Heterogeneous Learning for Brain Lesion Segmentation, Detection, and Classification

Sebastian Nørgaard Llambias*, Mads Nielsen, and Mostafa Mehdipour Ghazi

Pioneer Centre for Artificial Intelligence, Department of Computer Science, University of Copenhagen

Abstract

Brain lesions detected in magnetic resonance images often vary in type and rarity across different cohorts, posing a challenge for deep learning techniques that are typically specialized in recognizing single lesion types from homogenous data. This limitation restricts their practicality in diverse clinical settings. In this study, we explore different deep-learning approaches to develop robust models handling both subject and imaging variability, while recognizing multiple lesion types. Our research focuses on segmentation and detection tasks across four distinct datasets, encompassing six cohorts of subjects with white matter hyperintensities, multiple sclerosis lesions, or stroke abnormalities. Our findings reveal that a cascade approach, comprising a fully convolutional network and a fully connected classifier, offers optimal accuracy for robust multiclass lesion segmentation and detection. Notably, our proposed model remains competitive with models trained solely on one dataset and applied to the same dataset while showing robustness against domain shifts. Additionally, in related tasks, our model consistently produces results comparable with the state-of-the-art methods. This study contributes to advancing clinically applicable deep learning techniques for brain lesion recognition, offering a promising solution for handling lesion diversity in uncontrolled clinical environments.

1 Introduction

Brain image analysis utilizing deep learning is currently revolutionizing the way we assess data from neurological disorders [1]. This technology assumes a pivotal role in performing three fundamental tasks: segmentation, detection, and classification, with the overarching objective of precisely delineating diverse brain structures and lesions, identifying abnormalities, and helping with diagnoses based on the found characteristics and segmented regions [2]. Nevertheless, this domain confronts several formidable challenges, encompassing the necessity for large annotated data, model interpretability, and generalization to diverse populations [3]. Addressing these challenges is crucial to harness the full potential of deep learning in brain image analysis, ultimately improving patient outcomes.

One significant challenge in this domain pertains to the imperative of training deep learning models on multiple heterogeneous datasets. For instance, magnetic resonance imaging (MRI) data exhibits substantial variations in acquisition protocols, image quality, and demographic attributes. Consequently, the development of robust models capable of learning from diverse sources, while maintaining consistently high accuracy, becomes an urgent need [4]. Moreover, achieving model generalization to rare diseases or labels, as well as the detection of uncommon brain lesions, poses another obstacle [5]. Data acquisition with high variability for training and validation purposes from neurological conditions, many of which are rare, is problematic.

The segmentation of brain lesions encompassing a variety of abnormalities such as stroke, white matter (WM), and multiple sclerosis (MS), stands as a pivotal application of deep learning within brain imaging. White matter hyperintensities, often associated with neurodegenerative diseases, have seen remarkable progress with convolutional neural networks, which can achieve a Dice similarity coefficient (DSC) of up to 0.8, underscoring the potential of deep learning for brain lesion recognition [6]. MS lesions present a challenge in terms of accurate segmentation due to their considerable variability. For instance, deep networks applied to MS lesion segmentation have demonstrated DSC values surpassing 0.6 [7]. Furthermore, stroke lesion segmentation is a particularly challenging task, because of significant variability in lesion size, shape, and intensity. The current state-of-the-art methods utilizing 3D U-Net models [8, 9] achieve DSCs exceeding 0.3 when applied to the segmentation of ischemic lesions based on multimodal MRIs [10, 11]. These outcomes illustrate the intricacies of detecting brain lesions due to their diversity, small size, and class imbalance.

In the context of medical image segmentation, the reported DSC values for brain lesion segmentation fall within the range often characterized as poor for the stroke lesion type with a DSC less than 0.5, moderate for the MS lesion type with a DSC between 0.5 and 0.7, and good for WMH type with a DSC from 0.7 to 0.9 in the literature. However, it is important to note that the tasks contain a highly heterogeneous set of lesions, including irregularly shaped

*Corresponding Author: snl@di.ku.dk

Proceedings of the 5th Northern Lights Deep Learning Conference (NLDL), PMLR 233, 2024.
2024 Sebastian Nørgaard Llambias, Mads Nielsen, & Mostafa Mehdipour Ghazi. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

and overlapping lesions, which pose significant segmentation challenges. Hence, the categorization of Dice scores should depend on the specific application and the clinical goals, where a DSC below 0.5 may reflect the complexity of the data/task and the potential clinical utility of the segmentation method in challenging clinical scenarios.

