Article

Semi-Supervised Learning in Medical Image Segmentation: Brain Tissue with White Matter Hyperintensity Segmentation using FLAIR image

ZunHyan Rieu[1,†], Donghyeon Kim[1*,†], JeeYoung Kim[2,†], Regina EY Kim[1], Minho Lee[1], Min Kyoung Lee[3], Se Won Oh[2], Sheng-Min Wang[4], Nak-Young Kim[4], Dong Woo Kang[5], Hyun Kook Lim[4*],

1 Research Institute, NEUROPHET Inc, Seoul, Republic of Korea
2 Department of Radiology, Eunpyeong St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea
3 Department of Radiology, Yeouido St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea
4 Department of Psychiatry, Yeouido St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea
5 Department of Psychiatry, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea
* Correspondence: donghyeon.kim@neurophet.com, drblues@catholic.ac.kr

‡ These authors contributed equally to this work.

Abstract: White matter hyperintensity (WMH) has been considered the primary biomarker from small-vessel cerebrovascular disease to Alzheimer's disease (AD) and has been reported for its correlation of brain structural changes. To perform WMH related analysis with brain structure, both T1-weighted (T1w) and (Fluid Attenuated Inversion Recovery(FLAIR) are required. However, in a clinical situation, it is limited to obtain 3D T1w and FLAIR images simultaneously. Also, the most of brain segmentation technique supports 3D T1w only. Therefore, we introduced the semi-supervised learning method that can perform brain segmentation using FLAIR image only. Our method achieved a dice overlap score of 0.86 for brain tissue segmentation on FLAIR, with the relative volume difference between T1w and FLAIR segmentation under 4.8%, which is just as reliable as the segmentation done by its paired T1w image. We believe our semi-supervised learning method has a great potential to be used to other MRI sequences and provide encouragement to people who seek brain tissue segmentation from a non-T1w image.

Keywords: segmentation; deep-learning; FLAIR; T1w; white matter hyperintensity

1. Introduction

Automated quantification of structural MRI, such as cortical volume or thickness, has been commonly used as an objective indicator of neurodegeneration related to aging, stroke, and dementia.

 doi:10.20944/preprints202104.0004.v1

Recently, combining other MRI-based biomarkers such as white matter hyperintensities (WMH) and susceptibility-weighted images have been tried in AD or aging [1,2] research.

Among them, WMH indicates that bright area appear in the white matter on T2 fluid-attenuated inversion recovery sequences (T2-FLAIR). The etiologies of WMH are diverse and considered primarily markers of small-vessel cerebrovascular disease. The WMH may represent increased blood-brain barrier permeability, plasma leakage, and degeneration of axons and myelin [3].

Larger WMH are associated with an accelerated cognitive decline and increased risk for AD [4]. Recent studies suggest that WMH may play a role in AD's clinical symptoms and a synergistic contribution of both medial temporal lobe atrophy (MTA) and WMH on cognitive impairment and dementia severity in AD [5]. Patients with mild cognitive impairment (MCI) or early AD had concurrent WMH, which shows more significant cognitive dysfunction than those with a low WMH burden [5]. WMH may predict conversion from MCI to AD [6]. Besides cognitive impairment, WMH has reported a relationship with structural changes and cognitive performance, specially in processing speed, even in cognitively unimpaired participants [7].

In clinical practices, WMH burden is usually estimated by visual scale such as Fazekas' rating scale, but it is difficult to use as an objective indicator. Quantification of WMH is essential to evaluate the association of WMH burden with cognitive dysfunction and longitudinal change of WMH volume. So, a reliable automated method for measuring WMH and cortical volume is helpful in clinical practices. Recently, it is reported that WMH progression is associated with more rapid cortical thinning. Therefore, the automated WMH burden and cortical volume measurement method on FLAIR is clinically valuable for trace the longitudinal change in patients with cognitive impairment [8].

Thus, brain structural analysis, specially volumetric analysis, combined WMH could provide more descriptive information to reveal the the relationship between cognitive performance and MRI-based biomarkers.

There are various brain tissue segmentation tools in 3D T1-weighted (T1w) images, such as FreeSurfer [9], SPM [10], and FSL [11]. However, the other MR sequence (Fluid Attenuated Inversion Recovery(FLAIR), Susceptibility weighted imaging(SWI), Gradient echo sequence(GRE), etc.) rarely developed brain tissue segmentation because it does not intend to measure brain volume or analyze brain morphology precisely. In reality, it is rare for clinicians to obtain both 3D T1w and FLAIR images due to the Burden of scanning time. Ultimately, this fact hinders the approach of performing brain tissue segmentation on non-T1w sequences.

In this study, we propose a brain tissue segmentation method to obtain a trainable brain label on FLAIR. With given T1w and FLAIR paired datasets, we initially generate brain labels on FLAIR images. Then, improved the label quality using the semi-supervised learning method. Finally, we train a deep neural network-based brain segmentation model for FLAIR MRI data. The proposed method could be applied in SWI, GRE, and T2, which is hard to obtain brain segmentation labels by itself, to generate a trainable dataset.

