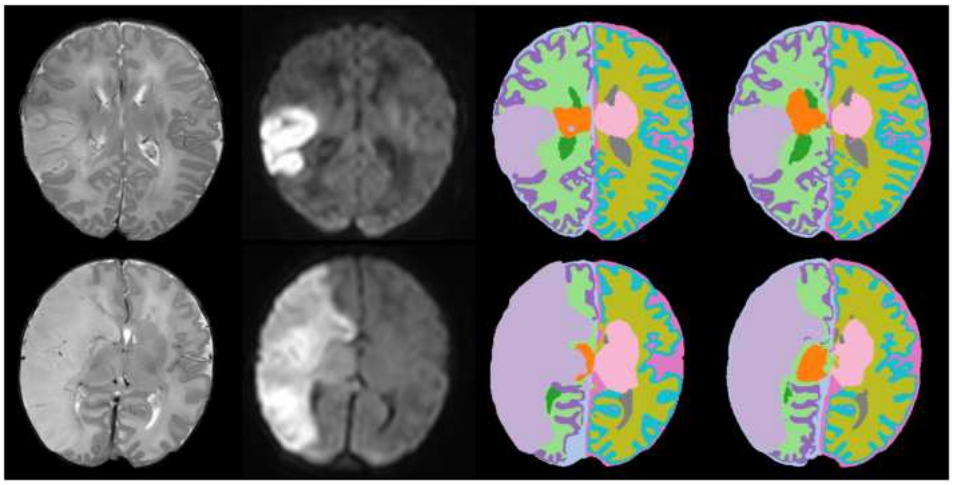


**Figure 5.3:** Example of brain tissue and ischemic lesion segmentations on the baseline scans for two patients (top and bottom row). From left to right: T2-weighted scan, DWI scan, reference segmentation and automatic segmentation. The top row shows an example on which the network successfully segmented the hemispheres and the brain tissues. The bottom row shows a DWI scan on which the ischemic lesion is undersegmented and the basal ganglia are oversegmented by the automatic method.



formance of the segmentation was similar in the left and the right hemisphere for all tissue classes. Examples of segmentations for baseline and follow-up scans are respectively shown in Figure 5.3 and Figure 5.4.

The median and inter-quartile range of the brain tissue and ischemic lesion volumes obtained from the reference segmentations in the left and right hemispheres are shown in Figure 5.5 for both the baseline and the follow-up scans. The bias and limits of agreement between the reference and automatically quantified volumes for both the baseline and follow-up scan are shown in Figure 5.6. For the baseline scans, eCSF volumes in both hemispheres had a bias greater than the other tissue classes indicating limited oversegmentation. The left and right ischemic lesion respectively had a bias greater and smaller than zero, indicating limited over -and under-segmentation. On the follow-up scans the eCSF in both hemispheres had a greater bias than the other tissue classes indicating oversegmentation.

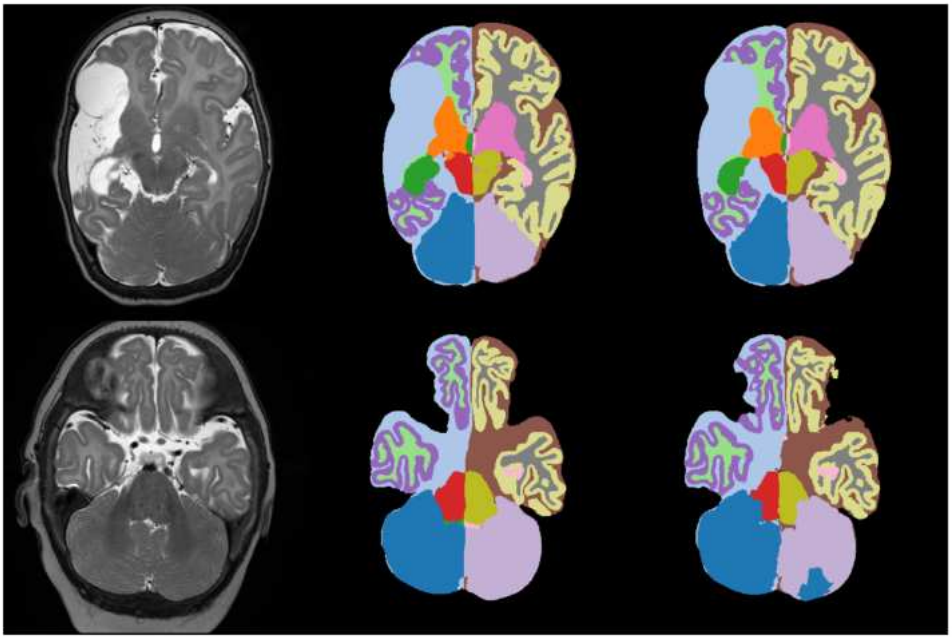


**Figure 5.4:** Example of brain tissue segmentations on follow-up scans for two patients (top and bottom row). From left to right: T2-weighted scan, reference segmentation and automatic segmentation. The top row shows a scan which was accurately segmented. The bottom row shows a scan on which the right cerebellum is incorrectly segmented.

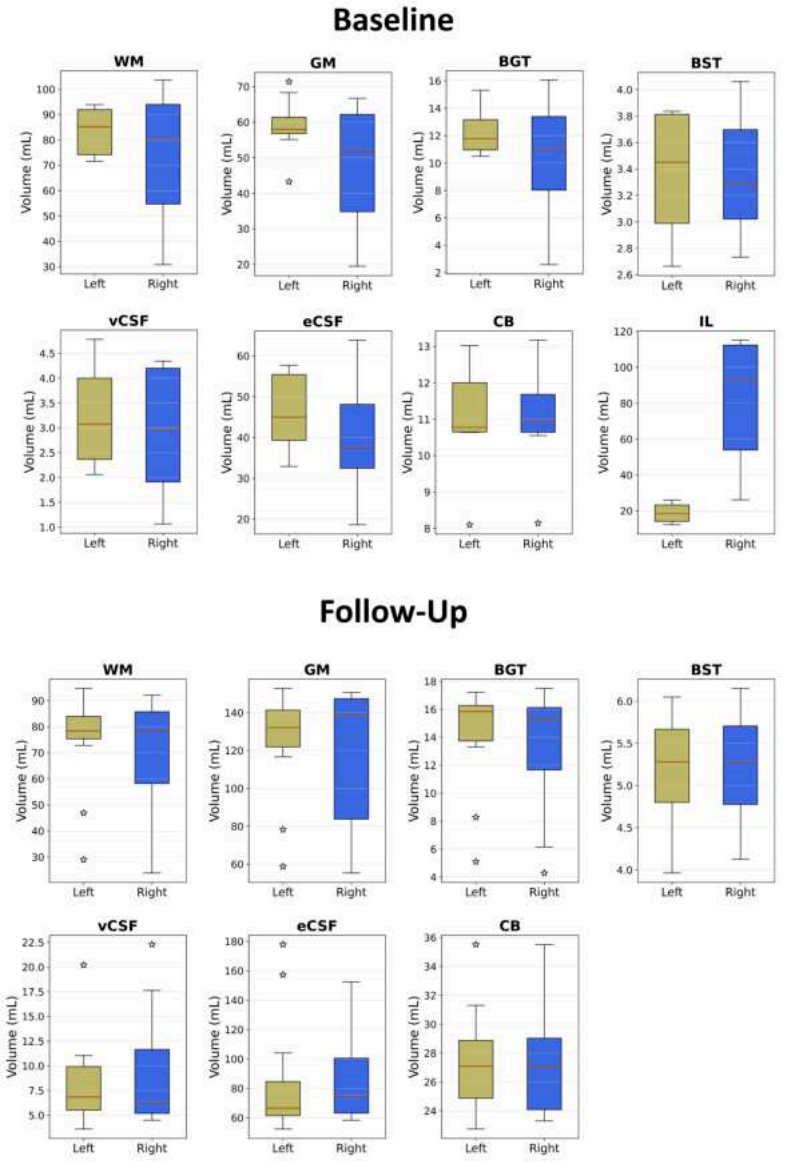



**Figure 5.5:** Boxplots showing the median and interquartile range (IQR) of the reference brain tissue volumes in mL in the baseline(n=9) and follow-up scans(n=12). Results are displayed for the left and right hemisphere in each plot: white matter (WM), gray matter (GM), basal ganglia and thalamus (BGT), brainstem (BST), the ventricular cerebro-spinal fluid (vCSF), extra-ventricular cerebrospinal fluid (eCSF), and the cerebellum (CB). In addition, the ischemic lesion (IL) is reported for the baseline images. Among the baseline scans four had ischemia in the left and five in the right hemisphere. Outliers on the follow-up scans are due to cerebrospinal fluid replacing brain tissue.

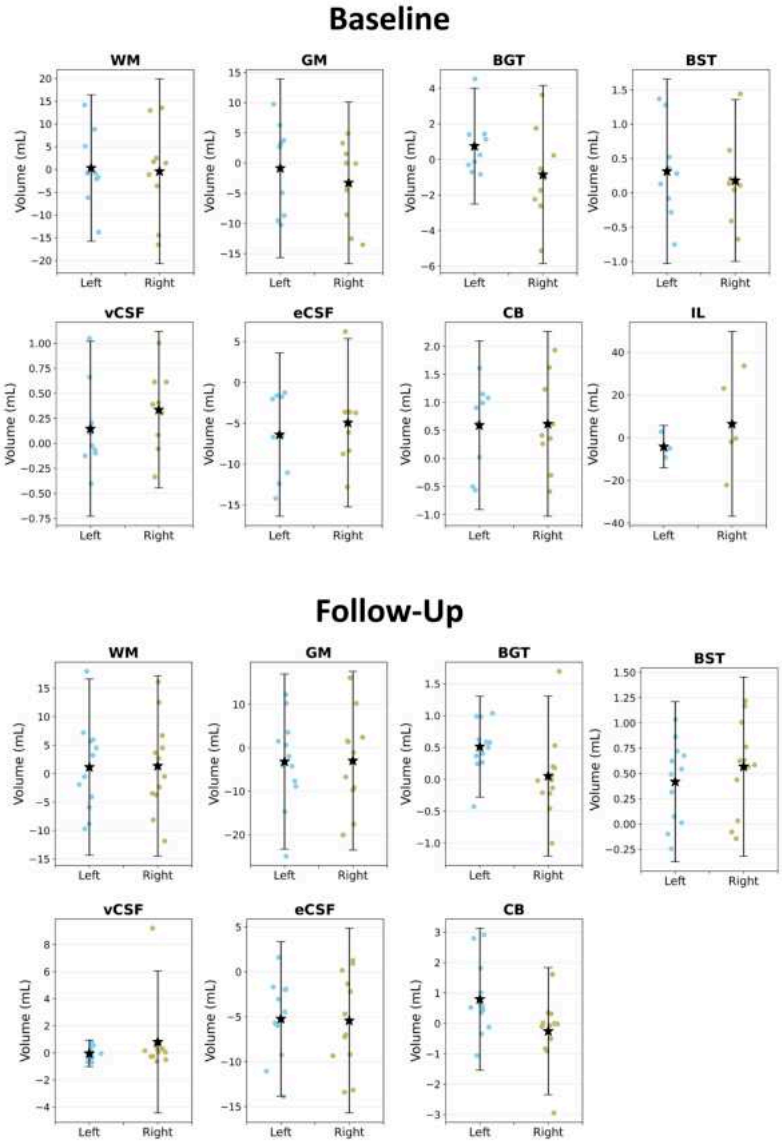



**Figure 5.6:** Bias and limits of agreement between the volumes obtained from the reference and automatic segmentation. Results are displayed for each brain tissue class in the left and right hemisphere: white matter (WM), gray matter (GM), basal ganglia and thalamus (BGT), brainstem (BST), the ventricular cerebro-spinal fluid(vCSF), extra-ventricular cerebrospinal fluid (eCSF), and the cerebellum (CB). In addition, results of the volumetric measurements for the ischemic lesion (IL) are reported for the baseline images. Among the baseline scans four had an infarct in the left and five in the right hemisphere.

