3D U-Net Application for White Matter Hyperintensity Analysis with Brain Lobe Identification

Kauê T. N. Duarte KAUE.DUARTE@UCALGARY.CA
David G. Gobbi DGOBBI@UCALGARY.CA
Richard Frayne RFRAYNE@UCALGARY.CA
Calgary Image Processing and Analysis Centre and Seaman Family, Dept of Radiology and Clinical Neurosciences, Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada T2N 2T9

Editors: Under Review for MIDL 2022

Abstract

White-matter hyperintensity (WMH) is associated with many disorders including small vessel diseases where it is suggestive of underlying cerebrovascular abnormalities. The fluid-attenuated inversion recovery (FLAIR) magnetic resonance (MR) imaging sequence is commonly used to visualize WMH, since it provides good image contrast not only between WMH and normal tissue, but also between WMH and cerebrospinal fluid. Manual segmentation of these presumed lesions in brain volumes, on an image slice-by-slice basis, is time-consuming, however, this process remains the broadly accepted gold standard. 3D U-shaped convolutional neural networks (3D U-Net CNNs) are a type of semantic segmentation that processes images volumetrically rather than by slice. We used a 3D U-Net model to segment the WMH in brain volumes obtained from twenty presumed healthy participants. VGG16, VGG19, ResNet and EfficientNet architectures were used as feature extractors. Results were grouped by brain lobe (frontal, insula, occipital, parietal, temporal) to identify regions that were more affected by the WMH. In general, the predicted volumes had an intersection over union (IoU) measure $> 95\%$ compared to manual segmentation. This metric demonstrated that 3D U-Net CNN models are reliable for WMH identification. Assessments of these architectures have shown that the correspondence between WMH and the brain lobes was meaningful and might provide future insights about presumed abnormalities.

Keywords: U-Net, semantic segmentation, white matter hyperintensity (WMH), fluid-attenuated inversion recovery (FLAIR) imaging.

1. Introduction

White matter hyperintensities (WMHs) are image features (presumed lesions) commonly found in older participants (Wardlaw et al., 2015), which are characterized by brighter-than-normal regions within the white matter on T2-weighted (T2-w) or fluid-attenuated inversion recovery (FLAIR) magnetic resonance (MR) images. The WMH are often of vascular origin and are associated with potential neuropathological problems including hypertension and dementia (Alber et al., 2019).Manual identification of voxel hyperintensities is time-consuming and can be error-prone. Neural network-based approaches (NN) have recently been adopted to detect WMH in either a supervised or unsupervised manner. Among the types of NNs, a variation of convolutional neural networks (CNN) called the U-Net has shown promising outcomes. However, most work has focused on 2D implementations of lesion segmentation, which ignores some of the spatial information present in the images. In

© 2022 K. Duarte, D.G. Gobbi & R. Frayne.

addition, it is often important to determine location of the WMH as this can impact clinical interpretation. Here we develop a method that segments and then assigns the presumed lesions to distinct brain lobes.

We propose a 3D U-Net approach for segmentation and brain lobe analysis of WMH on MR FLAIR images. Predicting and accurately locating the WMH that will enable physicians to perform a more reliable assessment. Our main contributions are: (1) automatic prediction, labeling of the volume voxels, and assigning voxels to the brain lobes; (2) identifying the best architecture (VGG16, VGG19, ResNet and EfficientNet) for WMH segmentation; and (3) determining the accuracy of the prediction by brain lobe (frontal, insula, occipital, parietal, temporal). The remainder of this work is organized as: Section 2 explains the state-of-the-art and the main advances in WMH segmentation. Section 3 defines the data set and WMH segmentation approach. Section 4 provides an analysis of the segmentation experiments (per-lobe analysis), and Section 5 discusses the results. Finally, Section 6 summarizes our work and conclusions, and describes possible future work.