2. Materials and Methods

2.1. Subjects

The following study was approved by Institutional Review Board(IRB). As shown in Table 1, Our dataset was supported by Catholic Aging Brain Imaging (CABI) database, which holds brain MRI

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2021 doi:10.20944/preprints202104.0004.v1

scans of outpatients at the Catholic Brain Health Center, Yeouido St Mary's Hospital and Eunpyeong St. Mary's Hospital, the Catholic University of Korea.

Table 1. Summary of the datasets used for train and validation.

Dataset	MRI	No. of Subjects	Matrix Size	Pixel Spacing (mm)	Purpose
CABI	T1w	68	256 x 256 x 256	1.0 x 1.0 x 1.0	Brain tissue segmentation
CABI	FLAIR	68	348 x 384 x 28	0.57 x 0.57 x 6	Brain tissue segmentation
CABI	FLAIR	308	768 x 768 x 32	0.27 x 0.27 x 5	WMH segmentation

For Brain tissue segmentation, a total of 68 subjects with paired T1w image and FLAIR image were obtained from Yeouido St Mary's Hospital and Eunpyeong St. Mary's Hospital, the Catholic University of Korea, from 2017 to 2019. The images had a consistent matrix size of 256 x 256 x 256, along with its pixel spacing of 1.0mm x 1.0mm x 1.0mm.

For WMH segmentation, initially, a total of 396 FLAIR images with its clinically confirmed WMH label were obtained from Eunpyeong St. Mary's Hospital, the Catholic University of Korea, from 2020 to 2021. The images had a consistent matrix size of 348 x 384 x 28, along with its pixel spacing of 0.57mm x 0.57mm x 5mm. However, due to the various spectrum of WMH in the dataset, we excluded the FLAIR images with neglectable WMH regions, along with mislabeled cases, which ended up a total of 308 FLAIR images at the end.

2.2. Overview of Proposed Method

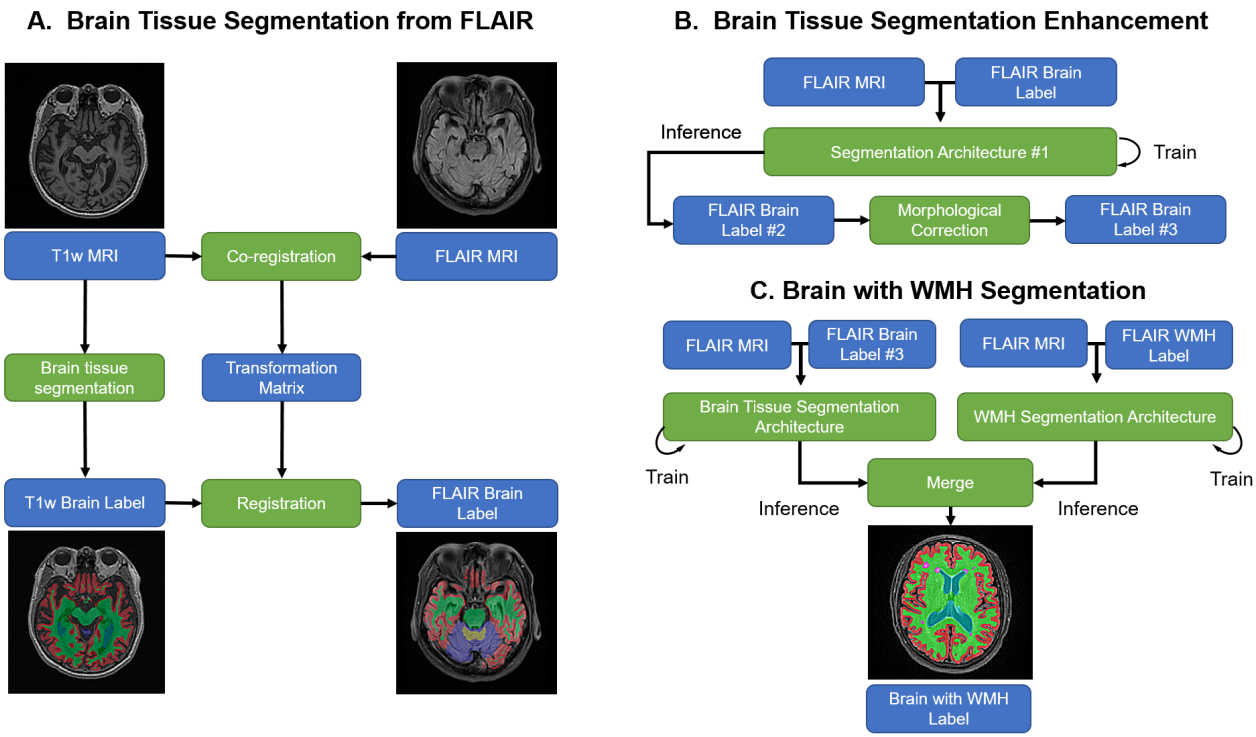

Figure 1. The pipeline of the proposed method. Blue box: input or output data; Green box: computation process.

