Robust Identification of White Matter Hyperintensities in Uncontrolled Settings Using Deep Learning

Author name(s) withheld EMAIL(S) WITHHELD
Address withheld

Abstract

White matter hyperintensities (WMH) are associated with an increased risk of stroke, cognitive decline, and dementia. A robust, yet accurate detection of WMH can help with the prevention of more lesions from forming. The task is still challenging as the lesions are often small and irregular. Hence, we propose a robust deep learning-based method for the automatic segmentation of WMH only using fluid-attenuated inversion recovery (FLAIR) scans and MRI-specific data augmentation and compare it with state-of-the-art methods. The methods are tested on public and private data, and we show that our model is more robust to domain shift and achieves higher segmentation accuracy than the alternatives.

Keywords: Deep learning, domain shift, data augmentation, white matter hyperintensity.

1. Introduction

White matter hyperintensities (WMH) of presumed vascular origin are common findings on brain magnetic resonance imaging (MRI), typically assessed in fluid-attenuated inversion recovery (FLAIR) sequences (Pantoni, 2010). These are associated with vascular risk factors and predict an increased risk of stroke, dementia, depression, cognitive impairment, and mobility, both in cross-sectional and longitudinal studies (Wardlaw et al., 2015). Hence, segmentation and detection of WMH are crucial in the analysis of the brain.

Automatic segmentation of WMH attempts to replace the time-consuming, expensive process of manual annotation. Still, the task is challenging because WMH are often small and irregular, making the classes highly imbalanced. Besides, data availability and variability make the problem more complex due to dealing with sensitive data and differences in pathologies, anatomies, MRI scanners, and acquisition protocols.

In this study, we use 3D U-Net models (Isensee et al., 2020) and train them on FLAIR images for WMH segmentation using the common image data augmentation techniques used in (Isensee et al., 2020) and MRI-specific data augmentation proposed by Mehdipour Ghazi and Nielsen (2022). We compare the segmentation accuracy of these two methods with the state-of-the-art method (Li et al., 2018) known as the winner of the MICCAI 2017 WMH Segmentation Challenge (Kuijf et al., 2022), which benefits from both T1-weighted and FLAIR MRI scans for WMH segmentation. The obtained results show that the proposed method identifies WMH in two different datasets significantly better than the alternatives.

2. Methods

2.1. Study Data

The MICCAI 2017 WMH Segmentation Challenge training ($n = 60$) and testing ($n = 110$) datasets were used for training and testing purposes, respectively. The used datasets contain

© 2023 CC-BY 4.0, Anonymous.

FLAIR images from five different scanners, two of which were excluded from the training set and assigned for testing. They were manually annotated as background, WMH, and other (non-WMH) pathologies. we used FLAIR images Additionally, we used an external test set of 22 FLAIR images from an in-house dataset acquired from the US and India.

Since the WMH load has a major impact on the model performance (Gaubert et al., 2023), we divided our test sets into low-load and high-load subsets concerning the volume of WMH and other abnormalities (OA). We refer to images with a load of other abnormalities greater than 1mL as WMH+OA and images without significant other abnormalities as WMH. Either group can be further divided into a "high WMH load" set if the WMH volume is greater than or equal to 10mL, or into a "low WMH load" set otherwise. The threshold values are obtained based on visual inspection.

2.2. Evaluation Metrics

Given the ground truth annotations and predicted segmentations, we evaluate the model performances using the overlap-based metric of Dice similarity coefficient (DSC) (Dice, 1945) and the distance-based metric of volume symmetry (VS) (Taha and Hanbury, 2015).

2.3. Deep Learning Models

We compare three different deep learning-based models for WMH segmentation. We use the state-of-the-art method of (Li et al., 2018), known as the winner of the MICCAI 2017 WMH Segmentation Challenge, which uses an ensemble of 2D U-Nets (Ronneberger et al., 2015). Moreover, the nnUNet (Isensee et al., 2020) is used as a segmentation framework using U-Nets with deep supervision, standard image data augmentation, and auto-configuration of network parameters. Finally, we train 3D U-Nets with MRI-specific data augmentation proposed by Mehdipour Ghazi and Nielsen (2022).