Despite notable achievements of deep learning models in medical imaging, attaining high-level accuracy remains a challenge in the context of small or subtle lesion segmentation. To this end, the primary objectives of this article are to develop and evaluate robust deep learning models for brain lesion segmentation and characterization, with a specific emphasis on addressing the challenges posed by imbalanced datasets, data variability in subjects and images, and the recognition of multiple lesion types. Our study makes four key contributions: 1. We develop and train a set of robust deep-learning models to address the challenges associated with brain lesion segmentation, including issues related to dataset imbalances, data variability, and the recognition of diverse lesion types. 2. We explore constructive strategies, such as MRI-specific data augmentation and balanced dataset training, to enhance the model's ability to generalize across different datasets, thus improving its robustness. 3. We examine multiclass and cascade model training to effectively segment and detect brain lesions across four heterogeneous datasets and tasks, spanning six cohorts of subjects with WM lesions, MS lesions, or stroke abnormalities. 4. We demonstrate the advantages of our proposed cascade approach, which combines a fully convolutional network with a fully connected classifier, in achieving robust multiple lesion recognition. These approaches enhance the model's performance and its potential for clinical applications.

2 Study data

The study utilizes fluid-attenuated inversion recovery (FLAIR) images from 4 resources spanning 3 lesion types and 6 cohorts. These images were supported by manual segmentations provided by expert annotators. The first two datasets are small and contain MS lesions. These datasets are obtained from the MICCAI 2016 MS Lesion Segmentation Challenge (MSSEG-1) [12] and the 2008 MICCAI MS Lesion Segmentation Challenge (MS08) [13]. The third dataset is obtained from the MICCAI 2022 Ischemic Stroke Lesion Segmentation Challenge (ISLES22) [14] and contains brain infarcts in acute and sub-acute stroke. The last dataset includes the multi-vendor and multi-site MICCAI 2017 WMH Segmentation Challenge [15], which introduces white matter hyperintensities (WMH) in three distinct cohorts from Amsterdam, Singapore, and Utrecht. A summary of the utilized datasets is shown in Table 1. We use the original training/test splits from MSSEG-1 and WMH datasets for model training and inference while using 25% of the MS08 and ISLES22 training data for testing as the test labels were not released.

Table 1. Overview of the datasets used in this study. Country codes are ISO Alpha-2 and vendor abbreviations refer to GE (G), Philips (P), and Siemens (S).

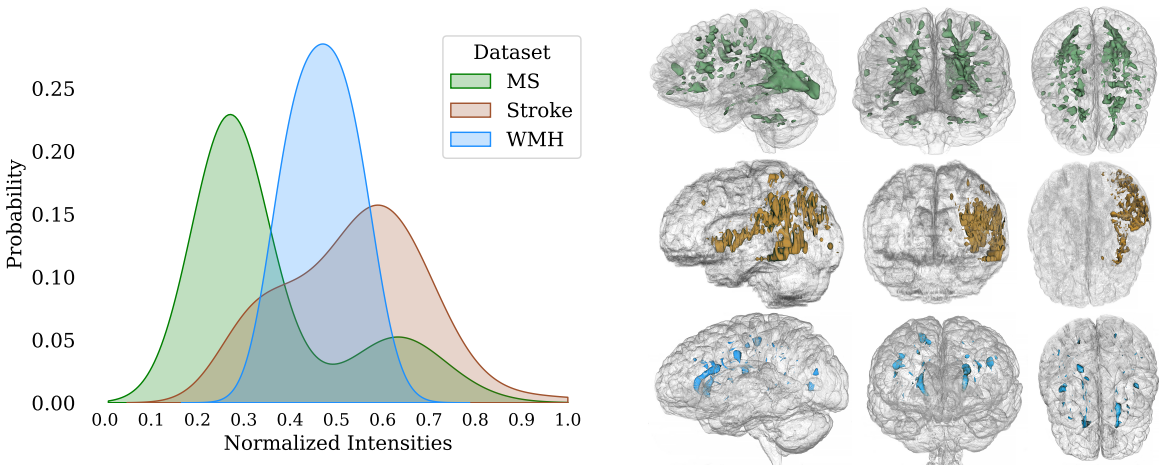
Dataset	Task	Train/Test	Site	Strength	Vendor
MSSEG-1 [12]	MS Lesion	15/38	FR	1.5-3T	G, P, S
MS08 [13]	MS Lesion	14/5	US	3T	S
ISLES22 [14]	Stroke Lesion	187/63	CH, DE	1.5-3T	P, S
WMH [15]	WM Lesion	60/110	NL, SG	1.5-3T	G, P, S