2. Review of the Related Work

CNNs are a major topic in current literature. They are used for both classification and segmentation tasks. Their generalizability is one of their strongest characteristics, because it allows the use of previously unseen data in your network, such as from different scanner manufactures or with varying image dimensions. 2D U-Nets are the most prevalent type of CNN in the literature, primarily because they have lower computational costs compared to 3D implementations. However, 2D U-Nets do not use the full spatial information inherent to 3D images. To bridge the gap between 2D and 3D, (Sundaresan et al., 2021) proposed a tri-planar U-Net ensemble network (TrUE-Net) to identify WMH abnormalities in deep or periventricular regions by using three orthogonal image planes (axial, sagittal, coronal) using FLAIR and T1-weighted (T1-w) brain MR imaging. The image plane predictions were assembled to form a 3D brain volume based on majority voting. Their comparisons using distinct datasets as input demonstrated promising generalization characteristics of U-Net shaped architectures.

Li et al. (2021) addressed the importance of locating and measuring WMH. They proposed an ensemble of predictive models such as 2D U-Net, SE-Net, and multifeatures to aid the segmentation of FLAIR and T1-w MR images. They compared not only the importance of identifying the lesions, but also the importance of the scan resolution. Their pipeline was standardized, consisting of skull stripping, data augmentation, and normalization via co-registration with T1-w MR imaging. Later, the ensemble was trained with different datasets and achieved promising results.

Guerrero et al. (2018) recently distinguished stroke and WMH-related pathologies by means of 2D U-Net segmentation using axial FLAIR images. The residual information was addressed by adopting the uResNet architecture, although they found no significant difference in performance between using summation and concatenation in skip connections, thus skipping some layers of the architecture in order to avoid overfitting the models.

The trade-off between computational cost and quality of the segmentation needs to be addressed when choosing 2D or 3D U-Net models. Duong et al. (2019) adopted a 3D U-Net to identify WMH in nineteen distinct brain pathologies using 3D FLAIR images.

Their preprocessing step removed the skull information using T1-w MR imaging, and their registration step allowed removal of the skull for the FLAIR images. Later, a data augmentation step was applied to increase the amount of training data, resampled to 1 mm^3 and divided into volumetric patches, thus reducing peak computational effort. The 3D U-Net is adopted with the patches to segment the presumed lesions, which are later assembled back to the entire volume. The generalization is also addressed in this study, which confirms the benefits of using U-Nets. Rudie et al. (2019) addressed the importance of using real-world scenarios in their multi-disease U-Net segmentation to compare brain gliomas and WMH. The use of T1-w, T2-w, and FLAIR MR imaging also improved their method and allowed the comparison of distinct scenarios. The segmentation using a 3D U-Net identified a reliable segmentation of the tissues, although they highlighted the importance of a proper quantification to efficiently assess a clinical decision-making. Our work differs from the literature in the following aspects: (1) the comparison of distinct well-known U-Net shaped architectures, and (2) assignment of the predicted WMH regions to brain lobes.

3. Materials and Methods

3.1. Participants Data and MR Protocol

The Calgary Normative Study (CNS) is a private longitudinal study that aims to evaluate MR pulse sequences and analysis techniques as potential biomarkers toward neurodegenerative classification. (McCreary et al., 2020) This dataset includes quantitative MR mapping of brain (*e.g.*, diffusion and perfusion measurements), as well as T1-w, T2-w, and FLAIR images. We selected twenty participants for our preliminary study, including both T1-w and FLAIR imaging. Male participants corresponds to 11 (55%) of the participants, with an average age 44.0 ± 17.0 years (mean $\pm$ std). Female participants had an average age of 43.2 ± 16.6 years. We intend to increase the number of participants in future studies. T1-w scans were acquired with a 3-T MR scanner (Discovery 750, General Electric Healthcare, Waukesha, WI) using a 3D-FSPGR pulse-sequence in one of two configurations: Fourteen (14, 70%) participants had TE=2.6 ms, TI=650 ms, TR=6.4 ms, flip=8°, and isotropic voxel spacing of 1.0 mm; and six (6, 30%) participants had TE=3.1 ms, TI=400 ms, TR=7.4 ms, flip=10°, and isotropic voxel spacing of 1.0 mm. A volumetric stack of 2D FLAIR images was also acquired on the same scanner using the following parameters: TE=140 ms, TI=2250 ms, TR=9000 ms, in-plane pixel spacing=0.94 mm, and slice thickness=3 mm.