### 5.3.3 Qualitative Evaluation

The ratings of the baseline image quality with respect to presence of imaging artifacts and automatic segmentation performance by Observer 1 are shown in Figure 5.7. In total, 74, 19, 8 and 5 scans were rated as having None , Mild, Moderate or Severe artifacts.

Specifically, the results show that the segmentation of brain tissue classes is accurate on images that have no visible artefacts, i.e. with the majority of the ratings being Excellent (useable after no or minor correction). Hemisphere separation (HS) is also accurate with most segmentations being rated as Excellent. The results also show that the segmentation of the ischemic lesion is the most challenging, with 61% percent of ischemic lesion segmentations on images without artefacts being rated as Excellent and 22.6% being rated as Moderate.



The ratings of the follow-up image quality with respect to presence of imaging artifacts and automatic segmentation performance by Observer 2 are shown in Figure 5.8. In total, Observer 2 graded 66, 18, 13 and 0 images as having None, Mild, Moderate or Severe imaging artifacts, respectively.

The segmentation of the brain tissues was mostly rated as Excellent in images without artefacts. Furthermore, on scans with no artefacts, 65% and 100% percent of cases the tissue surrounding former ischemia and hemisphere separation, respectively, were rated as Excellent.

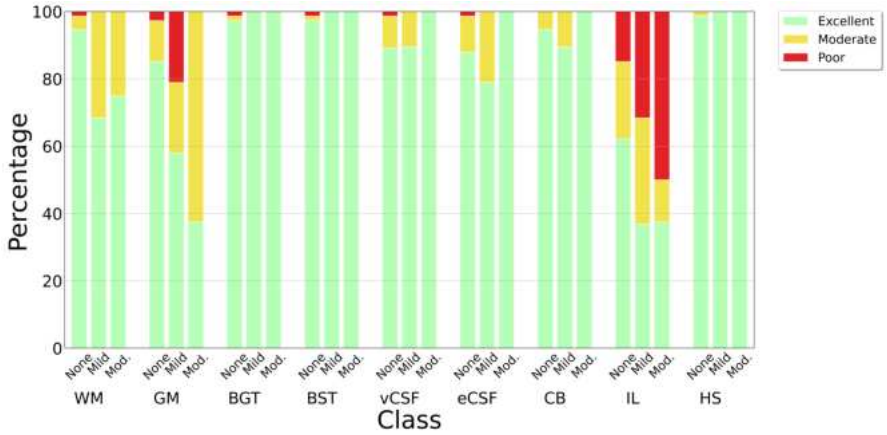



**Figure 5.7:** Qualitative evaluation of the automatic segmentations for each brain tissue class: white matter (WM), gray matter (GM), basal ganglia and thalamus (BGT), brainstem (BST), the ventricular cerebrospinal fluid (vCSF), extra-ventricular cerebrospinal fluid (eCSF), and the cerebellum (CB),the ischemic lesion (IL) and, the hemisphere separation (HS) on the baseline scans by Observer 1. The ratings of each tissue type are divided by the severity of the artefacts that afflicted them (None, Mild and Moderate). Note that scans with severe artifacts were considered unsuitable for automatic analysis and were therefore not rated. Results are given in percentages.



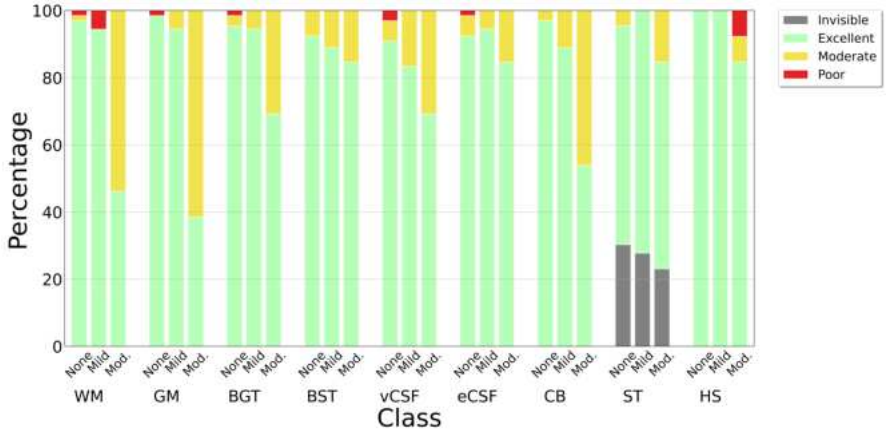



**Figure 5.8:** Qualitative evaluation of the automatic segmentations for each brain tissue class: white matter (WM), gray matter (GM), basal ganglia and thalamus (BGT), brainstem (BST), the ventricular cerebro-spinal fluid(vCSF), extra-ventricular cerebrospinal fluid (eCSF), and the cerebellum (CB),the surrounding tissue and, the hemisphere separation (HS) on the baseline scans by Observer 2. The ratings of each tissue type are divided by the severity of the artefacts that afflicted them (None, Mild and Moderate). Note that scans with severe artifacts were considered unsuitable for automatic analysis and were therefore not rated. Results are given in percentages.

| | Rater | CB | eCSF | vCSF | BST | BGT | WM | GM | IL | ST | HS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Baseline | Intrarater | 96.3 | 88.9 | 92.3 | 96.3 | 100.0 | 96.3 | 84.6 | 80.0 | | 96.3 |
| | Interrater | 93.1 | 88.9 | 92.9 | 96.4 | 96.6 | 73.1 | 84.0 | 89.5 | | 100.0 |
| follow-up | Interrater | 93.3 | 93.3 | 86.2 | 93.3 | 93.3 | 96.7 | 90.0 | - | 83.3 | 100 |

**Table 5.6:** For the baseline and follow-up scans the accuracy between the ratings of Observer 1 and Observer 2 are shown on a randomly selected subset of 30 scans. The classes for which the accuracy is calculated are: white matter (WM), gray matter (GM), basal ganglia and thalamus (BGT), brainstem (BST), the ventricular cerebrospinal fluid (vCSF), extra-ventricular cerebrospinal fluid (eCSF), the cerebellum (CB), the area surrounding the formerly ischemic lesion (ST), the ischemic lesion (IL), and the hemisphere separation (HS). For the baseline scans, accuracy between two ratings of the same observer (Observer 1) is additionally shown.

The inter- and intrarater accuracy was calculated on a subset of 30 randomly selected baseline scans. The interrater accuracy was also calculated on a subset of 30 randomly selected follow-up scans. The inter- and intrarater accuracy for the brain tissue segmentation, ischemic lesion segmentation, tissue surrounding the formerly ischemic region, and hemisphere separation are shown in Table 5.6.



## 5.4 Discussion

We have presented a deep learning method for segmentation of brain tissue classes and the ischemic lesions in both hemispheres of brain MRI scans in infants with PAIS. The segmentation was applied to scans acquired at two time points, i.e. to scans made after the onset of stroke and to scans made at 3-months follow-up. Quantitative evaluation of the automatic segmentation showed that on average the brain tissue classes and the ischemic lesions had good spatial and volumetric agreement with the manual expert segmentations. Furthermore, the qualitative analysis on a larger set of MRIs showed that the automatic brain tissue segmentations obtained on the baseline and follow-up scans were clinically useable after no or minor correction. However, the segmentation of the ischemic lesions on the baseline scans was more challenging more often requiring manual correction. Segmentation of the ischemic lesions was especially compromised in scans corrupted by artefacts. Hence, additional scrutiny of the segmentations should be applied when us-

ing scans corrupted by artefacts.

Several aspects may have contributed to compromised segmentation results. First, contrast between the ischemic lesions and the unaffected regions on DWI varied greatly between patients. This may have been due to differences in the timing of MRI. After a peak in signal intensity on DWI, the contrast slowly normalizes around day 7, so-called pseudonormalization. This diversity in contrast differences may not have been represented in our training data. Second, small false positive segmentations occurred for the ischemic lesion segmentation, but rarely for the brain tissue classes. While challenging for the automatic segmentation, this type of error is easily manually corrected. Third, ischemic lesions in the left hemisphere were often oversegmented. Although these false positive segmentations and oversegmentation may lead to an inaccurate estimation of ischemic volume, it takes little effort to manually correct them.



Previous automatic segmentation methods reported accurate segmentations, but neither differentiated brain tissues per hemisphere. The study by Moeskops et al. segmented the same brain tissue classes and unmyelinated white matter in preterm infants imaged at term-equivalent age. Given nearly the same tissue classes definition and age of the infants, this set enables comparison with performance on the baseline scans in the current study [219]. Despite the presence of pathology in the images in our study and more complex per hemisphere analysis, we achieved comparable performance on the segmentation of the baseline scans (Average Dice coefficients reported in [219] vs ours: WM 0.92 vs 0.91, GM 0.88 vs 0.82, BGT 0.91 vs 0.87, BST 0.84 vs 0.87, vCSF 0.81 vs 0.81, eCSF 0.84 vs 0.80, CB 0.93 vs 0.92 ). The largest differences are observed in the tissue classes that are directly adjacent to ischemia in our set such as GM and CSF. Note that the differences in patient populations between these two studies preclude direct performance comparison and only provide an indication.The study by Ding et al. segmented three brain tissue classes (GM, WM, CSF) [207]. Furthermore, this study used more data for training and their grey matter class incorporated brain tissues other than cortical grey matter, such as the brainstem, cerebellum, amygdala, and

hippocampus. Due to the differences in tissue definitions and population of patients quantitative results are not comparable.

Our method is the first deep-learning-based method that segments ischemic lesions on DWI scans in perinatal ischemic stroke patients. Other work by Igum et al. used a linear discriminant classifier to segment the ischemic lesion. However, their method did not segment other brain tissue classes or indicated the hemisphere in which the lesion was located. Unfortunately, this work did not report overlap or boundary metrics [211].

The most comparable deep-learning-based methods for stroke lesion segmentation problem are those that segment ischemic lesions on DWI scans in adult patients [206; 230; 231]. We observe that our method achieves performance in range of automated stroke lesion segmentation methods for adults (Dice coefficient: 0.79 in our study vs. 0.79 [231], 0.67 [206], 0.85 for a single U-Net [230]) despite our network being trained on two orders of magnitude fewer data. However, none of the methods that were developed for scans of adult patients additionally provide brain-tissue segmentations, nor do they indicate which hemisphere is afflicted by the ischemia.