The datasets include several scanner vendors, acquisition parameters, and imbalanced cohorts, e.g., from 19 cases in MS08 to 250 cases in ISLES22. This mimics the presence of rare neurological conditions. Moreover, the imaging attributes are highly varied in both intra- and inter-dataset with slice thickness ranging from 0.5 mm to around 5.0 mm and voxel dimensions ranging from 128×128×25 to 512×512×512. To highlight the differences between the study tasks, Figure 1 (a) presents the overall distribution of lesion intensities among all scans and (b) exemplifies one scan from each lesion type with high lesion loads, illustrating the heterogeneity of the loads, intensities, locations, and densities of different lesion types.

Finally, Table 2 presents statistics of the lesions per scan across various datasets. The table highlights the substantial variability in both the number of connected regions (lesion load) and the lesion volumes observed in different scans. Besides, it highlights distinctions within cohorts featuring the same lesion type, as exemplified by the significant variations in lesion loads and volumes observed in MSSEG-1 and MS08. In summary, the domain shifts discussed in the study span intra- and inter-dataset variations in the cohort (anatomies and demographics), pathologies (lesion types, shapes, and loads), and acquisition parameters (vendors, field strengths, resolutions, and contrast).

3 Methods

Three main approaches are used to address the domain shift problem inherent in the utilized heterogeneous datasets and tasks: balanced dataset training, MRI-specific data augmentation, and multiclass and cascade model training. Balanced dataset training is applied using a weighted sampler to draw scans uniformly from each task to fine-tune the baseline model in a balanced way. Without the weighted sampler fine-tuned models showed a tendency to learn to ignore the underrepresented lesions and specialize on, e.g., WMH. The MRI augmentation



(a) Distribution of the lesion intensities

(b) Scans with high lesion loads

Figure 1. The probability density of the lesion intensities across all scans (a) and a multi-view visualization of three different scans with high loads of various lesion types (b).

Table 2. Statistics of the lesions per scan across various datasets.

Dataset	# Lesions (mean+SD)	# Lesions (min/max)	Volume in mm^3 (mean+SD)	Volume in mm^3 (min/max)	# Scans with no lesions
MSSEG-1 [12]	40 ± 29	7/111	848 ± 6090	1/105710	0
MS08 [13]	22 ± 20	5/90	1617 ± 4665	12/44846	0
ISLES22 [14]	9 ± 11	1/66	365 ± 2329	1/59658	2
WMH [15]	61 ± 32	16/152	98 ± 713	1/14396	0

technique applies a range of transformations on the fly during training, including generic augmentations such as 3D rotation, elastic deformation, and additive noise along with augmentations specific to and commonly found in MRIs such as bias fields and motion artifacts. The technique is used to obtain scans with more realistic variations and has been shown to train more robust deep learning models in previous studies [6, 16–18].

Three models are developed, referred to as the baseline, multiclass, and cascade models. We first train the baseline model by learning from all four datasets for binary lesion segmentation by combining the three foreground labels from the MS, Stroke, and WMH tasks into one universal lesion label, disregarding the task and dataset size. Next, to train the multiclass model, we fine-tune all layers of the baseline model using the weighted sampler for multiclass (Background, MS, Stroke, and WMH) segmentation to classify the pixels at once. Afterwards, to create the cascade model, we fine-tune all layers of the baseline model using the weighted sampler for binary segmentation. A fully connected network is then trained on quantitative measures obtained from the segmentations. The combination of the fine-tuned binary model and the fully connected network is referred to as the cascade model.