3.2. 3D UNET-based WMH Segmentation approach

The FLAIR data were semi-automatically segmented to generate WMH masks (ground-truth) using custom software (Cerebra-LesionExtractor (Gobbi et al., 2012)). This software utilizes seed-based region growing and manual editing.

Our approach comprises four major steps, as illustrated in Figure 1: (1) *Data Preparation*: Obtaining FLAIR images and the preprocessing steps, (2) *3D CNN U-Net*: Performing volumetric segmentation, (3) *Anatomical Labeling*: Automatic parcellation of the brain T1-w images into white matter regions, and (4) *Volume Registration*: Registration of the parcellated T1-w images to the FLAIR space.

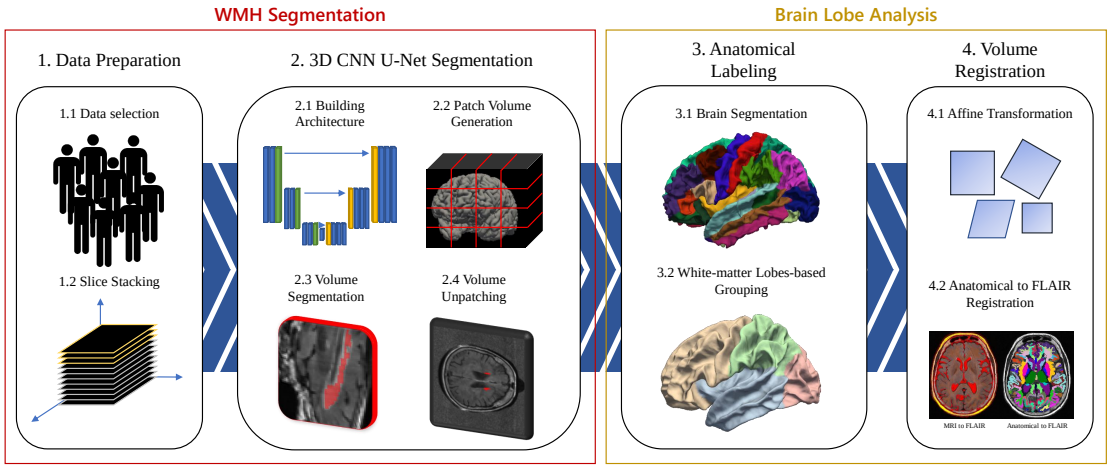


Figure 1: Flowchart of the proposed method.

3.2.1. Data Preparation

We examined the FLAIR volumes and the corresponding hand-generated WMH mask volumes. Each voxel in the mask volume was labeled either *True* or *False* according to whether it was marked as WMH. The FLAIR data were acquired as 256×256 axial images, with the number of slices varying from participant to participant. Because 3D semantic segmentation requires a power-of-two in each image dimension, we padded each FLAIR data volume with additional "blank" slices to achieve a final volume of 256×256×64.

3.2.2. 3D CNN U-Net Segmentation

The ability of a CNN to learn the patterns required to achieve a specific objective depends strongly on the network architecture. We tested ResNet with 152 layers (ResNet152), EfficientNet version B0 (EfficientNetB0), and two VGG-based architectures (VGG16 and VGG19) because they are widely used for feature extraction in medical image processing. The FLAIR volumes were split into $64 \times 64 \times 64$ patches, providing 16 patches per volume. To reduce class-imbalance, we only chose patches that contained at least one labeled WHM voxel, as proposed by Guerrero et al. (2018). We adopted a U-Net segmentation using a learning rate $l_0 = 5 \times 10^{-4}$. The loss function employed was a combination of Dice coefficient and binary focal loss. Dice loss (DL) reduces data imbalance issues between foreground and background, and it has been widely used in medical field. DL is described in Equation 1.

$$DL(TP, FP, FN) = \frac{(1+\beta^2) \cdot TP}{(1+\beta^2) \cdot FP + \beta^2 \cdot FN + FP} \tag{1}$$

where β corresponds to coefficient of balance, TP and FP represents the True and False Positive, and FN corresponds to the False Negative.

