



**HAL**  
open science

# Validating stroke lesion segmentation methods using MRI in children : transferability of deep learning models

Emma Lhermitte, Mickaël Dinomais, Rodrigo Araneda Oyaneder, Yannick Bleyenheuft, Andrea Guzetta, Sylvain Brochard, François Rousseau

## ► To cite this version:

Emma Lhermitte, Mickaël Dinomais, Rodrigo Araneda Oyaneder, Yannick Bleyenheuft, Andrea Guzetta, et al.. Validating stroke lesion segmentation methods using MRI in children : transferability of deep learning models. ISBI 2024 - 21st IEEE International Symposium on Biomedical Imaging, May 2024, Athènes, Greece. hal-04607398

**HAL Id: hal-04607398**

**<https://hal.science/hal-04607398v1>**

Submitted on 10 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# VALIDATING STROKE LESION SEGMENTATION METHODS USING MRI IN CHILDREN: TRANSFERABILITY OF DEEP LEARNING MODELS

Emma Lhermitte<sup>1,2,10</sup> Mickaël Dinomais<sup>3,4</sup> Rodrigo Araneda Oyaneder<sup>5,6</sup> Yannick Bleyenheuft<sup>5</sup>  
Andrea Guzetta<sup>7,8</sup> Sylvain Brochard<sup>1,2,9,10</sup> François Rousseau<sup>1,11</sup>

<sup>1</sup> LaTIM UMR 1101, Inserm, Brest, France

<sup>2</sup> Université de Bretagne Occidentale, Brest, France

<sup>3</sup> Université d'Angers, Laboratoire Angevin de Recherche en Ingénierie des Systèmes (LARIS) – EA7315 F-49000 France

<sup>4</sup> CHU Angers, Département de Médecine Physique et de Réadaptations, CHU Angers-Capucins, F- 49933, France

<sup>5</sup> Institute of Neuroscience, Université catholique de Louvain, Brussels, Belgium

<sup>6</sup> Exercise and Rehabilitation Science Institute, Faculty of Rehabilitation Science, Universidad Andres Bello, Santiago, Chile

<sup>7</sup> Department of Developmental Neuroscience, IRCCS Fondazione Stella Maris, Pisa, Italy

<sup>8</sup> Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy

<sup>9</sup> Fondation Ildys, Brest, France

<sup>10</sup> University Hospital of Brest, Brest, France

<sup>11</sup> IMT Atlantique, Brest, France

## ABSTRACT

Accurate segmentation of stroke lesions from paediatric brain MRI scans is a challenging task due to the heterogeneity in size, shape and texture of the injuries. Deep learning and atlas-based techniques developed for adult patients may lead to a reduced segmentation performance when applied to children because they do not account for the changes in brain shape that occur during childhood development. The objectives of this work are 2-folds: to investigate the learning transferability of stroke lesion segmentation models trained on the adult domain and applied to the children domain, and to identify the deep learning architecture that produces the most accurate segmentation results.

**Index Terms**— Stroke lesion, segmentation, deep learning

## 1. INTRODUCTION

Perinatal ischemic stroke constitutes the leading cause of unilateral cerebral palsy (CP) in term-born children [1]. The umbrella term CP regroups a large range of heterogeneous disorders of the development of movement and posture occurring early in brain development and persisting through the lifespan [2]. The arrival of non-invasive neuro-imaging techniques considerably improved the diagnostic, etiological elucidation and individualization of treatment of cerebral palsy [3]. Magnetic resonance imaging (MRI) is a powerful qualitative tool that provide detailed information about the location and extent of brain injuries. On a voxel-by-voxel basis, MRI scans can quantify the structural changes in the brain. Voxel-based morphometry (VBM) is an automatic technique that relies on statistical tests performed at

every voxel of the image to assess volumetric changes [4]. Surface-based morphometry (SBM) transforms volumetric tissue segmentation into surface meshes, to extract morphological shape metrics such as cortical thickness, gyrification indices, or sulcal depths [5]. Both methods allowed many studies to reveal that the brain undergoes a structural reorganization of the intact contra-lesional and the ipsi-lesional hemispheres after ischemic stroke [6].