The quantitative measures are calculated based on a set of properties for each connected component in the 3D regions. To do this, we first identify connected components in the foreground regions using 6-voxel minimal connectivity, in which the voxels would be part of the same object (lesion) if their faces touch in one of the 6 directions: in, out, left, right, up, and down. Next, we calculate the volumes (number of the voxels) and surface area (the distance around the boundary), as well as the average centroids (center of mass), orientation (Euler angles), and principal axis length (length of the major axes of the ellipsoid) of the connected regions and concatenate them as a feature vector. Finally, we train a network using two fully connected layers with 400 nodes and a Softmax layer in the output to obtain the probability scores associated with the lesion classes. The network is optimized using a weighted cross-entropy loss and Adam optimizer for 100 epochs with a learning rate of 0.1.

In all experiments, we use a backbone deep learning architecture based on the U-Net [8, 9]. Models are trained on a single A100 GPU for 1000 iterations, with one iteration consisting of 250 batches and a batch size of 64 slices. The initial learning rate is set to 0.001, gradually decreasing by a cosine annealing schedule. A combination of Dice loss and cross-entropy loss is used for optimization with the stochastic gradient descent algorithm. Initially, we utilized 2D, 2.5D, and 3D networks, but the test results revealed 2D models to achieve higher Dice scores and be more robust. Hence, we selected 2D networks for conducting the final experiments.

Models are evaluated in three tasks: lesion detection, segmentation, and classification. Lesion

Table 3. Scan classification, lesion detection, and segmentation performance (mean ± SD) of the cascade model on different test sets.

	Scan Classification			Lesion Detection			Segmentation	
Dataset	**Precision**	**Sensitivity**	**F1**	**Precision**	**Sensitivity**	**F1**	**Dice**	**VS**
MS	0.875	0.833	0.854	0.590±0.33	0.474±0.26	0.496±0.25	0.499±0.26	0.671±0.32
Stroke	0.828	0.841	0.835	0.416±0.36	0.386±0.34	0.335±0.26	0.324±0.31	0.551±0.36
WMH	0.981	0.927	0.953	0.772±0.23	0.671±0.22	0.686±0.21	0.691±0.22	0.823±0.25

Table 4. Scan classification, lesion detection, and segmentation performance (mean ± SD) of the multiclass model on different test sets.

	Scan Classification			Lesion Detection			Segmentation	
Dataset	**Precision**	**Sensitivity**	**F1**	**Precision**	**Sensitivity**	**F1**	**Dice**	**VS**
MS	0.827	1.000	0.905	0.690±0.29	0.421±0.20	0.459±0.16	0.451±0.19	0.620±0.23
Stroke	1.000	0.857	0.923	0.433±0.37	0.334±0.31	0.310±0.24	0.320±0.30	0.538±0.36
WMH	1.000	1.000	1.000	0.825±0.13	0.596±0.18	0.667±0.12	0.709±0.12	0.846±0.14

detection is the task of detecting individual lesions, where each lesion is delimited and identified based on connected components analysis of the segmented areas. A lesion is considered as detected if there is any overlap between the voxels of the ground-truth lesion and any predicted lesion. Classification is the task of classifying the primary pathology or major lesion type in a given scan and is obtained from the segmented areas based on the type of lesion with the largest volume in each scan. Finally, segmentation is the task of labeling each voxel of the scan.

In this study, we assess the accuracy of the models based on predictions from full scans or 3D volumes, where the 3D segmentations are obtained by concatenating the segmentation results from 2D models applied sequentially to each slice of the axial plane of the MRI scan. Five metrics are used for a thorough evaluation of the tasks: precision, sensitivity, F1 measure, Dice score, and volume similarity (VS). While the Dice coefficient measures the spatial similarities or overlaps between voxels of the ground-truth labels (T) and the predicted ones (Y) using $2\frac{|Y \cap T|}{|Y|+|T|}$, VS calculates the absolute difference between the volumes of the ground truth labels and predicted ones using $1 - \frac{||Y|-|T||}{|Y|+|T|}$.

4 Results

The segmentation and detection results from the cascade and multiclass models are shown in Table 3 and Table 4, respectively. We also predicted the major pathology for a given scan and evaluated it as the scan classification performance, where the multiclass model performed almost perfectly. This indicates the models have learned to distinguish between the lesions. For multiple lesion detection and segmentation tasks, the cascade model achieves higher sensitivity and F1-score while the multiclass model obtains a higher precision.

Table 5. Lesion segmentation performance (average per scan) of different data-specific models on different test sets vs. the fine-tuned binary model.