3. Experiments and Results

We trained the 3D models (nnU-Net and the proposed) using a combination of cross-entropy loss and Dice loss and optimized them based on the SGD method with Nesterov momentum of $\mu = 0.99$ and an initial learning rate of $\alpha = 0.01$. The training data were randomly split into training and validation sets in a 5-fold cross-validation fashion, where in each fold 80% assigned for training (48 samples) and 20% (12 samples) for validation.

The obtained segmentation results are shown in Tables 1 and 2. As can be seen, by only using FLAIR images, the proposed model achieves the best WMH segmentation accuracy in both test sets compared to the alternatives. We also observe that the WMH load of other abnormalities has less impact on the accuracy of the MICCAI test set (see Fig. 1).

Table 1: Test segmentation accuracy (mean±SD) of different models. The best results are in boldface.

	MICCAI data		In-house data	
	DSC	VS	DSC	VS
Challenge winner (Li et al., 2018)	.64 ± .17	.76 ± .17	.42 ± **.21**	.71 ± .28
3D nnUNet (Isensee et al., 2020)	.73 ± .13	.91 ± .09	.41 ± .27	.63 ± .29
Our proposed	**.80** ± **.10**	**.92** ± **.09**	**.49** ± .23	**.72** ± **.28**

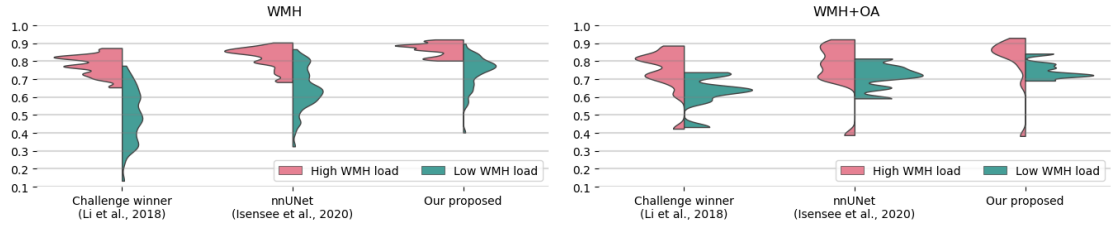

Figure 1: The obtained DSC distribution of different models on WMH segmentation of the MICCAI challenge test set with/without other abnormalities and w.r.t. the WMH load.

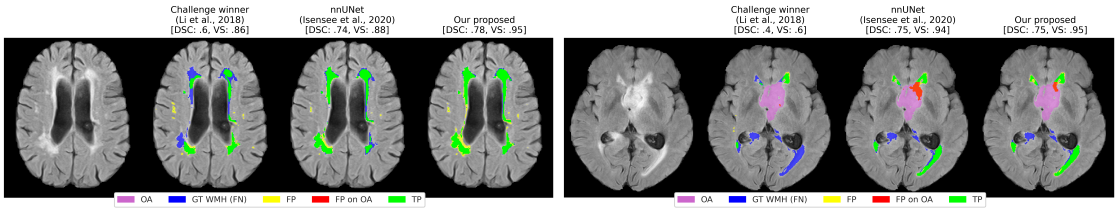

Figure 2: Two samples of the in-house data with high WMH load: a scan with WMH on the left (WMH 15.7mL), and a scan with WMH and other abnormalities on the right (WMH 19.7mL, tumor 15mL).

We achieve a total DSC of 0.80 on the MICCAI test set, which is 16% better than the challenge winner. The total DSC on the in-house test set is 0.49, which is 7% better than the challenge winner. More specifically, as shown in Table 2 and Fig. 2, we obtain DSCs of 0.87 (+9%) and 0.78 (+8%), on subjects with high and low WMH load, respectively. Still, when other abnormalities are present, the accuracy drops to 0.81 (+6%) and 0.72 (+19%).

4. Conclusion

We proposed a robust deep learning-based method for WMH segmentation using FLAIR images. The robustness was tested using two different datasets with scans from different devices and acquisition parameters unseen to the trained models and compared the results with two state-of-the-art methods. We showed that FLAIR sequences are enough to achieve a higher WMH segmentation accuracy than the state-of-the-art, even for out-of-distribution data. However, the models cannot generalize well to the data with a low WMH load.

Table 2: Segmentation accuracy (mean±SD) of different models and loads. The best results are in boldface.