Our goal is to produce a brain tissue and WMH segmentation on the FLAIR image exclusively. Since the FLAIR image itself lacks the structural information and give the hardship to create the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2021 doi:10.20944/preprints202104.0004.v1

ground-truth label, we will proceed with the following process shown in Figure 1: A. Brain Tissue Segmentation from FLAIR, B. Brain Tissue Segmentation Enhancement, and C. Brain with WMH Segmentation.

2.3. Brain Tissue Segmentation from FLAIR

2.3.1. Segmentation

In this study, each tissue (brain tissue and WMH) was labeled T1w and FLAIR image separately. For the T1w image, we used FreeSurfer (6.0, Boston, USA) reconstruction(recon-all), then extracted brain labels consisting of cerebral gray matter (CblGM), cerebral white matter (CblWM), cerebellum gray matter (CbrGM), cerebellum white matter(CbrWM), and lateral ventricle (Vent) from aseg+aparc.mgz [9].

Reference segmentation of WMH were performed by manual outlining on the FLAIR images. A total of 308 datasets were manually segmented on FLAIR images, producing binary masks with the value of 0 (non-WMH class) or 1 (WMH class). The manual segmentation process was performed through a consensus among four certified radiologists (JY Kim, SW Oh, MK Lee), who did not have access to T1w images for subjects. Chronic infarcts were hypointense with a hyperintense rim (probably cystic) lesions on the FLAIR image and were excluded in WMH.

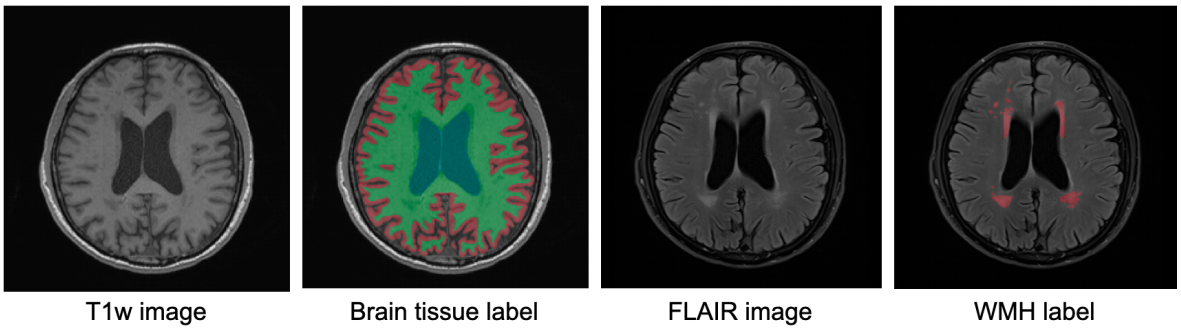

Figure 2. The example of image and segmentation label of T1w and FLAIR image.

2.3.2. Co-registration

Co-registration is a method to align two individual MRIs obtained from the same subject. This process is required when especially alignment needs to be made with MRIs with a different modality from the same subject. In our case, this is used to align the T1w image to the FLAIR image. Since the primary purpose of the Brain tissue segmentation from FLAIR is to generate initial brain tissue labels on FLAIR images, we calculated the transform matrix from T1-weighted (T1w) MRI to FLAIR MRI using the spatial co-registration method with SimpleITK [12]. Then we segmented T1w brain tissue using T1w brain segmentation tools. We performed the brain tissue label registration on the T1w image and the transform matrix based on FLAIR images' brain tissue label. The co-registrated brain tissue label is on FLAIR images. However, due to differences in image spacing and dimension, it did not delineate brain tissue structure accurately. Therefore, we iteratively enhanced the brain tissue segmentation labels of FLAIR images.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2021 doi:10.20944/preprints202104.0004.v1

2.4. Brain Tissue Segmentation Enhancement

For brain tissue segmentation enhancement, we trained a convolutional neural network(CNN) and segmented brain tissue from FLAIR. As for the initial segmentation model, we used U-Net [13] with EvoNorm layer [14]. Histogram Equalization, Rescale-Intensity, and Z Normalization were used for the pre-processing, and we set the input shape as 196 x 196.

With the following process, the CNN was able to learn the geometrical representation of brain morphology. Thus, the delineation was more precise afterward than the brain tissue label obtained T1w brain segmentation label. However, a small region distant from the brain could miss-labeled, or the peri-ventricular area could be assigned as none tissue due to the initial brain tissue label's impreciseness. Therefore, we performed a morphological correction (removing isolated label) to generate the final brain tissue label.