The automatic analysis of brain tissue volumes and the ischemic lesion in follow-up and baseline scans would allow neuro-regenerative treatments to be evaluated. For example, treatment of PAIS by intranasal administration of mesenchymal stromal cells has recently been shown to be feasible and safe [202]. However, large-scale placebo controlled trials still need to be conducted. Given that our automatic method analyses a scan in less than 4 minutes, the method would significantly reduce the analysis burden and could facilitate large scale studies.

Despite the accurate segmentations provided by our method, our study has several limitations. First, our study used a limited number of training and quantitative testing data with manual reference annotations at the baseline (n = 9) and follow-up (n = 12). Furthermore, these scans did not contain imaging artefacts. Hence, we have not been able to quantitatively evaluate

performance on a large set of scans or scans that contain imaging artefacts. To ameliorate this limitation in our evaluation, we conducted a qualitative evaluation in a large set of MRIs that showed that most brain tissue class segmentations could be used without any or after minor manual correction.

Second, our network instances were trained on scans acquired under very limited circumstances. For example, the scans were acquired on two types of MRI scanners; A 1.5T and a 3T Phillips Achieva scanner. By only having data from three scanners available to train on, the performance of our networks on data acquired by different scanner types from different manufacturers may not generalize. Furthermore, the scans were made at specific time points; the baseline scan was usually made the week after birth and the follow-up was made two to three months after the baseline. During the first year of life the brain develops rapidly. Hence, by training our networks only on images acquired at specific time points performance may not generalize to MRI made at other infant ages. Future research should include scans from various stages of the brain development and from a greater variety of scan types and scanner vendors.



Third, our results indicated that our method segmented the ischemic lesion in the right hemisphere more accurately than in the left hemisphere. This was likely because the ischemic lesions in the right hemisphere were larger than the ischemic lesions in the left hemisphere. Hence, these lesions were more difficult for the method instance to segment accurately. On a larger test set this bias would likely disappear.

Fourth, the intra- and interrater reliability were assessed by using the accuracy. This metric was used because the data was skewed due to the majority of the automatic brain tissue segmentations being rated as Excellent. However, caution must be applied when interpreting the accuracy alone because it does not correct for observations that occur due to chance alone.

To conclude, we presented a method for automated segmentation of brain tissue classes and ischemic lesions in each brain hemisphere. The automatic

segmentation method may allow evaluation of the brain development and efficacy of treatment methods in large datasets imaging infants affected by perinatal arterial ischemic stroke.

# Bibliography

[202] L. M. Baak, N. Wagenaar, N. E. van der Aa, F. Groenendaal, J. Dudink, M. L. Tataranno, U. Mahamuud, L. S. de Vries, F. van Bel, M. J. N L Benders, C. Development, R. K. Jellema, L. M. Baak, N. Wagenaar, N. E. van der Aa, F. Groenendaal, J. Dudink, M. Luisa Tataranno, U. Mahamuud, C. H. Verhage, R. M. J C Eijsermans, L. S. Smit, R. K. Jellema, T. R. de Haan, H. J. ter Horst, W. P. de Boode, S. J. Steggerda, H.-J. Prins, C. G. de Haar, L. S. de Vries, F. van Bel, C. J. Heijnen, C. H. Nijboer, and M. J. N L Benders. Feasibility and safety of intranasally administered mesenchymal stromal cells after perinatal arterial ischaemic stroke in the Netherlands (PASSIoN): a first-in-human, open-label intervention study. *The Lancet Neurology*, 21(6):528–536, 6 2022.

[203] S. Bava, S. L. Archibald, and D. A. Trauner. Brain structure in prenatal stroke: Quantitative magnetic resonance imaging (MRI) analysis. *Journal of Child Neurology*, 22(7):841–847, 2007.

[204] M. J. Benders, N. E. Van Der Aa, M. Roks, H. L. Van Straaten, I. Isgum, M. A. Viergever, F. Groenendaal, L. S. De Vries, and F. Van Bel. Feasibility and safety of erythropoietin for neuroprotection after perinatal arterial ischemic stroke. *The Journal of pediatrics*, 164(3), 2014.

[205] S. Chabrier, B. Husson, M. Dinomais, P. Landrieu, and S. Nguyen The Tich. New insights (and new interrogations) in perinatal arterial ischemic stroke. *Thrombosis research*, 127(1):13–22, 1 2011.

[206] L. Chen, P. Bentley, and D. Rueckert. Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. *NeuroImage: Clinical*, 15:633–643, 1 2017.

[207] Y. Ding, R. Acosta, V. Enguix, S. Suffren, J. Ortmann, D. Luck, J. Dolz, and G. A. Lodygensky. Using Deep Convolutional Neural Networks for Neonatal Brain Image Segmentation. *Frontiers in Neuroscience*, 14:207, 3 2020.

[208] C. Gale, Y. Statnikov, S. Jawad, S. N. Uthaya, N. Modi, N. Modi, E. Statnikov, N. Patel, S. Hegarty, M. Percival, K. Todd, M. Jolly, J. Walker,

T. Kelly, M. Upton, B. Harlev-Lam, K. Robbins, M. Knight, D. Odd, D. Siassakos, M. Magro, D. Peebles, N. Robertson, K. Luyt, P. Clarke, M. Forrester, J. Frohlich, A. Cameron, E. Petch, J. Dopran, T. Austin, J. P. Boardman, and F. Cowan. Neonatal brain injuries in England: population-based incidence derived from routinely recorded clinical data held in the National Neonatal Research Database. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, 103(4):F301, 10 2018.

[209] N. Ghosh, Y. Sun, B. Bhanu, S. Ashwal, and A. Obenaus. Automated detection of brain abnormalities in neonatal hypoxia ischemic injury from MR images. *Medical Image Analysis*, 18(7):1059–1069, 10 2014.

[210] I. Išgum, M. J. Benders, B. Avants, M. J. Cardoso, S. J. Counsell, E. F. Gomez, L. Gui, P. S. Huppi, K. J. Kersbergen, A. Makropoulos, A. Melbourne, P. Moeskops, C. P. Mol, M. Kuklisova-Murgasova, D. Rueckert, J. A. Schnabel, V. Srhoj-Egekher, J. Wu, S. Wang, L. S. de Vries, and M. A. Viergever. Evaluation of automatic neonatal brain segmentation algorithms: The NeoBrainS12 challenge. *Medical Image Analysis*, 20(1):135–151, 2 2015.

[211] I. Isgum, N. E. Van Der Aa, F. Groenendaal, L. S. De Vries, M. J. Benders, and M. A. Viergever. MRI-based delineation of perinatal arterial ischemic stroke. *14th International Conference on Medical Image Computing and Computer Assisted Intervention.*, Image Anal.

[212] N. Khalili, N. Lessmann, E. Turk, N. Claessens, R. d. Heus, T. Kolk, M. A. Viergever, M. J. Benders, and I. Išgum. Automatic brain tissue segmentation in fetal MRI using convolutional neural networks. *Magnetic Resonance Imaging*, 64:77–89, 12 2019.

[213] N. Khalili, E. Turk, M. Zreik, M. A. Viergever, M. J. Benders, and I. Išgum. Generative adversarial network for segmentation of motion affected neonatal brain MRI. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11766 LNCS:320–328, 6 2019.

[214] D. P. Kingma and J. L. Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015*

- *Conference Track Proceedings*. International Conference on Learning Representations, ICLR, 12 2015.

[215] R. Laugesaar, A. Kolk, T. Tomberg, T. Metsvaht, M. Lintrop, H. Varendi, and T. Talvik. Acutely and retrospectively diagnosed perinatal stroke: A population-based study. *Stroke*, 38(8):2234–2240, 8 2007.

[216] J. Lee, L. A. Croen, C. Lindan, K. B. Nash, C. K. Yoshida, D. M. Ferriero, A. J. Barkovich, and Y. W. Wu. Predictors of outcome in perinatal arterial stroke: A population-based study. *Annals of Neurology*, 58(2):303–308, 8 2005.

[217] D. C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming 1989 45:1*, 45(1):503–528, 8 1989.

[218] A. Makropoulos, E. C. Robinson, A. Schuh, R. Wright, S. Fitzgibbon, J. Bozek, S. J. Counsell, J. Steinweg, K. Vecchiato, J. Passerat-Palmbach, G. Lenz, F. Mortari, T. Tenev, E. P. Duff, M. Bastiani, L. Cordero-Grande, E. Hughes, N. Tusor, J. D. Tournier, J. Hutter, A. N. Price, R. P. A. Teixeira, M. Murgasova, S. Victor, C. Kelly, M. A. Rutherford, S. M. Smith, A. D. Edwards, J. V. Hajnal, M. Jenkinson, and D. Rueckert. The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. *NeuroImage*, 173:88–112, 6 2018.

[219] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. De Vries, M. J. Benders, and I. Isgum. Automatic Segmentation of MR Brain Images with a Convolutional Neural Network. *IEEE Transactions on Medical Imaging*, 35(5):1252–1261, 5 2016.

[220] K. Murphy, N. E. van der Aa, S. Negro, F. Groenendaal, L. S. de Vries, M. A. Viergever, G. B. Boylan, M. J. Benders, and I. Išgum. Automatic quantification of ischemic injury on diffusion-weighted MRI of neonatal hypoxic ischemic encephalopathy. *NeuroImage. Clinical*, 14:222–232, 2017.

[221] K. B. Nelson and J. K. Lynch. Stroke in newborn infants. *The Lancet. Neurology*, 3(3):150–158, 3 2004.

[222] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, 32, 2019.

[223] G. B. Praveen, A. Agrawal, P. Sundaram, and S. Sardesai. Ischemic stroke lesion segmentation using stacked sparse autoencoder. *Computers in Biology and Medicine*, 99:38–52, 8 2018.

[224] O. Ronneberger, P. Fischer, and T. Brox. 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)*, volume 9351, pages 234–241, Munich, Germany, 2015. Springer Verlag.

[225] C. T. Rueden, J. Schindelin, M. C. Hiner, B. E. DeZonia, A. E. Walter, E. T. Arena, and K. W. Eliceiri. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, 18(1):1–26, 11 2017.

[226] S. Schulzke, P. Weber, J. Luetschg, and H. Fahnenstich. Incidence and diagnosis of unilateral arterial cerebral infarction in newborn infants. In *Journal of Perinatal Medicine*, volume 33, pages 170–175. J Perinat Med, 2005.

[227] L. N. Smith. Cyclical learning rates for training neural networks. In *Proceedings - 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017*, pages 464–472. Institute of Electrical and Electronics Engineers Inc., 5 2017.

[228] S. M. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143, 11 2002.