Reliable extraction of these bio-markers depends on prior accurate segmentation. However, most automated morphometry tools are developed for healthy adult brains, and fail to consider the distortions caused by the lesion, their variability, and the overall morphological differences in children's brains. Pagnozzi et al. [7] highlighted the need for an effective tool for stroke lesion segmentation on preschool children brains, showing that commonly used morphometric analysis methods suffers a significant decrease in accuracy as the severity of the injury increases. A crucial aspect of morphometric studies on children's brains with stroke lesions is the segmentation (or delineation) of the lesion. To date, the gold standard remains the manual lesion tracing by experts. However, this is a time-consuming, tedious and expensive task requiring domain expertise. For the brain tumor segmentation task, [8] reported an average intra-rater variability of  $20\% \pm 15\%$  (conditioned to the time spent on the segmentation), and an average inter-rater variability of  $28\% \pm 12\%$ . With large cohorts of subjects, manual labelling can become a bottleneck in the image processing pipeline. Therefore, to reduce both workload and variability, and to speed up the image analysis, there is a clear need for easy and accurate automatic brain lesion labelling technique in children.

Typical deep learning models for stroke lesion segmentation are generally based on fully-convolutional networks

(FCN), such as U-Nets [9], with some variations of the architecture. U-Net models consist of an encoder, which extracts the features of the image, and a decoder, which reconstructs the image to its original size to output a segmentation map. Isensee et al. [10] introduced the nnU-Net, a deep learning framework that has been successfully tested on different biomedical imaging segmentation tasks, gaining in robustness, performance and generalizability, and is now considered the state of the art for bio-medical imaging. Given the lack of biomedical pediatric data, transferring knowledge from adult data is an alternative approach, yet challenging. Draai et al. [11] have shown that models trained on adult data can achieve good tumor segmentation results on children data.

To our knowledge, the ATLAS database [12] is the largest publicly available database of stroke patients with T1-weighted images manually delineated into lesion/non-lesion mask, and there is no dedicated learning base for children stroke lesion segmentation. Our aim is to validate existing stroke lesion segmentation methods based on deep learning for pediatric MRI. We studied different models performances in learning transferability from adult data to children data, and compared their results to a traditional, non-deep learning-based method results. We evaluated the impact of different factors (lesion size, training data sampling strategy, in-domain training and fine-tuning) on the segmentation results. Our secondary aim is to identify the method that produces the best accuracy.

## 2. MATERIALS AND METHODS

### 2.1. Datasets

#### 2.1.1. ATLAS dataset

ATLAS v2.0 is composed of 955 T1-weighted brain MRI of adults patients presenting acute to chronic stroke lesions. The dataset was previously separated into test images (n=300) and training images (n=655) with lesion masks traced by two specialists, using ITKsnap (Yushkevich et al., 2006). The images are all defaced, their intensity normalized, and registered into standard space (MNI-152 template). The data comes from 33 different cohorts and the lesions ranges from 0.003mL to 469mL, with an average of 26.8mL.

#### 2.1.2. CAP dataset

The CAP brain MRI dataset comes from a cohort of 110 children aged 1 to 4 with cerebral palsy resulting from various pathologies [13]. The dataset, with images acquired on 4 different sites, should provide a better understanding of pathological brain development during childhood and evaluate emerging developmental interventions for children. The children underwent MRI scans before and after intensive motor therapy, with a 90-day interval between the two sessions. In our study, we recovered the two images of the children

with a stroke lesion, and whose images that did not show excessive artifacts, for a total of 40 images (8 boys and 12 girls, mean age : 37.15 month +- 12.94). The 20 images from the session 1 were manually segmented by a medical expert, using ITKsnap. They were then registered to the session 2 to get the corresponding transformation matrix. The matrix was used to register the manually traced mask from the first session to the second session image, to get its corresponding mask. All the registered masks were visually checked. The registration steps were achieved using FSL FLIRT[14]. The lesions sizes from the CAP dataset ranges from 2.5mL to 172mL, with an average of 50.6mL.