2.4.1. Morphological label correction

After training brain tissue segmentation in FLAIR image, segmentation rather delineates brain tissue clearly than the brain tissue label obtained from T1w images. However, there some noise that incomplete training label data could influence. Thus, we perform a simple morphological correction method based on brain structure characteristics to enhance brain tissue label. All brain tissue (including ventricles) must be connected 26 ways in 3 dimensions. Thus we performed connected component-based noise reduction. Furthermore, the ventricles label could be miss-labeled as none tissue due to the similarity between ventricles and none-tissue. Thus we use the fill-hole method to compensate it [15]. The morphological refined brain tissue does not have any tissue label isolated and a hole in the ventricles.

2.5. Brain with WMH Segmentation

The following study aims to design a deep learning-based segmentation learning framework instead of proposing a novel segmentation network. Therefore, we used three well-known segmentation architectures for Brain tissue segmentation: U-Net [13], U-Net++ [13], and HighRes3DNet [16]. For WMH segmentation, we only used U-Net architecture with fine-tuning [17]. We used the same number of kernels size in each architecture in the original article. We set the input and output shape to 196 x 196 and used the EvoNorm layer [14] instead of the batch normalization and activation function. For U-Net, we did not use the deep-supervision method, and for HighRes3DNet, we did not use the Monte Carlo sampling strategy using the dropout layer for pure architecture comparison.

2.5.1. Pre-Processing

Even though Brain tissue segmentation and WMH are trained individually, we still had to resample the whole dataset into a 1mm isometric space, since our purpose is to merge the segmentation result of the brain tissue and WMH. For WMH segmentation, we used the skull-stripping method with HD-BET [18] on the FLAIR images of the WMH dataset to focus our train on the white-matter region. Also, to deal with the differences in MRI's intensity variance, we used the min-max normalization technique with percentile cut-off of (0.05, 99.95), histogram normalization, and z-score normalization by using TorchIO [19]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2021 doi:10.20944/preprints202104.0004.v1

2.5.2. Data Augmentation

Data augmentation is an inevitable process to increase the robustness of segmentation. As the word itself describes, the input data gets augmented from the original dataset and added during the training. All of our data augmentations were applied by using TorchIO [19].

- RandomAffine: a scale parameter in range of 0.85 1.15
- RandomMotion: a degrees value to 10 with a translation value to 10 in millimeter
- RandomBiasField: a magnitude coefficient parameter in range of -0.5 to 0.5
- RandomNoise: a mean value of gaussian distribution in range of 0 to 0.025
- RandomFlip: a spatial transform value to 2, which inverse the Z axis

2.5.3. Experiment Setup

All deep neural network learning were achieved using torch deep learning library [20] on a workstation with Intel i9-9900X 3.5Ghz CPU, 128G RAM, and two NVIDIA RTX 2080 11GB GPU. We used torchIO [19] medical deep learning library for pre-processing, augmentation, patch-based learning based on torch.

For Brain Tissue Segmentation, we distributed 68 subjects with a split ratio of 0.8; 54 subjects for training and 14 subjects for validation. In the data pre-processing, we used the following normalization technique; min-max normalization with percentile cut-off (0.05, 99.95), histogram normalization, and z-score normalization. We used the medical image-based augmentation technique to improve the robustness of CNN segmentation architecture as mentioned above. After that, we used patch-based training, which crops 128 samples per FLAIR image at a random location4.. The size of cropped image is 128x128. Lastly, we train CNN segmentation network using cross entropy loss function [21] and adamW optimizer [22] with learning rate=0.001, betas=(0.9, 0.999), eps=1e-08, and weight decay=0.01. As for the training process, U-Net and U-Net++ was terminated at 500 epochs. But considering the different architecture that HighRes3DNet had, we gave twice more epochs to get its loss to be converged. After finish training architecture, we used grid-based sampling and aggregation to inference segmentation results.

For WMH Segmentation, we distributed 308 subjects with a split ratio of 0.9; 277 subjects for training and 31 subjects for validation. We resampled the image spacing of X-axis and Y-axis of the FLAIR image and its paired ground-truth to 1.0 and 1.0. (Z space was excluded from this process since our training process requires the 2D Plane data) We skull-stripped the resampled MRI image using HD-BET [18], an open-source brain extraction tool to focus on the WM region. As for the data augmentation, we excluded the histogram normalization from the augmentation process to deal with FLAIR MRI images taken with different sequences, such as fat suppression. Instead, we used the normalization technique, min-max normalization with (0, 100) percentile and z-score normalization. No medical image-based augmentation technique was used on WMH image processing to avoid any confusion on the already sensitive object and the other experiment setup is the same as Brain tissue segmentation. For the loss function, we used DiceBCE loss function, which is a combination of Dice loss function [23] and Binary Cross Entropy loss function [21] to deal with the varying size that WMH has. Also, we used adamW optimizer [22] with learning rate=0.001, betas=(0.9, 0.999), eps=1e-08, and weight decay=0.01.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2021 doi:10.20944/preprints202104.0004.v1