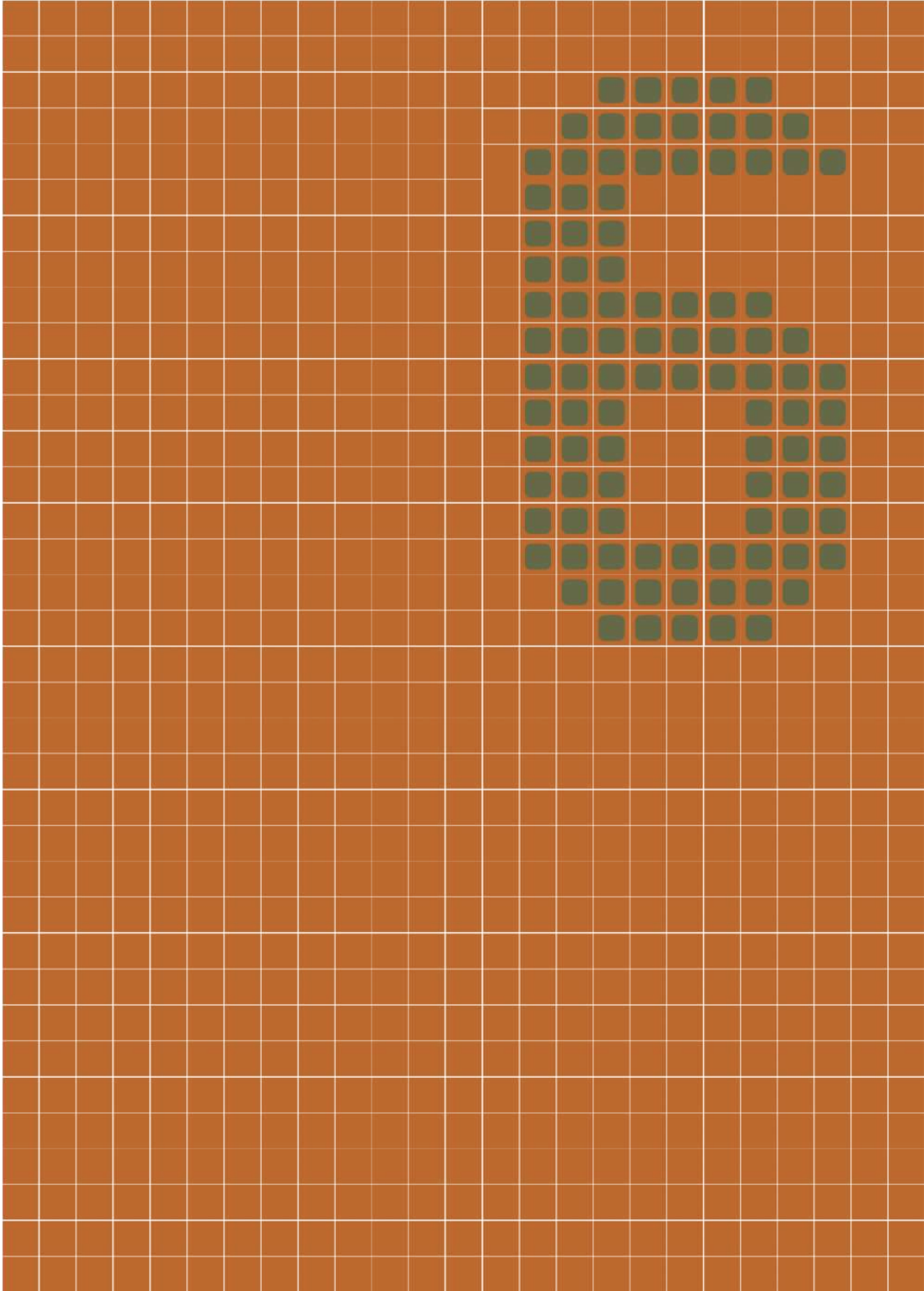
[229] A. L. Sorg, R. Von Kries, M. Klemme, L. Gerstl, U. Felderhoff-Müser, and M. Dzietko. Incidence Estimates of Perinatal Arterial Ischemic Stroke in Preterm- And Term-Born Infants: A National Capture-Recapture Cal-

culation Corrected Surveillance Study. *Neonatology*, 118(6):727–733, 12 2021.

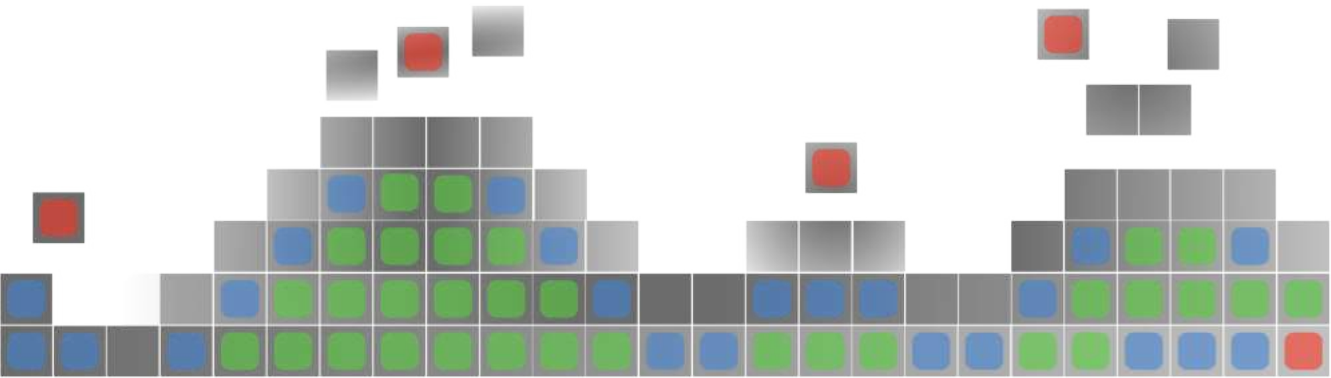
[230] I. Woo, A. Lee, S. Chai Jung, H. Lee, N. Kim, S. Jin Cho, D. Kim, J. Lee, L. Sunwoo, and D.-W. Kang. Fully Automatic Segmentation of Acute Ischemic Lesions on Diffusion-Weighted Imaging Using Convolutional Neural Networks: Comparison with Conventional Algorithms. *Korean J Radiol*, 20(8):1275–1284, 2019.

[231] R. Zhang, L. Zhao, W. Lou, J. M. Abrigo, V. C. Mok, W. C. Chu, D. Wang, and L. Shi. Automatic Segmentation of Acute Ischemic Stroke From DWI Using 3-D Fully Convolutional DenseNets. *IEEE Transactions on Medical Imaging*, 37(9):2149–2160, 9 2018.

[232] Z. Zhou, R. Siddiquee, N. Tajbakhsh, and J. Liang. UNet++: A Nested U-Net Architecture for Medical Image Segmentation.



# Chapter 6
## Summary



# Summary

Developing deep learning-based algorithms that accurately segment structures in scans that are relevant to treatment or evaluation of the outcome of uncommon stroke is a difficult task. The difficulty is due to the presence of image artefacts, few data being available to train the networks, and the small volume of some of the target structures. Hence, the aim of this thesis was to investigate, develop, and evaluate deep learning-based algorithms for automatic segmentation of images of uncommon sub-types of stroke.

Deep transfer learning is a method by which information learned by a deep neural network on one problem can be re-used to improve performance on another problem. It is often used to ameliorate training of neural networks if few data are available. In a transfer learning scenario, a neural network is pre-trained on a source task and domain and re-used, by either fine-tuning or feature extraction, on a target task and domain. In chapter 2, we have evaluated the effect of pre-training by using various source domain and task combinations on target segmentation task performance. Convolutional neural networks were pre-trained on three types of tasks; segmentation, classification, and auto-encoding. The domain used was the same as the target task (T1-weighted MR scans) or different from the target task (natural images). The pre-trained convolutional neural networks were fine-tuned and their performance was evaluated on three target segmentation tasks: Multiple sclerosis lesion, stroke lesion and brain tissue segmentation. Our results showed that pre-training on a segmentation source task on the same domain resulted in a greater improvement in spatial agreement from transfer learning than the other source tasks and domain combinations. Pre-training on a similar task and domain resulted in a greater improvement in spatial overlap on two of the three target segmentation tasks, even when it was compared to pre-training on ImageNet [233], which is a dissimilar source domain and task and consisted of ten times more data. However, our results have also shown that the choice of source task and domain has an inconsistent effect on stroke and multiple sclerosis lesion detection accuracy. Based on this study, we can recommend that pre-training for target medical segmentation tasks should

be done on a similar source task and domain if spatial overlap is the most relevant metric.

Final lesion volume on follow-up NCCT is a common surrogate outcome measure of anterior circulation stroke. Due to the scarcity of available data in posterior circulation stroke, final lesion volume is understudied as a surrogate outcome measure for this pathology. To reduce the manual annotation burden of posterior stroke lesion segmentation, we have developed an algorithm that used deep transfer learning to automatically segment infarct lesions due to a posterior circulation stroke in chapter 3. The model that used deep transfer learning was pre-trained on segmentations of infarct lesions due to a stroke in the anterior circulation. We compared the method that used deep transfer learning to models that were trained on only anterior circulation stroke patients, only on posterior circulation stroke patients, and on patients suffering from a stroke in either one of the regions. The model that used deep transfer learning achieved a greater volumetric agreement and a higher spatial overlap between the automatic and reference lesion segmentation than the other methods. Furthermore, the deep transfer learned method also improved lesion detection relative to the other methods. This method facilitates investigation of final lesion volume as a surrogate outcome measure in posterior circulation stroke.