### 2.2. Models

#### 2.2.1. U-Net

U-Nets [9] are widely used in the field of bio-medical imaging segmentation, therefore, we evaluated the performance of this architecture for our task. We relied on the Monai [15] implementation of a 3D U-Net with 5 layers with the sum of cross-entropy and DICE as loss function, a fixed learning rate at 0.01 and an Adam optimizer, implemented using PyTorch (<https://pytorch.org/>) with the default parameters.

#### 2.2.2. Pre-trained Res-U-Net

Transfer learning uses data from other domains, generally differing in data distribution or modality, to rely on a large amount of training data for models that can then be refined for specific tasks. Chen et al [16] aggregated datasets from several medical challenges to build the 3DSeg-8 dataset, composed of diverse modalities, target organs, and pathologies. 3D residual networks were trained on this dataset to create publicly available pre-trained models reusable as encoders for multiple segmentation tasks. We used a ResNet-10 pre-trained on the 3DSeg-8 dataset as an encoder of our network.

#### 2.2.3. nnU-Net

We compared U-Net architectures with the nnU-Net [10] framework. nnU-Net is an entire pipeline based on U-Net models, that automatically configures pre-processing, network architecture, training and post-processing for any new task, basing itself on the data fingerprint and network capacity (e.g. GPU limit memory). We used the standard parameters of nnU-Net.

### 2.3. Baseline

Deep learning methods were compared with LesionGnb [17], a toolbox for stroke lesion segmentation based on a gaussian bayesian classifier trained on images transformed into feature maps. The maps encode information about

the probability that the voxel belongs to missing or abnormal tissue. LesionGnb can be implemented within the SPM (Statistical Parametric Mapping) software package (<http://www.fil.ion.ucl.ac.uk/spm/>). We first segmented and normalized the images into standard space using SPM12 and then transformed the images into feature maps.

## 2.4. Implementation

To evaluate the performance of existing methods used in the adult domain when inferred to the paediatric domain, we trained all of our models on the full ATLAS dataset. The U-Net and pre-trained Res-U-Net were trained separating the ATLAS training data into 80% (n=524) training and 20% (n=131) validation, for 200 epochs. The ATLAS dataset contains a large range of lesions sizes, including very small lesions (<1mL). In contrast, the lesions of the CAP dataset are more evenly distributed between 2.5mL and 172mL. The distribution of the ATLAS dataset towards small lesions incorporates a bias in the data. Given the sensitivity of the deep learning models on the data distribution, we hypothesized that a more precise sampling strategy of the training data would improve the segmentation on the target data. We tested our hypothesis setting a threshold on the ATLAS data, that excludes lesions with a volume inferior to 2.5mL, as well as cerebellum lesions. Our initial ATLAS dataset is therefore reduced to 305 cases with an average lesion size of 42.2mL. All of our models were trained on the ATLAS subset.

We hypothesize that deep learning models performs better when they are trained on the same domain as the target. Our hypothesis was tested by training our models on 5 randomly selected CAP images, with the same settings as for adult data. In this context of limited paediatric data available, we evaluated the performance of in-domain training with models relying on much less data. Transferring knowledge from the ATLAS adult data and fine-tuning to the CAP paediatric data has the advantage of exploiting the strengths of both datasets : using the amount of data available in the ATLAS dataset while learning the domain specificities of the CAP data. To assess whether fine-tuning the weights of the models on the target domain can be used to refine segmentation maps, we used the above-mentioned models initialized with the weights learnt on the ATLAS full dataset and subset, and then fine-tuned with an additional training on 5 images of the CAP dataset for 100 epochs.

For all training procedures, the data was preprocessed and randomly augmented, with elastic deformations and addition of Gaussian noise, bias field artefacts, or motion artefacts on the images.

The segmentation masks were evaluated using DICE similarity coefficient (DSC) and Hausdorff distance (HD), with the manually traced masks as ground truth. The DSC measures the overlap between two segmentation maps and ranges from 0 (no overlap) to 1 (complete overlap). The HD mea-

sures the maximum Euclidean distance between the prediction and the GT and is, hence, especially sensitive to outliers. A smaller HD indicates a better prediction.

## 3. RESULTS

Table 1 shows the different average DSC and HD obtained with the 4 methods on the ATLAS validation dataset and the CAP dataset.