MICCAI test data	WMH+OA				WMH			
	High WMH load		Low WMH load		High WMH load		Low WMH load	
Methods	DSC	VS	DSC	VS	DSC	VS	DSC	VS
Challenge winner (Li et al., 2018)	.75 ± **.12**	.84 ± **.13**	.63 ± .10	.76 ± .14	.78 ± .06	.88 ± .08	.50 ± .15	.65 ± .15
3D nnUNet (Isensee et al., 2020)	.77 ± .13	.89 ± .15	.72 ± .07	**.89** ± **.08**	.82 ± .06	.94 ± .05	.65 ± .12	.89 ± .10
Our proposed	**.81** ± .14	**.89** ± .16	**.75** ± **.05**	.88 ± .09	**.87** ± **.04**	**.95** ± **.04**	**.74** ± **.09**	**.90** ± **.09**

In-house test data	WMH+OA				WMH			
	High WMH load		Low WMH load		High WMH load		Low WMH load	
Methods	DSC	VS	DSC	VS	DSC	VS	DSC	VS
Challenge winner (Li et al., 2018)	.53 ± .18	.68 ± .10	.27 ± .18	.56 ± .34	.70 ± .09	.89 ± .04	.46 ± **.12**	**.86** ± **.10**
3D nnUNet (Isensee et al., 2020)	.69 ± .09	.89 ± **.06**	.17 ± **.15**	.44 ± **.28**	.77 ± .05	.92 ± .06	.50 ± .17	.69 ± .23
Our proposed	**.72** ± **.04**	**.89** ± .09	**.31** ± .18	**.61** ± .33	**.78** ± **.01**	**.96** ± **.02**	**.55** ± .16	.72 ± .24

References

Lee R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3): 297–302, July 1945. doi: 10.2307/1932409. URL https://doi.org/10.2307/1932409.

Malo Gaubert, Andrea Dell'Orco, Catharina Lange, et al. Performance evaluation of automated white matter hyperintensity segmentation algorithms in a multicenter cohort on cognitive impairment and dementia. *Frontiers in Psychiatry*, 13, January 2023. doi: 10.3389/fpsyt.2022.1010273. URL https://doi.org/10.3389/fpsyt.2022.1010273.

Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnU-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, December 2020. doi: 10.1038/s41592-020-01008-z. URL https://doi.org/10.1038/s41592-020-01008-z.

Hugo Kuijf, Matthijs Biesbroek, Jeroen De Bresser, Rutger Heinen, Christopher Chen, Wiesje Van Der Flier, Barkhof, Max Viergever, and Geert Jan Biessels. Data of the white matter hyperintensity (wmh) segmentation challenge, 2022. URL https://dataverse.nl/citation?persistentId=doi:10.34894/AECRSD.

Hongwei Li, Gongfa Jiang, Jianguo Zhang, Ruixuan Wang, Zhaolei Wang, Wei-Shi Zheng, and Bjoern Menze. Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images. *NeuroImage*, 183:650–665, December 2018. doi: 10.1016/j.neuroimage.2018.07.005. URL https://doi.org/10.1016/j.neuroimage.2018.07.005.

Mostafa Mehdipour Ghazi and Mads Nielsen. FAST-AID Brain: Fast and accurate segmentation tool using artificial intelligence developed for brain. *arXiv preprint arXiv:2208.14360*, 2022. doi: 10.48550/arXiv.2208.14360. URL https://doi.org/10.48550/arXiv.2208.14360.

Leonardo Pantoni. Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. *The Lancet Neurology*, 9(7):689–701, July 2010. doi: 10.1016/s1474-4422(10)70104-6. URL https://doi.org/10.1016/s1474-4422(10)70104-6.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. URL https://arxiv.org/abs/1505.04597.

Abdel Aziz Taha and Allan Hanbury. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, 15(1), August 2015. doi: 10.1186/s12880-015-0068-x. URL https://doi.org/10.1186/s12880-015-0068-x.

Joanna M. Wardlaw, Maria C. Valdés Hernández, and Susana Muñoz-Maniega. What are white matter hyperintensities made of? *Journal of the American Heart Association*, 4 (6), June 2015. doi: 10.1161/jaha.114.001140. URL https://doi.org/10.1161/jaha.114.001140.