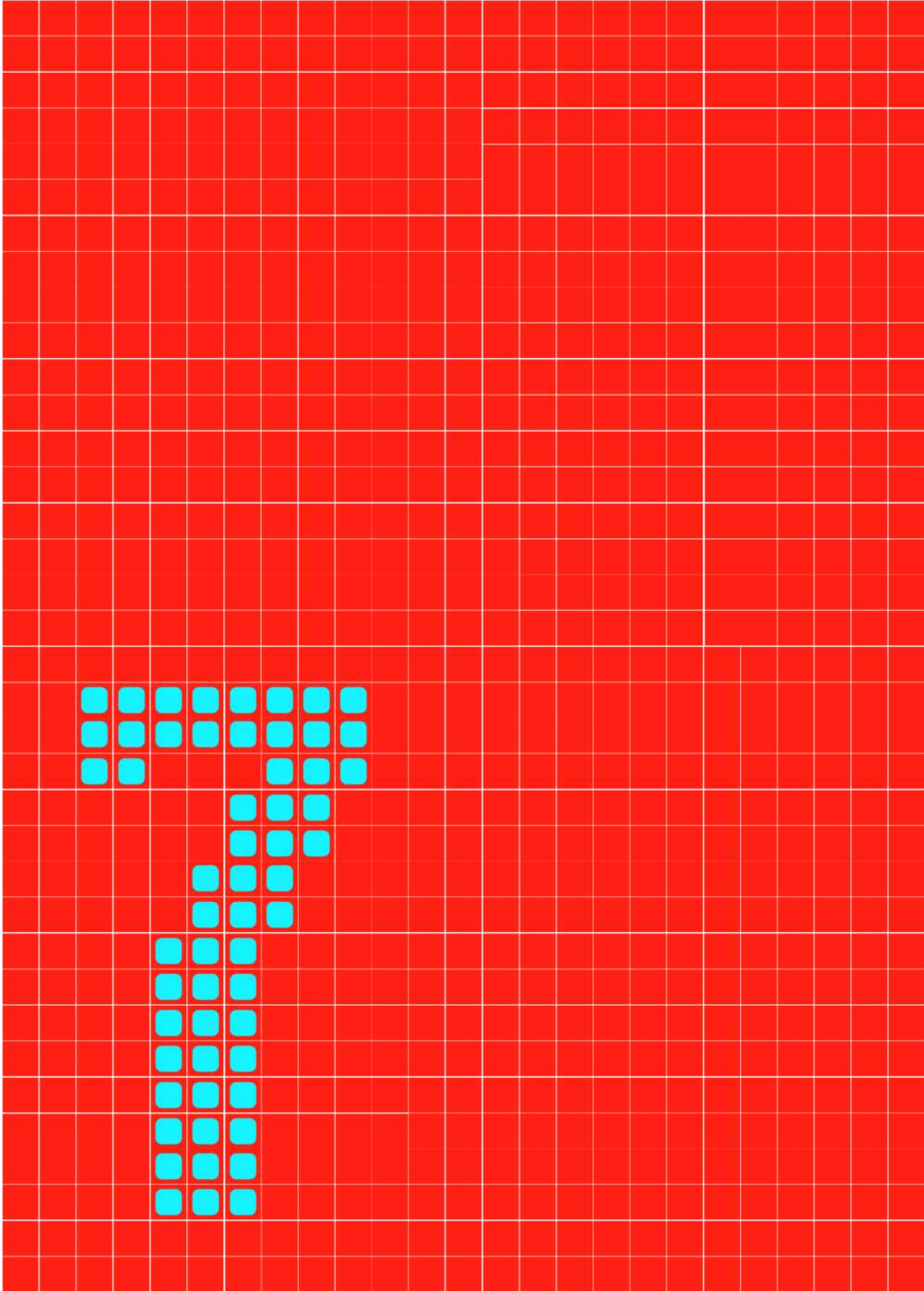


Segmentation-based thrombus image characteristics have been associated with stroke treatment outcome. Manual annotation of thrombi may limit the study of these associations. Hence, in chapter 4, we have developed an automatic method that localizes and segments thrombi causing a posterior circulation stroke. Segmentation methods for posterior thrombi that are not restricted to a specific region in the brain, segment a large number of false positive thrombi. We have shown that our method, which restricted itself to the area around the brainstem, improved performance relative to a standard UNet and reduced the number of false positives. Our method can be used to reduce the manual annotation burden for investigating the association between segmentation-based thrombus image characteristics and various outcome metrics.

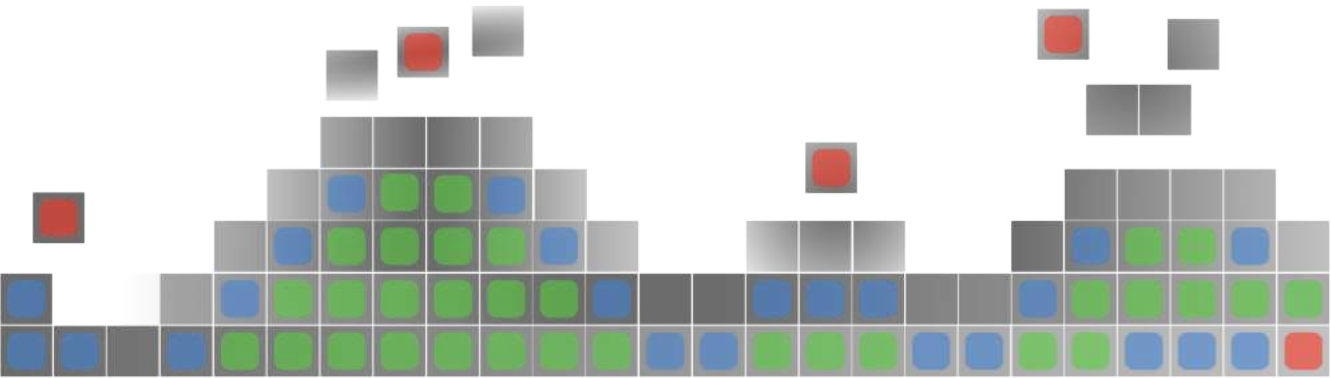
Brain-tissue and ischemic lesion volume per hemisphere can be used to evaluate treatment efficacy in patients suffering from perinatal arterial ischemic stroke. Therefore, in chapter 5, we developed two instances of a convolutional neural network to segment the white matter, gray matter, cerebrospinal fluid, brainstem, cerebellum, basal ganglia and thalamus, ventricles, and the ischemic lesion in scans of patients suffering from perinatal arterial ischemic stroke. One network instance automatically segmented scans acquired at baseline, the other instance automatically segmented scans acquired at follow-up. Our network instances achieved comparable spatial overlap and alignment to methods that were developed in related research for brain tissue segmentation in scans of healthy brains and ischemic lesion segmentation in scans of adult patients suffering acute ischemic stroke.

# Bibliography

[233] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252, 9 2014.



# Chapter 7
## General Discussion



# General Discussion

## Contributions to stroke research

An important aspect of modern stroke research involves analysing segmentation-based image characteristics derived from scans, such as thrombus volume and lesion volume. However, manual annotation of image data is labor- and time-intensive. Hence, there is an interest to automate the segmentation process. Currently, deep learning is the best available technique to automate image segmentation. A limiting factor to the development, application, and evaluation of deep learning-based segmentation methods is that they often require a large amount of data. Less common sub-types of stroke that have few available data have not yet benefited from deep learning-based segmentation. Therefore, the primary focus of this thesis was to investigate, develop, and evaluate deep learning-based algorithms for the segmentation of images of sub-types of stroke that have few data available.

Final lesion volume has been proposed as an outcome measure in acute stroke reperfusion clinical trials [258]. Subsequently, follow-up NCCT was shown to have a strong association to functional outcome in anterior circulation stroke [237; 260]. Due to the scarcity of available data in posterior circulation stroke, the association between final lesion volume and functional outcome is understudied. The study in chapter 3 has shown that deep learning can be used to automatically segment lesions due to a posterior circulation large vessel occlusion on follow-up NCCT of adult patients. The developed algorithm can be used to study associations between segmentation-based lesion characteristics, such as lesion volume, and functional outcome. The method can reduce the annotation burden by providing automatic segmentations. However, due to the low spatial overlap the resulting automatically generated segmentations should be inspected and, if necessary, manually corrected.

Thrombus image characteristics have been associated with treatment outcome in anterior circulation stroke in adult patients by prior studies. The studied thrombus image characteristics for anterior circulation stroke are per-

viousness, density, and length [236; 238; 242; 245; 249; 254; 255]. However, one study showed that the only thrombus characteristic that is associated with functional outcome in posterior circulation stroke is thrombus length [240]. A limitation of current applications of thrombus length, perviousness, and density assessment is that they are derived from Region-of-Interest (ROI) markers in the proximal, medial, and distal parts of the thrombus rather than complete segmentations. Prior research has already shown that thrombus segmentations provide more information than ROI markers [256]. Segmentations have been used in conjunction with techniques from radiomics to extract thrombus characteristics in patients with anterior circulation stroke. These characteristics were found to be more predictive than thrombus length, permeability, and density [253]. The algorithm developed in 4 makes it easier to test for these associations by reducing the annotation burden in posterior stroke clinical trials and registries that involve adult patients.

Effectiveness of novel neuro-protective and neuro-regenerative treatments, such as recombinant tissue erythropoietin [235] and mesenchymal stromal cells [234], can be tested by comparing brain tissue volumes between each hemisphere on baseline and follow-up scans. This is possible due to most cases of stroke being confined to a single hemisphere. If treatments are effective, patients with perinatal arterial ischemic stroke treated with neuro-protective and neuro-regenerative agents should have a similar volume for all tissue classes in both hemispheres due to more brain tissue retention and growth after the stroke. Moreover, the volume of the ischemic region can be added to the analysis to control for possible interactions between lesion size and treatment. However, annotating each scan manually would require approximately a week of work per scan. The network instances developed in chapter 5 reduce the manual annotation burden by automating the segmentation process. A limitation of the presented network instances is that the resulting segmentations may require manual inspection and correction before being used in an analysis, specifically of the ischemic region on the baseline scans. Moreover, due to limited variability in the training dataset, generalizations to scans from other MRI machines and protocols may be limited. Hence, further development is required before possible clinical use of our network

instances.

## Contributions to medical image segmentation research

With the exception of the brain-tissue segmentation algorithms introduced in chapter 5, all methods make use of some variant of transfer learning and build on the findings in chapter 2. In chapter 2, we demonstrated that applying transfer learning to medical segmentation tasks yields the best results if the source domain and task are similar to the target domain and task. Hence, pre-training on automatically generated segmentations adds little cost relative to self-supervised pre-training, but has greater benefit if the data is similar.

Accurately segmenting small objects in medical scans is a difficult task due to the large class imbalance, limited available data, and variable presentation. A standard UNet approach applied to the entire scan volume results in a large number of false positive segmentations. By localizing small objects prior to segmenting small objects, false positive segmentation can be reduced. Various methods that localize a structure in a medical scan prior to segmenting it have been developed in earlier work [239; 244; 247; 250; 257; 261; 262]. An often used method of localization is by using a bounding box regression [239; 241; 247; 261]. Bouget et al. combined a 2D Mask R-CNN in conjunction with a 2D U-Net to localize and segment mediastinal lymph nodes and anatomical structures in the mediastinal area [239]. The 2D bounding boxes are converted to 3D by combining consecutive 2D bounding boxes. Liang et al. used a different method to combine bounding boxes to localize and segment organs at risk in head and neck images [247]. Their method combined features extracted from axial, coronal or sagittal slices as input. By using a voting method the features were combined before bounding box regression and segmentation. Zhang et al. used a fixed-size bounding box located around the esophagus to improve segmentation [261]. The bounding box location was found by using previously segmented structures that are close to the oesophagus. De Vos et al. used a neural network to predicted the probability that a structure was visible on slices from the three anatomical planes [241]. Using the starting location and ending location on each anatomical plane, a bounding box was constructed. Using a 3D bounding box would not have

been the best way of detecting thrombi in the posterior circulation, which was described in chapter 4. This is due to the variable orientation that the thrombus has, which would cause a bounding box to be much larger than required in order to contain the entire thrombus. This may result in a larger amount of false positive thrombus localization and segmentation.

## Future research directions

### *Posterior Circulation Stroke*

The results from our automated posterior stroke lesion segmentation method can be used to estimate the overall infarcted volume. However, the severity of clinical deficit not only depends on the total volume but may also depend on the specific affected substructures. A study has already shown that lesion volume in specific structures is associated with worse outcome 90 days after the stroke than other structures in anterior circulation stroke [243]. In posterior circulation stroke, hypo-attenuation of specific structures is predictive of worse outcome [252]. Similarly, future research could focus on developing a model to quantify the volume of these hypo-attenuated structures of the brain in the posterior region.