**Table 1.** Average DSCs  $\uparrow$  and HDs  $\downarrow$  of lesion masks of adult (ATLAS) and children (CAP) datasets. \* Obtained from [18]

	ATLAS		CAP			
	DSC	HD	no fine-tuning		fine-tuning	
	DSC	HD	DSC	HD	DSC	HD
<b>LesionGnb</b>	0.42*	58.19*	0.34	63.52		
Training on full ATLAS						
<b>U-Net</b>	0.45	75.10	0.43	96.88	0.63	85.80
<b>Res-U-Net</b>	0.44	100.47	0.40	90.24	0.67	26.75
<b>nnU-Net</b>	0.61	43.17	0.49	63.18	<b>0.73</b>	27.17
Training on subset of ATLAS						
<b>U-Net</b>	0.54	40.04	0.40	107.22	0.70	46.62
<b>Res-U-Net</b>	0.68	62.18	0.48	65.09	0.63	<b>26.15</b>
<b>nnU-Net</b>	<b>0.79</b>	<b>27.41</b>	<b>0.57</b>	<b>37.95</b>	0.71	30.23

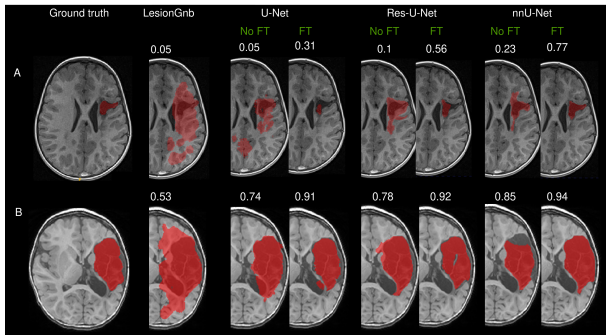
The best results are obtained with the nnU-Net model, meaning that a robust pre-processing of the training images, with data augmentation, and a fine choice of model parameters can have a large impact on the segmentation results. We observe a global reduced performance of models trained on adult data when applied to children data, and high HD for the U-Net and Res-U-Net which indicates the presence of outliers. These scores highlight the differences in intensity distribution or shape between children and adult brain images, which makes the features extraction task and voxel classification difficult for the models. The DSCs obtained from the models on the ATLAS dataset are consistent with those found in the literature for more complex architectures, where the scores generally ranges from 0.47 to 0.67 on ATLAS v1.2 [12]. nnU-Net DSC is located at the high end of the range, even though it is based on a classic U-Net model. All of the models trained on the ATLAS subset increase their performance on adult and children data segmentation (except for the U-Net), but we still observe a reduced performance when we infer the models on children data. These results demonstrate that a selection of the training data according to the target domain data distribution is a good strategy to improve segmentation performance, especially in a context of limited amount of data.

The training on a few CAP images did not improve the segmentation scores of the U-Net, as indicated in Table 2.

**Table 2.** Comparison of the mean DSCs and HDs of lesion masks from CAP dataset with models trained on CAP dataset.

	DICE	HD
<b>U-Net</b>	0.42	44.97
<b>Res-U-Net</b>	0.50	84.00
<b>nnU-Net</b>	0.70	25.5

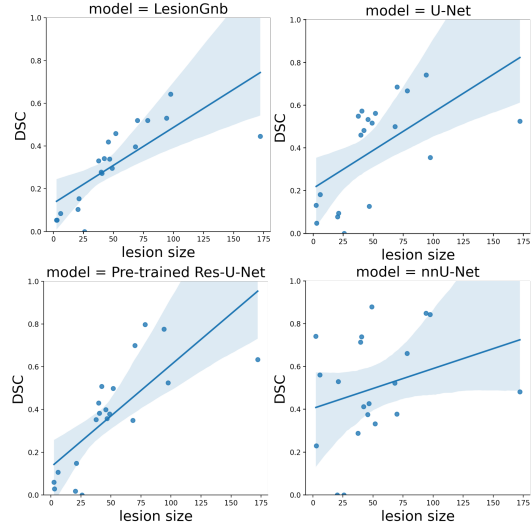
The Res-U-Net, being pre-trained on a large amount of medical data, is not impacted by the small training sample, compared to the U-Net, that learnt the segmentation from scratch. It shows that the U-Net needs more training data to learn from diverse lesions and brains. However, the HD decreases which indicates less outliers and more precise lesion localisation. nnU-Net is not impacted by the small training sample and can extract precise features from only 5 training images. The Table 1 shows that the fine-tuning phase on the CAP dataset as a strong positive impact on the segmentation results for all methods. These results demonstrate that the models can extract good foundational knowledge from adult data, but still need to learn specific target domain features to perform best.