2.5.4. Metrics for Evaluation

Our goal is to compare the segmentation done with the FLAIR image to the segmentation done with the T1w image and WMH. Therefore, our primary evaluation method is the dice overlap score [24], which measures the similarity between the ground-truth and the prediction. In our case, the ground-truth (A) is the original label and volume from the reference FLAIR, and the prediction (B) is the label and volume prediction made with our trained model with the FLAIR image.

$$\bullet \text{ Dice Overlap Score (A, B)} = \frac{2\,|A \cap B|}{|A| + |B|}$$

Since the dice overlap is the comparison between the reference FLAIR (ground-truth) and the predicted segmentation, we will need to measure the difference between the reference T1w and the predicted segmentation to see if the prediction would match the most fundamental ground-truth.

$$\bullet \text{ Relative Difference (X, } X_{reference}) = \frac{X - X_{reference}}{X_{reference}} * 100$$

3. Results

In this section, we will compare each model's detailed segmentation performance by comparing the average dice overlap score of individual brain tissues and comparing the relative difference of segmented volume with the reference T1 and the segmentation result of each model done with its paired FLAIR. Also, we will report the segmentation performance of WMH segmentation, which was done with an individual dataset.

3.1. The Dice Overlap Score Comparison in Brain Segmentation

	Cerebrum GM	Cerebrum WM	Cerebellum GM	Cerebellum WM	Ventricle	Average
U-Net++	0.76±0.02	0.87±0.01	0.91±0.01	0.83±0.03	0.89±0.03	0.85±0.06
U-Net	0.76±0.02	0.87±0.01	0.91±0.01	0.86±0.03	0.90±0.03	0.86±0.06
HighRes3DNet	0.69±0.02	0.84±0.01	0.87±0.02	0.79±0.04	0.87±0.05	0.81±0.08

Table 2. The comparison of the average dice overlap score of each model on individual brain tissues.

To evaluate the average dice overlap score, we demonstrate the following Table 2, which has the average dice overlap score of each model on each brain tissue. All the models exceeded the average dice overlap score of 0.80 as shown in Table 2. Also, HighRes3DNet, which had an entire 3D convolutional network, a different segmentation architecture than U-Net and U-Net++ could get a similar segmentation result of the U-Net series.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2021 doi:10.20944/preprints202104.0004.v1

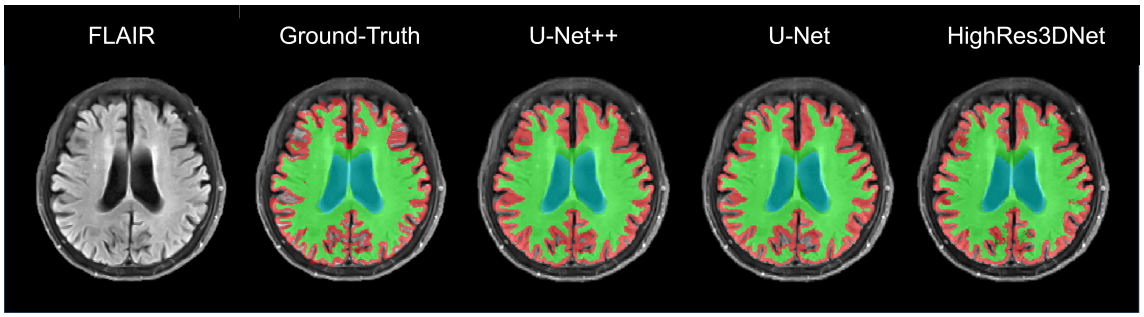

Figure 3. The comparison of the FLAIR reference and the predicted segmentation from each model (U-Net++, U-Net, and HighRes3DNet)

Figure 3 shows the segmentation results of each model. By comparing each model's predicted segmentation results to its ground-truth, we could observe that the results were similar to each other as the average dice overlap score, as it was demonstrated in Table 2. The U-Net based architectures had a similar prediction on their segmentation, which is hard to notice the difference visually. On the other hand, the predicted segmentation of HighRes3DNet had a noticeable difference in its Cerebrum GM. Even with visual comparison, HighRes3DNet had a lesser region in Cerebrum GM and labeled more in Cerebrum WM region instead.

3.2. Measured Volume Comparisons between T1w and FLAIR Images

Measurement	Brain Tissue	Reference (T1)	Reference (FLAIR)	U-Net++	U-Net	HighRes3DNet
Volume (ml, mean±SD)	Cerebellum GM	430.8±45.7	444.5±47.3	458.0±45.4	455.4±42.0	408.0±34.6
	Cerebellum WM	499.3±55.8	516.3±57.5	510.4±52.7	519.5±54.4	559.3±56.4
	Cerebrum GM	100.2±10.2	103.5±10.7	102.2±9.6	101.1±9.6	96.7±8.9
	Cerebrum WM	23.2±3	24.0±3.1	22.5±3.1	23.7±2.8	22.9±3.6
	Lateral Ventricles	41.2±20.8	42.5±21.6	39.1±20.6	39.9±20.7	41.5±21.3
Relative Difference (%, mean±SD)	Cerebellum GM	-	3.2±1.1	6.4±2.5	5.9±2.8	5.4±4.0
	Cerebellum WM	-	3.4±0.9	2.5±2.0	4.1±2.0	12.2±3.6
	Cerebrum GM	-	3.3±1.4	2.7±2,2	2.3±1.7	4.3±3.1
	Cerebrum WM	-	4.3±3.2	6.1±3.7	6.8±6.7	9.7±7.9
	Lateral Ventricles	-	3.1±1.4	6.1±4.3	4.7±4.3	5.4±6.0
Average Difference (%, mean±SD)	-	-	3.4±0.5	4.8±2.0	4.8±1.7	7.4±3.4