One interesting research direction that could be explored to improve the quantification of the thrombus characteristics, arterial filling and stenosis detection during a posterior circulation stroke is deep learning-based artery centerline tracking. Wolterink et al. developed a method to track the coronary arteries [259]. The automatically extracted centerlines can help visualization of the artery by straightening it using Multi-Planar Reformatting (MPR). Subsequent research used MPR straightened arteries to localize plaque type and the clinical significance of any stenosis [263]. Likewise, for posterior stroke such an approach could be used to segment the thrombus, stenosis, and the lumen of the involved arteries.

Deep learning approaches that involve segmentation in the posterior fossa are hampered by CT image artifacts, such as beam hardening. An interesting challenge could be to make deep learning-based methods more robust to CT imaging artifacts. Specifically, a data augmentation method could be devised

to mimic CT artifacts during training. Such data augmentation techniques already exist for MRI [246; 251] and CT .

### *Perinatal Arterial Ischemic Stroke*

The methods that were developed in chapter 5 segment brain tissue in patients that suffer or have suffered from perinatal arterial ischemic stroke. These methods currently rely on two separate instances of a deep neural network; one for term and one for three month follow-up scans. We developed separate network instances for term and follow-up scans for two reasons. First, the degree of myelination increases as the infant develops. This causes the appearance of the brain on T2 scans acquired at term versus follow-up to differ. Second, during the term scan the DWI shows the hypo-perfused area in the brain. Hence, this sequence is added to the network instance trained at term but not at follow-up. Future trials and registries for this type of stroke may, however, have imaging performed at other stages of development of the infant. Because the brain changes rapidly during the first year of life, generalization of the network instances to images acquired at other stages of development cannot be assumed. Therefore, a more general method to segment brain tissue on MRI scans acquired during the first year of life, which has already been developed for scans of patients not afflicted by severe pathology [248], is of interest for brains afflicted by perinatal arterial ischemic stroke.

# Bibliography

[234] L. M. Baak, N. Wagenaar, N. E. van der Aa, F. Groenendaal, J. Dudink, M. L. Tataranno, U. Mahamuud, L. S. de Vries, F. van Bel, M. J. N L Benders, C. Development, R. K. Jellema, L. M. Baak, N. Wagenaar, N. E. van der Aa, F. Groenendaal, J. Dudink, M. Luisa Tataranno, U. Mahamuud, C. H. Verhage, R. M. J C Eijsermans, L. S. Smit, R. K. Jellema, T. R. de Haan, H. J. ter Horst, W. P. de Boode, S. J. Steggerda, H.-J. Prins, C. G. de Haar, L. S. de Vries, F. van Bel, C. J. Heijnen, C. H. Nijboer, and M. J. N L Benders. Feasibility and safety of intranasally administered mesenchymal stromal cells after perinatal arterial ischaemic stroke in the Netherlands (PASSIoN): a first-in-human, open-label intervention study. *The Lancet Neurology*, 21(6):528–536, 6 2022.

[235] M. J. Benders, N. E. Van Der Aa, M. Roks, H. L. Van Straaten, I. Isgum, M. A. Viergever, F. Groenendaal, L. S. De Vries, and F. Van Bel. Feasibility and safety of erythropoietin for neuroprotection after perinatal arterial ischemic stroke. *The Journal of pediatrics*, 164(3), 2014.

[236] A. B. Bilgic, R. Gocmen, E. M. Arsava, and M. A. Topcuoglu. The Effect of Clot Volume and Permeability on Response to Intravenous Tissue Plasminogen Activator in Acute Ischemic Stroke. *Journal of Stroke and Cerebrovascular Diseases*, 29(2):104541, 2 2020.

[237] A. M. Boers, I. Jansen, l. F. M Beenen, T. Devlin, L. san roman, J. hoe heo, M. Ribó, S. Brown, M. Almekhlafi, D. Liebeskind, J. Teitelbaum, h. F. Lingsma, W. van Zwam, P. Cuadras, M. Beaumont, M. M. Brown, a. J. Yoo, r. J. van Oostenbrugge, B. K. Menon, G. Donnan, J. louis Mas, Y. B. W M roos, C. Oppenheim, A. van der lugt, r. J. Dowling, M. D. Hill, A. Davalos, T. Moulin, N. Agrinier, a. M. Demchuk, D. K. Lopes, L. aja rodríguez, D. W. J Dippel, B. V. Campbell, P. J. Mitchell, F. Al-ajlan, T. Jovin, J. Madigan, g. W. Albers, S. Soize, F. Guillemin, V. K. Reddy, S. Bracard, J. Blasco, K. W. Muir, R. Nogueira, P. M. White, M. Goyal, s. M. Davis, H. Marquering, and c. B. M Majoie. Association of follow-up infarct volume with functional outcome in acute ischemic stroke:

a pooled analysis of seven randomized trials. *J NeuroIntervent Surg*, 10:1137–1142, 2018.

[238] J. Borst, O. A. Berkhemer, E. M. Santos, A. J. Yoo, M. Den Blanken, Y. B. Roos, E. Van Bavel, W. H. Van Zwam, R. J. Van Oostenbrugge, H. F. Lingsma, A. Van Der Lugt, D. W. Dippel, H. A. Marquering, and C. B. Majoie. Value of Thrombus CT Characteristics in Patients with Acute Ischemic Stroke. *AJNR. American journal of neuroradiology*, 38(9):1758–1764, 9 2017.

[239] D. Bouget, A. Jørgensen, G. Kiss, H. O. Leira, and T. Langø. Semantic segmentation and detection of mediastinal lymph nodes and anatomical structures in CT data for lung cancer staging. *International Journal of Computer Assisted Radiology and Surgery*, 14(6):977–986, 3 2019.

[240] A. A. E. Bruggeman, J. Brouwer, . Nerea, A. Terreros, N. Boodt, S. J. Den Hartog, A. Pirson, . Ludo, F. M. Beenen, P.-J. Van Doormaal, R. J. Van Oostenbrugge, J. Staals, . Wim, H. Van Zwam, . Wouter, J. Schonewille, . Lucianne, C. M. Langezaal, J.-A. Vos, . Aad Van Der Lugt, . Diederik, W. J. Dippel, J. M. Coutinho, . Yvo, B. W. E. M. Roos, H. A. Marquering, B. J. Emmer, . Charles, and B. L. M. Majoie. Thrombus Imaging Characteristics and Outcomes in Posterior Circulation Stroke Patients Treated With EVT. *Stroke: Vascular and Interventional Neurology*, 2(3), 5 2022.

[241] B. D. De Vos, J. M. Wolterink, P. A. De Jong, T. Leiner, and M. A. Viergever. ConvNet-Based Localization of Anatomical Structures in 3D Medical Images.

[242] B. G. Dutra, M. L. Tolhuisen, H. C. Alves, K. M. Treurniet, M. Kappelhof, A. J. Yoo, I. G. Jansen, D. W. Dippel, W. H. Van Zwam, R. J. Van Oostenbrugge, A. J. Da Rocha, H. F. Lingsma, A. Van Der Lugt, Y. B. Roos, H. A. Marquering, and C. B. Majoie. Thrombus Imaging Characteristics and Outcomes in Acute Ischemic Stroke Patients Undergoing Endovascular Treatment. *Stroke*, 50(8):2057–2064, 8 2019.

[243] M. Ernst, A. M. Boers, A. Aigner, O. A. Berkhemer, A. J. Yoo, Y. B. Roos, D. W. Dippel, A. Van Der Lugt, R. J. Van Oostenbrugge, W. H. Van Zwam,

J. Fiehler, H. A. Marquering, and C. B. Majoie. Association of Computed Tomography Ischemic Lesion Location with Functional Outcome in Acute Large Vessel Occlusion Ischemic Stroke. *Stroke*, 48(9):2426–2433, 9 2017.

[244] K. He, X. Cao, Y. Shi, D. Nie, Y. Gao, and D. Shen. Pelvic Organ Segmentation Using Distinctive Curve Guided Fully Convolutional Networks. *IEEE Transactions on Medical Imaging*, 38(2):585–595, 2 2019.

[245] M. Kappelhof, M. L. Tolhuisen, K. M. Treurniet, B. G. Dutra, H. Alves, G. Zhang, S. Brown, K. W. Muir, A. Davalos, Y. B. Roos, J. L. Saver, A. M. Demchuk, T. G. Jovin, S. Bracard, B. C. Campbell, A. Van Der Lugt, F. Guillemin, P. White, M. D. Hill, D. W. Dippel, P. J. Mitchell, M. Goyal, H. A. Marquering, and C. B. Majoie. Endovascular Treatment Effect Diminishes with Increasing Thrombus Perviousness: Pooled Data from 7 Trials on Acute Ischemic Stroke. *Stroke*, 52(11):3633–3641, 11 2021.

[246] N. Khalili, N. Lessmann, E. Turk, N. Claessens, R. d. Heus, T. Kolk, M. A. Viergever, M. J. Benders, and I. Išgum. Automatic brain tissue segmentation in fetal MRI using convolutional neural networks. *Magnetic Resonance Imaging*, 64:77–89, 12 2019.

[247] S. Liang, K. H. Thung, D. Nie, Y. Zhang, and D. Shen. Multi-View Spatial Aggregation Framework for Joint Localization and Segmentation of Organs at Risk in Head and Neck CT Images. *IEEE Transactions on Medical Imaging*, 39(9):2794–2805, 9 2020.

[248] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. De Vries, M. J. Benders, and I. Isgum. Automatic Segmentation of MR Brain Images with a Convolutional Neural Network. *IEEE Transactions on Medical Imaging*, 35(5):1252–1261, 5 2016.

[249] M. Mokin, M. Waqas, J. Fifi, R. De Leacy, D. Fiorella, E. I. Levy, K. Snyder, R. Hanel, K. Woodward, I. Chaudry, A. T. Rai, D. Frei, J. E. Delgado Almandoz, M. Kelly, A. S. Arthur, B. W. Baxter, J. English, I. Linfante, K. M. Fargen, A. Turk, A. H. Siddiqui, and J. Mocco. Clot perviousness is associated with first pass success of aspiration thrombectomy in the COMPASS trial. *Journal of NeuroInterventional Surgery*, 13(6):509–514, 6 2021.

[250] J. Noothout, E. Postma, S. Boesveldt, B. D. de Vos, P. Smeets, and I. Išgum. Automatic segmentation of the olfactory bulbs in MRI. In *115961J*, volume 11596, page 51. SPIE-Intl Soc Optical Eng, 2 2021.

[251] F. Pérez-García, R. Sparks, and S. Ourselin. TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods and Programs in Biomedicine*, 208, 9 2021.

[252] V. Puetz, P. N. Sylaja, S. B. Coutts, M. D. Hill, I. Dzialowski, P. Mueller, U. Becker, G. Urban, C. O'Reilly, P. A. Barber, P. Sharma, M. Goyal, G. Gahn, R. Von Kummer, and A. M. Demchuk. Extent of hypoattenuation on CT angiography source images predicts functional outcome in patients with basilar artery occlusion. *Stroke*, 39(9):2485–2490, 9 2008.

[253] W. Qiu, H. Kuang, J. Nair, Z. Assis, M. Najm, C. McDougall, B. McDougall, K. Chung, A. T. Wilson, M. Goyal, M. D. Hill, A. M. Demchuk, and B. K. Menon. Radiomics-Based Intracranial Thrombus Features on CT and CTA Predict Recanalization with Intravenous Alteplase in Patients with Acute Ischemic Stroke. *American Journal of Neuroradiology*, 40(1):39–44, 1 2019.