**Fig. 1.** Segmentation maps obtained from models trained on the full ATLAS dataset, overlaid on brain MRI axial slices, with corresponding DSC. (A) : low DSCs (lesion of 2.9mL) and (B) high DSCs (lesion of 94.1mL). FT : Finetuning.

Figure 1 shows examples of segmentation of low and high DSCs. LesionGnb, U-Net and Res-U-Net methods have a tendency to segment outside of the lesion, yet, they can localise the lesion in the brain. nnU-Net seems more robust and precise to delineate the lesion, but often under-estimates its size, missing some parts of large lesions. These results show that stroke lesions complicate neural network learning due to their heterogeneity in size, shape and location.

We studied the impact of the lesion size on the segmentation performance, comparing DSC with the size of the lesion. The results can be found on Figure 2. We notice a major decrease in performance for the small lesions for all methods, which might have been biased by the sensitivity of the DSC. However, the nnU-Net is more robust to the lesion size.



**Fig. 2.** Effect of the lesion sizes on the CAP images DSC.

#### 4. CONCLUSION

We evaluated methods to transfer stroke segmentation knowledge acquired from adult data to children data, according to the DSC and HD metrics. A characteristic of stroke lesions being their heterogeneity, it remains difficult for segmentation methods to generalize within a single domain. The difficulty increases when transferred to another domain, i.e. children data imaging. Experimental results show that the best performance is obtained with nnU-Net, thanks to its robust preprocessing steps. We revealed that extracting features from adult data may yield acceptable results on pediatric stroke segmentation tasks, and manually segmenting a few images from the target domains to fine-tune the models can considerably improve the segmentation results.

#### 5. COMPLIANCE WITH ETHICAL STANDARDS

Full ethical approval has been obtained from the involved centers; Belgium (B403201316810, Comité d’éthique Hospitalo-Facultaire/Université catholique de Louvain, Brussels), France (29BRC19.0050/N2019-A01173-54, Comité de Protection des personnes, Ile de France) and Italy (244/2019, Comitato Etico Regionale per la Sperimentazione Clinica della Regione Toscana. Sezione: Comitato Etico Pediatrico).

#### 6. ACKNOWLEDGMENTS

The research leading to these results has received funding from the “La Fondation Paralysie Cérébrale” (CAP’ project) and ANR (AI4CHILD ANR-19-CHIA-0015-01). The authors would like to thank to “La Fondation Paralysie Cérébrale” (France) for their support with this project. Members of Early HABIT-ILE group: Julie Paradis, Daniela

Ebner-Karestinos, Geoffroy Saussez, Anne Klöcker, Rodolphe Bailly, Sandra Bouvier, Josselin Demas, Inma Riquelme, Enora Trahmel.