Table 3. The volume of individual brain tissue of the reference T1, reference FLAIR, U-Net++, U-Net, and HighRes3DNet with the relative difference between the reference T1 and the designated column. GM, gray matter; SD, standard deviation; WM, white matter.

Table 3 shows the volume of segmentation (ml) and the relative difference between the segmented volume and the reference T1 (%). The volume of individual brain tissue shared the similarity in the U-Net and U-Net++ with the same average relative difference of 4.8%. On the other hand, HighRes3DNet shared a similar relative difference except for the Cerebellum WM (this is due to the overextended segmentation made on Cerebellum WM of HighRes3DNet, demonstrated in Figure 3)

3.3. The Dice Overlap Score Comparison in WMH Segmentation

	Dice Overlap Score	Precision	Recall	F1 Score
U-Net	0.81±0.07	0.86±0.06	0.84±0.08	0.84±0.04

Table 4. The average dice overlap score of WMH, along with precision, recall, and F1 score value.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2021 doi:10.20944/preprints202104.0004.v1

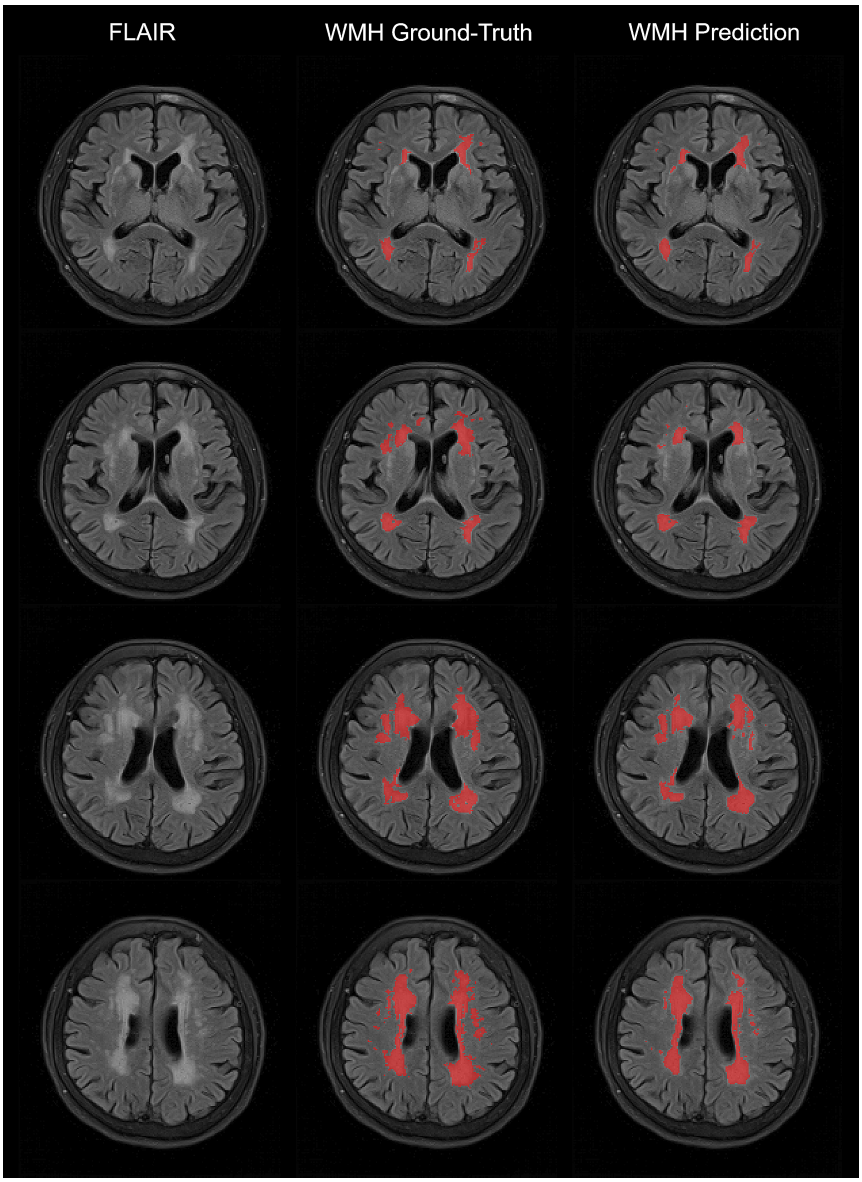

Figure 4. The comparison of WMH ground-truth and WMH prediction

The computation of the proposed WMH segmentation took about 18 hours total, and the DiceBCE losses of U-Net converged in 500 epochs. The dice overlap score of WMH was about 0.81, and the value of precision and recall were higher than the dice overlap score itself as shown in Table 4.