	Data-specific models		Binary model	
Dataset	**Dice**	**VS**	**Dice**	**VS**
MSSEG-1	0.635±0.20	0.819±0.20	0.590±0.17	0.809±0.15
MS08	0.347±0.16	0.742±0.28	0.363±0.17	0.623±0.22
ISLES22	0.310±0.30	0.514±0.35	0.336±0.30	0.554±0.33
WMH	0.772±0.10	0.918±0.09	0.743±0.11	0.888±0.09

To better interpret the results, Figure 2 compares the distributions of the Dice scores and detection F1 scores obtained by the cascade and multiclass models. It demonstrates the cascade model obtaining higher scores for both metrics in all cases, but it also depicts the difficulty of the stroke task. Besides, Figure 3 showcases some samples from the test set segmented by the cascade model against the ground truths. These samples were selected from the subjects with high and low lesion loads, representing the variability of both scans and lesion types. These results indicate the problem is diverse and challenging and the proposed model achieves a reasonable segmentation accuracy (orange areas) with low false positives (red areas).

To finalize our study, we additionally trained data-specific models and applied them to different test sets. A comparison between the results shown in Table 5 for the fine-tuned binary model (cascade model without the classifier head) and the data-specific models trained to exclusively solve a specific task, reveals the impact of heterogeneous learning in difficult settings. In the MS08 data, the model likely transfers knowledge from the related MSSEG-1 data. The ISLES22 data is only loosely related to the other datasets, where the model learns to extrapolate from heterogeneous data.

The results presented in this study show that the proposed models can obtain Dice scores compet-

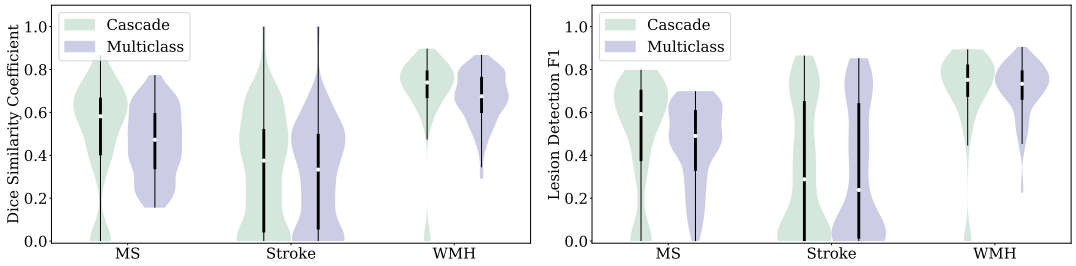


Figure 2. The violin plots for the lesion segmentation (left) and detection (right) accuracies for different test sets using the cascade and multiclass models.

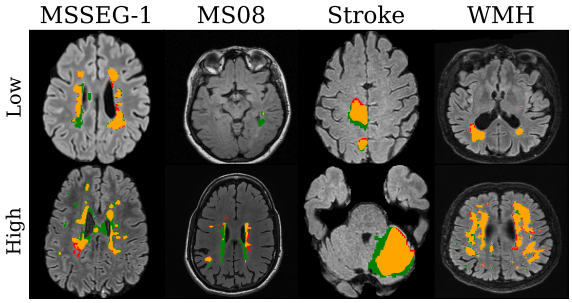


Figure 3. Test samples with high/low lesion loads, segmented by the cascade model against the ground truths. The green area is the ground truth (non-overlapping), the orange area is the overlapping segments between the ground truths and predicted labels, and the red area shows false positives.

itive with both the data-specific models and the state-of-the-art, as presented in Section 1. While remaining robust to large domain shifts, the binary model achieves higher Dice scores in 2/4 datasets and exceeds the state-of-the-art DSCs for the stroke while being within 0.01-0.06 for MS and WMH.

5 Discussion

A plausible explanation for the data-specific models surpassing the fine-tuned binary model in MSSEG-1 and WMH is that these datasets have higher quality in larger quantities. With curated homogeneous datasets, several variations found in the test set may also be expressed in the training set, like the noise introduced in the augmentation and heterogeneous training, which could be detrimental.