[254] E. M. Santos, J. W. Dankbaar, K. M. Treurniet, A. D. Horsch, Y. B. Roos, L. J. Kappelle, W. J. Niessen, C. B. Majoie, B. Velthuis, and H. A. Marquering. Permeable Thrombi Are Associated with Higher Intravenous Recombinant Tissue-Type Plasminogen Activator Treatment Success in Patients with Acute Ischemic Stroke. *Stroke*, 47(8):2058–2065, 8 2016.

[255] E. M. Santos, H. A. Marquering, M. D. Den Blanken, O. A. Berkhemer, A. M. Boers, A. J. Yoo, L. F. Beenen, K. M. Treurniet, C. Wismans, K. Van Noort, H. F. Lingsma, D. W. Dippel, A. Van Der Lugt, W. H. Van Zwam, Y. B. Roos, R. J. Van Oostenbrugge, W. J. Niessen, and C. B. Majoie. Thrombus Permeability Is Associated with Improved Functional Outcome and Recanalization in Patients with Ischemic Stroke. *Stroke*, 47(3):732–741, 3 2016.

[256] E. M. Santos, A. J. Yoo, L. F. Beenen, O. A. Berkhemer, M. D. den Blanken, C. Wismans, W. J. Niessen, C. B. Majoie, and H. A. Marquering. Observer

variability of absolute and relative thrombus density measurements in patients with acute ischemic stroke. *Neuroradiology*, 58(2):133–139, 2 2016.

[257] S. Wang, K. He, D. Nie, S. Zhou, Y. Gao, and D. Shen. CT male pelvic organ segmentation using fully convolutional networks with boundary sensitive representation. *Medical Image Analysis*, 54:168–178, 5 2019.

[258] S. J. Warach, M. Luby, G. W. Albers, R. Bammer, A. Bivard, B. C. Campbell, C. Derdeyn, J. J. Heit, P. Khatri, M. G. Lansberg, D. S. Liebeskind, C. B. Majoie, M. P. Marks, B. K. Menon, K. W. Muir, M. W. Parsons, A. Vagal, A. J. Yoo, A. V. Alexandrov, J.-C. Baron, D. J. Fiorella, A. J. Furlan, J. Puig, P. D. Schellinger, and M. Wintermark. Acute Stroke Imaging Research Roadmap III Imaging Selection and Outcomes in Acute Stroke Reperfusion Clinical Trials. *Stroke*, 47(5):1389–1398, 2016.

[259] J. M. Wolterink, R. W. Hamersvelt, M. A. Viergever, T. Leiner, and I. Išgum. Coronary artery centerline extraction in cardiac CT angiography using a CNN-based orientation classifier. *Medical Image Analysis*, 51:46–60, 1 2019.

[260] S. F. Zaidi, A. Aghaebrahim, X. Urra, M. A. Jumaa, B. Jankowitz, M. Hammer, R. Nogueira, M. Horowitz, V. Reddy, and T. G. Jovin. Final infarct volume is a stronger predictor of outcome than recanalization in patients with proximal middle cerebral artery occlusion treated with endovascular therapy. *Stroke*, 43(12):3238–3244, 12 2012.

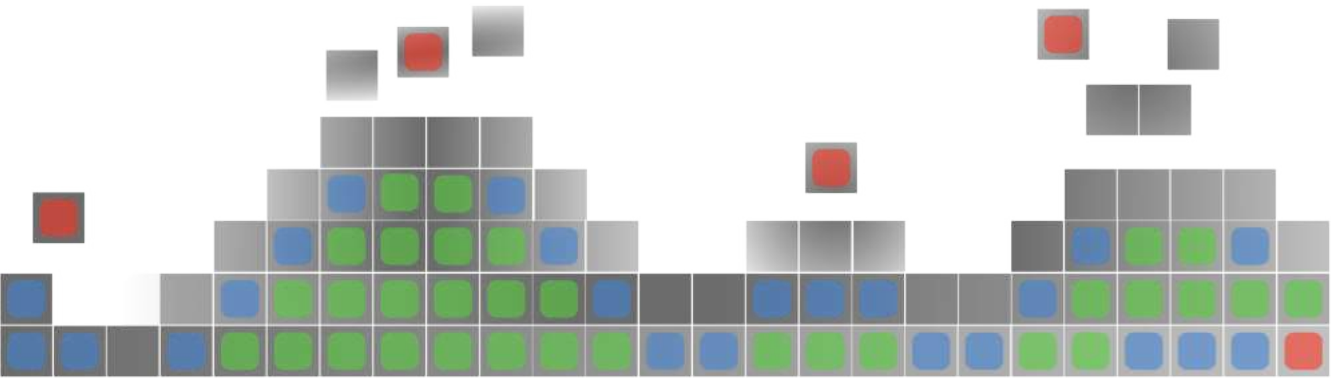
[261] L. Zhang, J. Zhang, P. Shen, G. Zhu, P. Li, X. Lu, H. Zhang, S. A. Shah, and M. Bennamoun. Block Level Skip Connections across Cascaded V-Net for Multi-Organ Segmentation. *IEEE Transactions on Medical Imaging*, 39(9):2782–2793, 9 2020.

[262] Y. Zhao, H. Li, S. Wan, A. Sekuboyina, X. Hu, G. Tetteh, M. Piraud, and B. Menze. Knowledge-Aided Convolutional Neural Network for Small Organ Segmentation. *IEEE Journal of Biomedical and Health Informatics*, 23(4):1363–1373, 7 2019.

[263] M. Zreik, R. W. Van Hamersvelt, J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum. A Recurrent CNN for Automatic Detection

and Classification of Coronary Artery Plaque and Stenosis in Coronary CT Angiography. *IEEE Transactions on Medical Imaging*, 38(7):1588–1598, 7 2019.

# Chapter 8
## Nederlandse Samenvatting



# Nederlandse Samenvatting

Het ontwikkelen van op deep learning gebaseerde algoritmen die accuraat structuren in scans kunnen segmenteren die relevant zijn voor de behandeling of de evaluatie van de uitkomst van ongebruikelijke beroerten is een moeilijke taak. Deze moeilijkheid komt door de aanwezigheid van artefacten in scans, de kleine hoeveelheid data die beschikbaar is voor het trainen van neurale netwerken en het kleine volume van de te segmenteren structuren. Daarom was het doel van deze these om op deep learning gebaseerde algoritmen voor het automatisch segmenteren van beelden van ongebruikelijke subtypen van beroerten te onderzoeken, ontwikkelen en evalueren.

Deep transfer learning is een methode waarmee informatie die geleerd is door een diep neuraal netwerk op een probleem kan worden hergebruikt om de prestaties te verbeteren op een ander probleem. Het wordt vaak gebruikt om het trainen van neurale netwerken te verbertern wanneer er weinig data beschikbaar is. In een transfer learning scenario, wordt een neuraal netwerk dat is voor-getraind op een oorsprongstaak en domein hergebruikt voor een doeltaak en doeldomein. Dit hergebruiken gebeurt doormiddel van het hercalibereren van het netwerk of door het toepassen van het netwerk om kenmerken te extraheren van de data uit het doeldomein met mogelijkerwijs een andere doeltaak. In hoofdstuk 2 hebben we het effect van het voor-trainen geëvalueerd op accuratesse van de segmentatie doeltaken, wanneer er verschillende oorsprongs-domeinen en taken werden gebruikt. Convolutionele neurale netwerken werden voor-getrained op drie verschillende oorsprongstaken; segmentatie, classificatie en auto-encoding. Het oorsprongsdomein dat werd gebruikt was hetzelfde als die van de doeltaak (T1-gewogen MRI scans) of het verschilde van die van de doeltaak (natuurlijke beelden). De voorgetrainede convolutionele neurale netwerken werden hergecalibreerd en hun prestaties geëvalueerd op drie segmentatie doeltaken: Laesies veroorzaakt door multipele sclerose, laesies veroorzaakt door een beroerte en hersenweefseltype segmentatie. Onze resultaten lieten zien dat voor-trainen op een segmentatie oorsprongstaak op hetzelfde domain resulteerde in een grotere verbetering in de spatiële overeenstemming door transfer learning dan andere

combinaties van oorsprongstaken en oorsprongsdomeinen. Het voor-trainen op een vergelijkbare taak en domein resulteerde zelfs in een grotere verbetering van de spatiële overeenstemming op twee van de drie segmentatie doeltaken wanneer het werd vergeleken met voor-trainen op de ImageNet dataset [264], een dataset die bestaat uit data van ander domein met een andere taak. Dit was ondanks dat de ImageNet dataset uit tien keer meer data bestond. Onze resultaten hebben echter ook laten zien dat de keuze van de oorsprongstaak en het oorsprongsdomein een inconsisten effect heeft op laesie detectie accuratesse door multipele sclerose of beroerten. Op basis van deze studie kunnen we aanbevelen dat het voor-trainen op medische segmentatie doeltaken gedaan zou moeten worden op een vergelijkbare oorsprongstaak en domein wanneer de spatiële overeenstemming de meest relevante maatstaf is.

Het uiteindelijke volume van de laesie op vervolg NCCT is een gebruikelijke surrogaat uitkomstmaat na een beroerte in de anterieure circulatie. Doordat data van patienten met een beroerte in de posterieure circulatie schaars is, is het uiteindelijke volume van de laesie van dit ziektebeeld onderbestudeerd als surrogaat uitkomstmaat. Om de handmatige annotatie last van het segmenteren van laesies door een beroerte in de posterieure circulatie te verminderen, hebben we in hoofdstuk 3 een algoritme ontwikkeld dat doormiddel van deep transfer learning automatisch laesies segmenteert die ontstaan zijn door een beroerte. Het model dat deep transfer learning gebruikte, was voor-getrained op segmentaties van infarct laesies die door een beroerte in de anterieure circulatie waren ontstaan. We vergeleken het model dat deep transfer learning gebruikte met drie modellen die met andere data waren getraind: Het eerste model was getrained enkel op data van patienten met een beroerte in de anterieure circulatie. Het tweede model was enkel getraind op data van patienten met een beroerte in de posterieure circulatie. Het derde model was getraind op data van patienten met een beroerte in ofwel de posterieure circulatie ofwel de anterieure circulatie. Het model dat deep transfer learning gebruikte behaalde een grotere volumetrische en spatiële overeenstemming tussen de automatische en handmatige laesie segmentaties dan de andere modellen. Bovendien verbeterde het model dat deep transfer learning gebruikte ook de laesie detectie ten opzicht van de andere modellen.

De ontwikkelde methode faciliteert verder onderzoek in het gebruik van het uiteindelijke laesie volume na een beroerte in de posterieure circulatie als surrogaat uitkomstmaat.