4. Discussion

In this study, we developed reliable automated segmentation method using FLAIR for WMH and cortical volume without 3D T1-weighted volume images. In clinical practices, it is not easy to get 3D T1-weighted volume images due to long scan time, MR machine performance, and patient's condition. On the other hand, FLAIR images is more common and essential sequence to evaluate the gaining brain and easy to get in routine practices, so this method is applicable to more patients.

The result of Brain Tissue Segmentation Enhancement suggests that semi-supervised learning enabled the direct training of brain tissue segmentation on FLAIR image only. By following the

procedure shown in Figure 1, the transformation between two different MRI modalities T1w and FLAIR could be made with improvement on labeling quality.

For Brain segmentation, there was no comparable reference of dice overlap score with segmentation done with FLAIR only. However, the comparison of volumes from the reference T1w and the segmentation result could be made, which is less than 10% for all three models; 0.86 for U-Net and 0.85 for U-Net++, and 0.81 for HighRes3DNet. Considering the relative difference of 3.4% already existed from the reference FLAIR, we could consider that our method is sufficient except the HighRes3DNet. At the same time, we believe the architectural difference of HighRes3DNet caused this issue. With more given time on training, HighRes3DNet would have similar results to what U-Net and U-Net++ achieved.

As we proposed, our method could measure the volume of brain tissue and WMH using only the FLAIR image. However, we did not evaluate the association of WMH burden with cortical volume or cognitive impairment. In a further study, the evaluation of the correlation WMH and cortical volume with cognitive decline dependency on age using FLAIR for clinical feasibility needs to be done.

T2, SWI, and GRE are also known for the difficulty of getting their brain tissue labels because it is pretty much impossible to get their structural information without its paired T1w image. As mentioned before, there are tremendous medical segmentation researches done with deep learning-based, yet not that many considered getting structural information from a single modality except T1w MRI. However, with the proposed semi-supervised learning method, T2, SWI, and GRE should be applicable as well.

5. Conclusions

We introduced the semi-supervised learning method for brain tissue segmentation using only FLAIR image. With our brain segmentation results, we demonstrated that our FLAIR segmentation is just as reliable as the segmentation done with its paired T1w image and could perform the brain tissue segmentation and WMH segmentation from a single FLAIR image. Furthermore, the results have shown that our semi-supervised learning method may not be limited to FLAIR and could also be applied to T2, SWI, and GRE with the difficulty of getting brain tissue labels without its paired T1w image. We believe our semi-supervised learning method could impact those who seek brain tissue segmentation from non-T1w image sequences.