While achieving competitive results, the data-specific models are not a feasible alternative to the cascade or multiclass approaches and largely fail to generalize to other datasets, displaying average Dice scores below 0.1, except for the model trained on WMH that generalizes well to the MS task. The remaining domain shifts incurring serious failures, often seen when deep learning models are applied to unknown domains, could be alleviated by the ensemble of models. However, this will require intelligent selection strategies to choose the correct segmentation on a case-by-case basis, where a model may produce a good segmentation while the remaining ones obtain bad segmentations when applied to unseen data. These challenges severely hamper the applicability of both data-specific models and ensemble approaches.

6 Conclusion

In this work, we examined several deep learning-based approaches for heterogeneous learning from different brain MRI datasets for lesion segmentation and detection. Overall, the proposed cascade model was found superior for detecting and classifying diverse brain lesions from different cohorts. It also proved to learn complex relations between lesions and provided strides toward high-level accuracy in realistic settings. These provide evidence that robust deep learning models, capable of handling multiple heterogeneous datasets and lesion types, can achieve results resembling the state-of-the-art models trained exclusively to solve a single task.

The results from the dataset-specific models affirmed the need to bridge the gap between datasets and to learn more from weakly related tasks, as was the case in the WMH to MS lesion generalization. In a comparison with the state-of-the-art results and related work, the cascade model's segmentation accuracy was shown to be competitive, and in some cases, even superior [19–21].

Finally, it is worth mentioning that although deep learning models have shown remarkable promise in improving the segmentation process in a fast and accurate manner, it is imperative to address the challenges related to robustness, generalization, and the need for large and diverse datasets to harness their full potential in clinical practice and reduce failures that can have significant implications for the study outcomes and downstream applications such as clinical diagnosis and treatment. Besides, it is essential to have rigorous validation and quality control processes in place and to consider human expert input when necessary to verify or correct automated segmentations.

Acknowledgments

This project has received funding from Innovation Fund Denmark under grant number 1063-00014B, Lundbeck Foundation with reference number R400-2022-617, and Pioneer Centre for AI, Danish National Research Foundation, grant number P1.

References

[1] 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". In: *Medical Image Analysis* 42 (2017), pp. 60–88. DOI: 10.1016/j.media.2017.07.005.

[2] H. Jiang, Z. Diao, T. Shi, Y. Zhou, F. Wang, W. Hu, X. Zhu, S. Luo, G. Tong, and Y.-D. Yao. "A review of deep learning-based multiple-lesion recognition from medical images: classification, detection and segmentation". In: *Computers in Biology and Medicine* (2023), p. 106726. DOI: https://doi.org/10.1016/j.compbiomed.2023.106726.

[3] A. S. Lundervold and A. Lundervold. "An overview of deep learning in medical imaging focusing on MRI". In: *Zeitschrift für Medizinische Physik* 29.2 (2019), pp. 102–127. DOI: 10.1016/j.zemedi.2018.11.002.

[4] M. E. Mayerhoefer, P. Szomolanyi, D. Jirak, A. Materka, and S. Trattnig. "Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: an application-oriented study". In: *Medical Physics* 36.4 (2009), pp. 1236–1243. DOI: 10.1118/1.3081408.

[5] E. D. Bigler, T. J. Abildskov, J. Petrie, T. J. Farrer, M. Dennis, N. Simic, H. G. Taylor, K. H. Rubin, K. Vannatta, C. A. Gerhardt, et al. "Heterogeneity of brain lesions in pediatric traumatic brain injury". In: *Neuropsychology* 27.4 (2013), p. 438. DOI: 10.1037/a0032837.

[6] A. Schiavone, S. N. Llambias, J. Johansen, S. Ingala, A. Pai, M. Nielsen, and M. Mehdipour Ghazi. "Robust Identification of White Matter Hyperintensities in Uncontrolled Settings Using Deep Learning". In: *Medical Imaging with Deep Learning, Short Paper Track*. 2023.

[7] A. Danelakis, T. Theoharis, and D. A. Verganelakis. "Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging". In: *Computerized Medical Imaging and Graphics* 70 (2018), pp. 83–100. DOI: 10.1016/j.compmedimag.2018.10.002.

[8] O. Ronneberger, P. Fischer, and T. Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer. 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[9] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature Methods* 18.2 (2021), pp. 203–211. DOI: 10.1038/s41592-020-01008-z.