Op segmentatie gebaseerde thrombus beeld eigenschappen zijn geassocieerd met de uitkomsten van de behandeling van een beroerte. Handmatige annotatie van thrombi kan het bestuderen van deze associaties limiteren. Om deze reden heben we in hoofdstuk 4 een methode ontwikkeld die thrombi die een beroerte in de posterieure circulatie veroorzaken automatisch lokaliseert en segmenteert. Segmentatie methoden voor thrombi in de posterieure circulatie, die niet worden beperkt tot een bepaalde regio in de hersenen, segmenteren een groot aantal vals positieve thrombi. We hebben laten zien dat onze methode, welke zichzelf limiteerde tot het gebied rondom de hersenstam, de prestaties verbeterde ten opzichte van een standaard U-Net. Onze methode kan worden gebruikt om de manuele annotatie last te verminderen welke nodig is om de associatie tussen op segmentatie gebaseerde thrombus beeld karakteristieken en verschillende uitkomstmaten te bestuderen.

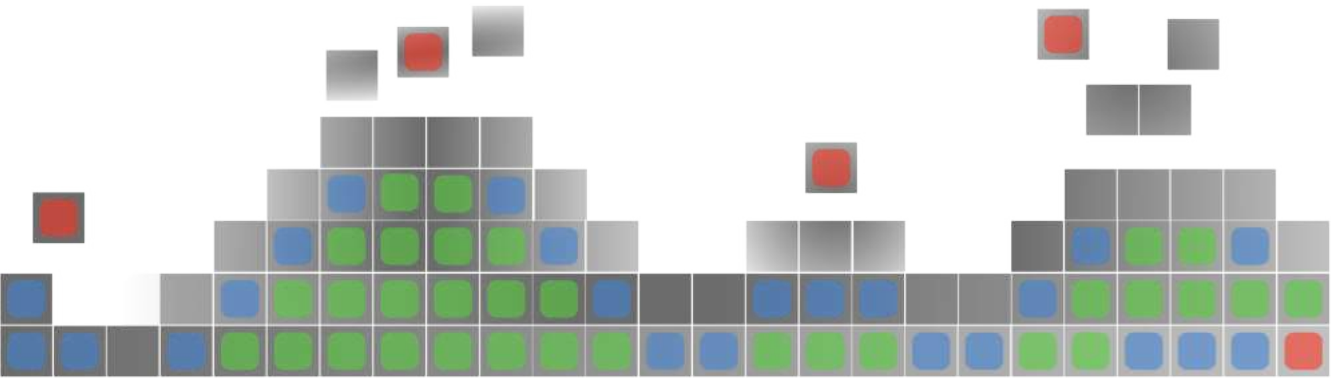
Het volume van het hersenweefsel en de ischemische laesie per hemisfeer kan worden gebruikt om de effectiviteit van de behandeling van patienten die leiden aan een perinatale arteriële ischemische beroerte te evalueren. Om deze reden hebben we in hoofdstuk 5 twee exemplaren van een convolutioneel neuraal netwerk ontwikkeld om witte stof, grijze stof, liquor, de hersenstam, het cerebellum, de basale ganglia en thalami, de ventrikelen en de ischemische laesie te segmenteren in de basislijn- en vervolgs-scans van patienten die leiden aan een perinatale arteriële ischemische beroerte. Een van de netwerk exemplaren segmenteerde automatisch de scans die verkregen waren tijdens de basislijn en het andere exemplare segmenteerde automatisch de vervolgscans. Onze netwerk exemplaren bereikten een vergelijkbare spatiële overeenstemming en uitlijning met methoden die in gerelateerd onderzoek werden ontwikkeld voor hersenweefsel segmentatie in scans van gezonde hersenen en ischemische laesie segmentatie in scans van volwassen patienten die aan een acute ischemische beroerte leden.

Twee problemen traden vaak op in de segmentaties die werden gecreëerd door onze netwerk exemplaren. Allereerst, waren er vals positieve segmentaties die buiten en niet verbonden aan de hersenen waren. Deze vals positieven werden verwijderd door de segmentatie map in twee klassen te scheiden: een klasse die bestond uit alleen achtergrond voxels en een klasse die bestond uit al het hersenweefsel en de ischemische laesie samengevoegd. Vervolgens, werd er een verbonden componeneten analyse gedaan en werd alleen het weefsel en de ischemische laesie gehouden die volledig binnen de grootste verbonden component viel. Als tweede waren er gedeelten van de hemisfeer segmentaties waarin het hersenweefsel en de ischemische laesie correct waren geklassificeerd, maar niet de hemisfeer. Dit werd gecorrigeerd door middel van morfologisch sluiten. Onze netwerk exemplaren bereikten een vergelijkbare spatiële overeenstemming en uitlijning met methoden voor hersenweefsel segmentatie in scans van gezonde hersenen en ischemische laesie segmentatie in scans van volwassen patienten die aan een acute ischemische beroerte leden.

# Bibliography

[264] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252, 9 2014.

# Appendix



# Abbreviations

| | |
|---|---|
| **95% CI** | 95% Confidence Interval |
| **ACS** | Anterior Circulation Stroke |
| **ADC** | apparent diffusion coefficient |
| **AIS** | Acute Ischemic Stroke |
| **ANN** | Artificial Neural Network |
| **ATLAS** | Anatomical Tracings of Lesions After Stroke |
| **BA** | Brain Anatomy |
| **BAHC** | Brain Age Healthy Cohort |
| **BGT** | basal ganglia and thalami |
| **BST** | brainstem |
| **CB** | cerebellum |
| **CNN** | Convolutional Neural Networks |
| **CT** | Computed Tomography |
| **CTA** | Computed Tomography Angiography |
| **DWI** | Diffusion Weighted Imaging |
| **eCSF** | extra-ventricular cerebro-spinal fluid |
| **FLV** | Final Lesion Volume |
| **FN** | False Negative |
| **FP** | False Positive |
| **FU-NCCT** | Follow-Up Non-Contrast Computed Tomography |
| **GAN** | Generative Adversarial Network |
| **GM** | gray matter |
| **HAS** | Hyper-dense Artery Sign |
| **HS** | Hemisphere separation |
| **ICC** | Intraclass Correlation Coefficient |
| **ISLES** | Ischemic Stroke LEsion Segmentation |
| **LVO** | Large Vessel Occlusion |
| **mIOU** | mean Intersection Over Union |
| **MPR** | Multi-Planar Reformatting |
| **MR** | Magnetic Resonance |
| **MRI** | Magnetic Resonance Imaging |
| **mRS** | modified Ranking Scale |

| | |
|---|---|
| **MS** | Multiple Sclerosis |
| **NCCT** | Non-Contrast Computed Tomography |
| **PAIS** | Perinatal Arterial Ischemic Stroke |
| **PC-ASPECTS** | Posterior Circulation Alberta Stroke Program Early Computed Tomography Score |
| **PCS** | Posterior Circulation Stroke |
| **rhEPO** | recombinant human erythropoietin |
| **ROI** | Region-of-Interest |
| **rTPA** | recombinant Tissue Plasminogen Activator |
| **TP** | True Positive |
| **VBR** | Volume-based Removal |
| **vCSF** | ventricular cerebro-spinal fluid |
| **VOI** | Volume-of-Interest |
| **WM** | white matter |

# Portfolio

# Portfolio

| |
|---|
| **Name PhD student:** Riaan Zoetmulder |
| **PhD period:** 2018 - 2022 |
| **Names of PhD supervisor(s) & co-supervisor(s):** Henk Marquering, Ivana Išgum & Efstratios Gavves |

**1. PhD training**

| | Year | ECTS |
|---|---|---|
| **General courses** | | |
| Project Management | 2020 | 0.6 |
| **Specific courses** | | |
| Scientific writing in English | 2019 | 1.5 |
| Entrepreneurship in the health and life sciences | 2021 | 0.5 |
| **Seminars, workshops and master classes** | | |
| **Presentations** | | |
| CV Eng, SOOS Talk, QIA/QurAI groups | - | - |
| **(Inter)national conferences** | | |
| MICCAI | 2020 | 1.4 |
| Other | - | - |

| 2. Teaching | | |
|---|---|---|
| | **Year** | **ECTS** |
| **Lecturing** | | |
| **Tutoring, Mentoring** | | |
| MAM 10 | 2018 | 2 |
| MAM 11 | 2019 | 2 |
| MAM 12 | 2020 | 2 |
| Statistics, Simulation & Optimization | 2018 | 3.5 |
| Statistics, Simulation & Optimization | 2019 | 3.5 |
| Deep Learning | 2020 | 3.5 |
| **Supervising** | | |
| Master Thesis (Mahsa Mojtahedi) | 2020 | 1.1 |
| **Other** | | |
| AI Journal Club AMC (Organisation) | 2018 - 2020 | 1.8 |
| SOOS Talk (Attendence) | 2018 - 2020 | 1.8 |

| 3. Publications | |
|---|---|
| **Peer reviewed** | **Year** |
| Automated Final Lesion Segmentation in Posterior Circulation Acute Ischemic Stroke Using Deep Learning *(Shared First Author)* | Diagnostics 2021 |
| Domain- and task-specific transfer learning for medical segmentation tasks *(First Author)* | Computer Methods & Programs in Biomedicine 2022 |
| Deep Learning Based Posterior Circulation Stroke Detection and Segmentation in Computerized Tomography Using a Moving Volume of Interest *(First Author)* | 2022 Diagnostics |
| **Submitted** | **Year** |
| Brain Segmentation in Patients with Perinatal Arterial Ischemic Stroke *( First Author)* | 2022 |

# Acknowledgements

PhD students have a reputation for being hermits that spend all of their time in a dimly lit room, slaving away at finding solutions to obscure scientific problems. Of course this is an exaggeration of life as a PhD student. In reality, PhD students work in fairly well lit rooms. They are also allowed to go out, but only to meet with their supervision team. As such, obtaining a PhD is more of a team effort than people are led to believe. Hence, I would like to extend my gratitude to the members of my supervision team: Henk A. Marquering, Efstratios Gavves & Ivana Išgum. Thank you for all the time and effort that you have invested in my personal development and research projects.

You cannot spell laboratory, without labor. This is why academics refer to laboratories as labs, such that they get associated with likeable dogs rather than hard work. I have had the luck of being a member of three labs. Unfortunately, none of them had any Labradors, but all of them had affable, smart, and supportive colleagues. These three labs were the department of biomedical engineering and physics, the qurAI group, and VIS-Lab. Here I would like to express appreciation to my fellow PhD candidates and Post-Docs. Despite the prolonged lock-downs, you have all been a very important source of support during my PhD candidacy. I hope to stay in touch with you after my PhD.

Finally, I would like to thank those that were not directly involved in the PhD process or part of any lab, but that were supportive during this time. One group of people that I would like to thank are those from the other labs for the many lunches, talks, and gym sessions that we have had. Another group that I would like to express my gratitude to are my friends, that had to put up with the stressed out and less-than-optimal version of me. Finally, I would like to thank my family that have always been supportive of me going through this process.

# About the Author

Riaan Zoetmulder was born in Amsterdam on the 19*th* of June 1991. After his secondary education he studied psychology and economics and finance at the University of Amsterdam. After obtaining his bachelors degrees, he switched to artificial intelligence and received his masterś degree cum laude in 2018. During his masters thesis he worked on the interpretability of neural networks for image classification problems. After completing his masters degree he started working as a PhD student at the Amsterdam University medical centers, location AMC. The results of which are described in this thesis.