Author Contributions: Conceptualization, Z.R., D.K., J.K., H.K.L.; methodology, Z.R. and D.K; validation, Z.R.; formal analysis, Z.R; investigation, Z.R, D.K., J.K. and H.K.L.; resources, J.K., H.K.L; data curation, M.K.L., S.W.O., S.W., N.K., and D.W.K; writing—original draft preparation, Z.R., D.K., and J.K.; writing—review and editing, M.L., R.E.K., M.K.L., S.W.O., S.W., N.K., and H.K.L; visualization, Z.R.; supervision, D.K. and H.K.L.; project administration, D.K.; funding acquisition, D.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Research and Business Development Program through the Korea Institute for Advancement of Technology(KIAT) funded by the Ministry of Trade, Industry and Energy(MOTIE) (grant number: P0009507) and supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare and Ministry of science and ICT, Republic of Korea (grant number: HU20C0315).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Caligiuri, M.E.; Perrotta, P.; Augimeri, A.; Rocca, F.; Quattrone, A.; Cherubini, A. Automatic Detection of White Matter Hyperintensities in Healthy Aging and Pathology Using Magnetic Resonance Imaging: A Review. *Neuroinformatics* **2015**, *13*, 261–76. doi:10.1007/s12021-015-9260-y.
2. Guerrero, R.; Qin, C.; Oktay, O.; Bowles, C.; Chen, L.; Joules, R.; Wolz, R.; Valdés-Hernández, M.C.; Dickie, D.A.; Wardlaw, J.; Rueckert, D. White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks. *NeuroImage: Clinical* **2018**, *17*, 918–934, [1706.00935]. doi:10.1016/j.nicl.2017.12.022.
3. Hase, Y.; Horsburgh, K.; Ihara, M.; Kalaria, R.N. White matter degeneration in vascular and other ageing-related dementias. *Journal of neurochemistry* **2018**, *144*, 617–633.
4. Au, R.; Massaro, J.M.; Wolf, P.A.; Young, M.E.; Beiser, A.; Seshadri, S.; D'Agostino, R.B.; DeCarli, C. Association of white matter hyperintensity volume with decreased cognitive functioning: the Framingham Heart Study. *Archives of neurology* **2006**, *63*, 246–250.
5. Garnier-Crussard, A.; Bougacha, S.; Wirth, M.; André, C.; Delarue, M.; Landeau, B.; Mézenge, F.; Kuhn, E.; Gonneaud, J.; Chocat, A.; others. White matter hyperintensities across the adult lifespan: relation to age, Aβ load, and cognition. *Alzheimer's research & therapy* **2020**, *12*, 1–11.
6. Prasad, K.; Wiryasaputra, L.; Ng, A.; Kandiah, N. White matter disease independently predicts progression from mild cognitive impairment to Alzheimer's disease in a clinic cohort. *Dementia and geriatric cognitive disorders* **2011**, *31*, 431–434.
7. Kearney-Schwartz, A.; Rossignol, P.; Bracard, S.; Felblinger, J.; Fay, R.; Boivin, J.M.; Lecompte, T.; Lacolley, P.; Benetos, A.; Zannad, F. Vascular structure and function is correlated to cognitive performance and white matter hyperintensities in older hypertensive patients with subjective memory complaints. *Stroke* **2009**, *40*, 1229–1236.
8. Kim, S.J.; Lee, D.K.; Jang, Y.K.; Jang, H.; Kim, S.E.; Cho, S.H.; Kim, J.P.; Jung, Y.H.; Kim, E.J.; Na, D.L.; others. The Effects of Longitudinal White Matter Hyperintensity Change on Cognitive Decline and Cortical Thinning over Three Years. *Journal of clinical medicine* **2020**, *9*, 2663.
9. Fischl, B. FreeSurfer. *NeuroImage* **2012**, *62*, 774–781. 20 YEARS OF fMRI, doi:https://doi.org/10.1016/j.neuroimage.2012.01.021.
10. Ashburner, J.; Friston, K.J. Unified segmentation. *Neuroimage* **2005**, *26*, 839–851.
11. Jenkinson, M.; Beckmann, C.F.; Behrens, T.E.; Woolrich, M.W.; Smith, S.M. Fsl. *Neuroimage* **2012**, *62*, 782–790.
12. Beare, R.; Lowekamp, B.; Yaniv, Z. Image Segmentation, Registration and Characterization in R with SimpleITK. *Journal of Statistical Software* **2018**, *86*. doi:10.18637/jss.v086.i08.
13. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support; Stoyanov, D.; Taylor, Z.; Carneiro, G.; Syeda-Mahmood, T.; Martel, A.; Maier-Hein, L.; Tavares, J.M.R.S.; Bradley, A.; Papa, J.P.; Belagiannis, V.; Nascimento, J.C.; Lu, Z.; Conjeti, S.; Moradi, M.; Greenspan, H.; Madabhushi, A., Eds.; Springer International Publishing: Cham, 2018; pp. 3–11.
14. Liu, H.; Brock, A.; Simonyan, K.; Le, Q.V. Evolving normalization-activation layers. *arXiv preprint arXiv:2004.02967* **2020**.
15. Lehmann, G. Label object representation and manipulation with ITK. *Insight Journal* **2008**, pp. 1–34.
16. Li, W.; Wang, G.; Fidon, L.; Ourselin, S.; Cardoso, M.J.; Vercauteren, T. On the compactness, efficiency, and representation of 3D convolutional networks: Brain parcellation as a pretext task. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2017**, *10265 LNCS*, 348–360, [1707.01992]. doi:10.1007/978-3-319-59050-9_28.
17. Duong, M.T.; Rudie, J.D.; Wang, J.; Xie, L.; Mohan, S.; Gee, J.C.; Rauschecker, A.M. Convolutional neural network for automated FLAIR lesion segmentation on clinical brain MR imaging. *American Journal of Neuroradiology* **2019**, *40*, 1282–1290.
18. Isensee, F.; Schell, M.; Pflueger, I.; Brugnara, G.; Bonekamp, D.; Neuberger, U.; Wick, A.; Schlemmer, H.P.; Heiland, S.; Wick, W.; Bendszus, M.; Maier-Hein, K.H.; Kickingereder, P. Automated brain extraction

of multisequence MRI using artificial neural networks. *Human Brain Mapping* **2019**, *40*, 4952–4964, [1901.11341]. doi:10.1002/hbm.24750.

19. Pérez-García, F.; Sparks, R.; Ourselin, S. TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *arXiv preprint arXiv:2003.04696* **2020**.
20. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; others. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703* **2019**.
21. Rubinstein, R.Y.; Kroese, D.P. *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning*; Springer Science & Business Media, 2013.
22. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* **2014**.
23. Sudre, C.H.; Li, W.; Vercauteren, T.; Ourselin, S.; Cardoso, M.J. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*; Springer, 2017; pp. 240–248.
24. Sorensen, T.A. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biol. Skar.* **1948**, *5*, 1–34.