[10] S. Winzeck, A. Hakim, R. McKinley, J. A. Pinto, V. Alves, C. Silva, M. Pisov, E. Krivov, M. Belyaev, M. Monteiro, et al. "ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI". In: *Frontiers in Neurology* 9 (2018), p. 679. DOI: 10.3389/fneur.2018.00679.

[11] X. Hu, W. Luo, J. Hu, S. Guo, W. Huang, M. R. Scott, R. Wiest, M. Dahlweid, and M. Reyes. "Brain SegNet: 3D local refinement network for brain lesion segmentation". In: *BMC Medical Imaging* 20 (2020), pp. 1–10. DOI: 10.1186/s12880-020-0409-2.

[12] O. Commowick, A. Istace, M. Kain, B. Laurent, F. Leray, M. Simon, S. C. Pop, P. Girard, R. Ameli, J.-C. Ferré, et al. "Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure". In: *Scientific Reports* 8.1 (2018), p. 13650. DOI: 10.1038/s41598-018-31911-7.

[13] M. Styner, J. Lee, B. Chin, M. Chin, O. Commowick, H. Tran, S. Markovic-Plese, V. Jewells, and S. Warfield. "3D segmentation in the clinic: A grand challenge II: MS lesion segmentation". In: *MIDAS Journal* 2008 (2008), pp. 1–6. DOI: 10.54294/lmkqvm.

[14] M. R. Hernandez Petzsche, E. de la Rosa, U. Hanning, R. Wiest, W. Valenzuela, M. Reyes, M. Meyer, S.-L. Liew, F. Kofler, I. Ezhov, et al. "ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset". In: *Scientific Data* 9.1 (2022), p. 762. DOI: 10.1038/s41597-022-01875-5.

[15] H. J. Kuijf, J. M. Biesbroek, J. De Bresser, R. Heinen, S. Andermatt, M. Bento, M. Berseth, et al. "Standardized Assessment of Automatic Segmentation of White Matter Hyperintensities and Results of the WMH Segmentation

Challenge". In: *IEEE Transactions on Medical Imaging* 38.11 (2019), pp. 2556–2568. DOI: 10.1109/TMI.2019.2905770.

[16] M. Mehdipour Ghazi and M. Nielsen. "FAST-AID Brain: Fast and Accurate Segmentation Tool using Artificial Intelligence Developed for Brain". In: *arXiv preprint arXiv:2208.14360* (2022). DOI: 10.48550/arXiv.2208.14360.

[17] N. R. Ferrer, M. Vendela Sagar, K. V. Klein, C. Kruuse, M. Nielsen, and M. Mehdipour Ghazi. "Deep Learning-Based Assessment of Cerebral Microbleeds in COVID-19". In: *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. 2023, pp. 1–4. DOI: 10.1109/ISBI53787.2023.10230832.

[18] S. N. Llambias, M. Nielsen, and M. Mehdipour Ghazi. "Data Augmentation-Based Unsupervised Domain Adaptation In Medical Imaging". In: *arXiv preprint arXiv:2308.04395* (2023). DOI: 10.48550/arXiv.2308.04395.

[19] R. Guerrero, C. Qin, O. Oktay, C. Bowles, L. Chen, R. Joules, R. Wolz, M. d. C. Valdés-Hernández, D. A. Dickie, J. Wardlaw, et al. "White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks". In: *NeuroImage: Clinical* 17 (2018), pp. 918–934. DOI: 10.1016/j.nicl.2017.12.022.

[20] K. Ong, D. M. Young, S. Sulaiman, S. M. Shamsuddin, N. R. Mohd Zain, H. Hashim, K. Yuen, S. J. Sanders, W. Yu, and S. Hang. "Detection of subtle white matter lesions in MRI through texture feature extraction and boundary delineation using an embedded clustering strategy". In: *Scientific Reports* 12.1 (2022), p. 4433. DOI: 10.1038/s41598-022-07843-8.

[21] P. Tran, U. Thoprakarn, E. Gourieux, C. L. Dos Santos, E. Cavedo, N. Guizard, F. Cotton, P. Krolak-Salmon, C. Delmaire, D. Heidelberg, et al. "Automatic segmentation of white matter hyperintensities: validation and comparison with state-of-the-art methods on both Multiple Sclerosis and elderly subjects". In: *NeuroImage: Clinical* 33 (2022), p. 102940. DOI: 10.1016/j.nicl.2022.102940.