## 7. REFERENCES

- [1] Adam Kirton and Gabrielle Deveber, “Life after perinatal stroke,” *Stroke*, vol. 44, no. 11, pp. 3265–3271, 2013.
- [2] Peter Rosenbaum, Nigel Paneth, Alan Leviton, Murray Goldstein, Martin Bax, Diane Damiano, Bernard Dan, Bo Jacobsson, et al., “A report: the definition and classification of cerebral palsy april 2006,” *Dev Med Child Neurol Suppl*, vol. 109, no. suppl 109, pp. 8–14, 2007.
- [3] Martin Bax, Clare Tydeman, and Olof Flodmark, “Clinical and mri correlates of cerebral palsy: the european cerebral palsy study,” *JAMA*, vol. 296, no. 13, pp. 1602–1608, 2006.
- [4] John Ashburner and Karl J Friston, “Voxel-based morphometry—the methods,” *Neuroimage*, vol. 11, no. 6, pp. 805–821, 2000.
- [5] Robert Dahnke, Rachel Aine Yotter, and Christian Gaser, “Cortical thickness and central surface estimation,” *Neuroimage*, vol. 65, pp. 336–348, 2013.
- [6] Karan Shinde, Brandon T Craig, Jordan Hassett, Nomazulu Dlamini, Brian L Brooks, Adam Kirton, and Helen L Carlson, “Alterations in cortical morphometry of the contralesional hemisphere in children, adolescents, and young adults with perinatal stroke,” *Scientific Reports*, vol. 13, no. 1, pp. 11391, 2023.
- [7] Alex M Pagnozzi, Yaniv Gal, Roslyn N Boyd, Simona Fiori, Jurgen Fripp, Stephen Rose, and Nicholas Dowson, “The need for improved brain lesion segmentation techniques for children with cerebral palsy: a review,” *International Journal of Developmental Neuroscience*, vol. 47, pp. 229–246, 2015.
- [8] Gloria P Mazzara, Robert P Velthuisen, James L Pearlman, Harvey M Greenberg, and Henry Wagner, “Brain tumor target volume determination for radiation treatment planning through automated mri segmentation,” *International Journal of Radiation Oncology\*Biophysics\*Physics*, vol. 59, no. 1, pp. 300–312, 2004.
- [9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.
- [10] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein, “nnu-net: a self-configuring method for deep learning-based biomedical image segmentation,” *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.
- [11] Maxime Draï, Benoit Testud, Gilles Brun, Jean-François Hak, Didier Scavarda, Nadine Girard, and Jan-Patrick Stellmann, “Borrowing strength from adults: Transferability of ai algorithms for paediatric brain and tumour segmentation,” *European Journal of Radiology*, vol. 151, pp. 110291, 2022.
- [12] Sook-Lei Liew, Bethany P Lo, Miranda R Donnelly, Artemis Zavaliangos-Petropulu, Jessica N Jeong, Giuseppe Barisano, Alexandre Hutton, Julia P Simon, Julia M Juliano, Anisha Suri, et al., “A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms,” *Scientific Data*, vol. 9, no. 1, pp. 320, 2022.
- [13] R Araneda, SV Sizonenko, CJ Newman, Mickaël Dinomais, G Le Gal, E Nowak, A Guzzetta, I Riquelme, S Brochard, Yannick Bleyenheuft, et al., “Functional, neuroplastic and biomechanical changes induced by early hand-arm bimanual intensive therapy including lower extremities (e-habit-ile) in pre-school children with unilateral cerebral palsy: study protocol of a randomized control trial,” *BMC neurology*, vol. 20, pp. 1–10, 2020.
- [14] Stephen M Smith, Mark Jenkinson, Mark W Woolrich, Christian F Beckmann, Timothy EJ Behrens, Heidi Johansen-Berg, Peter R Bannister, Marilena De Luca, Ivana Drobnjak, David E Flitney, et al., “Advances in functional and structural mr image analysis and implementation as fsl,” *Neuroimage*, vol. 23, pp. S208–S219, 2004.
- [15] M Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murrey, Andriy Myronenko, Can Zhao, Dong Yang, et al., “Monai: An open-source framework for deep learning in healthcare,” *arXiv preprint arXiv:2211.02701*, 2022.
- [16] Sihong Chen, Kai Ma, and Yefeng Zheng, “Med3d: Transfer learning for 3d medical image analysis,” *arXiv preprint arXiv:1904.00625*, 2019.
- [17] Joseph C Griffis, Jane B Allendorfer, and Jerzy P Szaflarski, “Voxel-based gaussian naïve bayes classification of ischemic stroke lesions in individual t1-weighted mri scans,” *Journal of neuroscience methods*, vol. 257, pp. 97–108, 2016.
- [18] Kaori L Ito, Hosung Kim, and Sook-Lei Liew, “A comparison of automated lesion segmentation approaches for chronic stroke t1-weighted mri data,” *Human brain mapping*, vol. 40, no. 16, pp. 4669–4685, 2019.