Focal loss (FL) is also addressed to remove class imbalance, such as tumor segmentation, by applying a modulating to the cross-entropy criterion. FL is defined as Equation 2.

$$FL(PT) = -GT\alpha(1-PT)^{\gamma}\log(PT) - (1-GT)\alpha PT^{\gamma}\log(1-PT) \tag{2}$$

where GT and PT are the ground-truth and predition, respectively; α and γ represent hyperparameters for calibration. We split our data, with 90% for training and 10% for validation. Ten-fold cross-validation was used. Intersection over union (IoU) and F-measure were employed to calculate the effectiveness of our approaches, where the best model per IoU metric was saved. IoU is a standard metric for image segmentation problems. It compares the predicted outcome and the gold standard. F-measure (F) is also commonly applied to image segmentation, and it is calculated as the harmonic mean between precision (P) and recall (R), defined in the following equation:

$$IoU = \frac{TP}{FP + TP + FN}, \quad R = \frac{TP}{TP + FN}, \quad P = \frac{TP}{TP + FP}, \quad F = 2 \times \frac{P \times R}{P + R} \tag{3}$$

where TP are the number of true-positive, TN are the number of the true-negative, FP are the number of the false-positive, and FN are the number of the false-negative findings. In total, 4,000 epoch iterations were used for training. In addition, we adopted a sigmoid activation function and we dichotomized the output at 0.5.

3.2.3. Anatomical Labeling and Volume Registration

In addition to the FLAIR volumetric information, we processed the T1-w MR scan for each participant in order to identify the anatomical brain regions. Freesurfer (Fischl et al., 2002) was employed to parcellate the brain into anatomical regions using the T1-w scan. We used Freesurfer's WMparc white-matter atlas to label the images, and grouped the WMparc labels by lobe (frontal, insula, occipital, parietal, temporal and unknown). We then registered the T1-w images to the FLAIR images using the mutual information cost function, and transferred the white-matter lobe labels to the FLAIR space.

3.3. Technical Details

We used a GPU (Nvidia Titan V with 12GB on Ubuntu 20.04 Linux) to train our U-Net. The training time varied depending on the selected architecture: VGG19 required 11h40m, VGG16 10h45m, ResNet152 the least amount of time at 7h40m, and EfficientNetB0 the most time at 50h52m). The implementation was performed using Python 3.6 and used Jupyter notebook to apply the learning step. The code is freely-available in our Github[1].

4. Results

This section consists of analyzing the efficiency of the training step of our approach, in addition to identifying the brain lobes where the predicted WMH were found. Two evaluation metrics were used both to optimize training efficiency and to analyze our results: IoU and F-measure. Figure 2 illustrates the training step over the epochs using the proposed metrics. The segmentation analysis is performed in 3D to evaluate whether the voxels are correctly labeled. Figure 3 presents some sample image segmentation. Figure 4 shows the F-measure for each participant and their corresponding Hausdorff distance (HD) calculation(Belogay et al., 1997). In addition, the labeling reliability by brain lobe is shown in Figure 5.

1. https://github.com/KaueTND/WMH_CNS

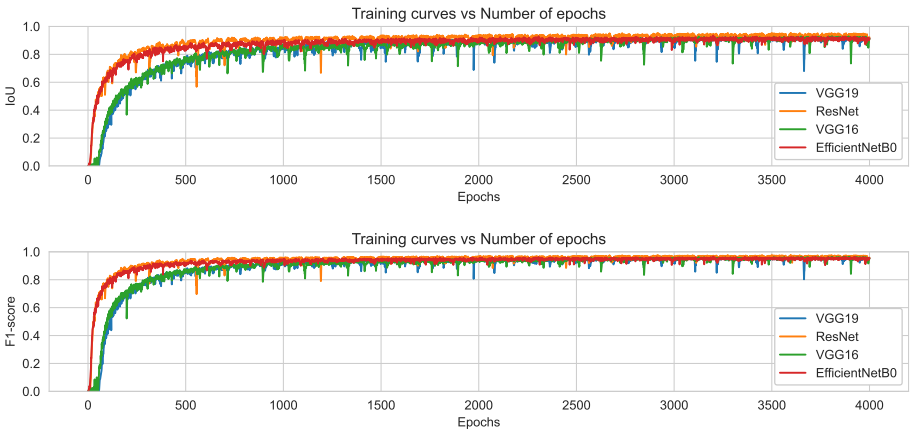


Figure 2: Learning step for each 3D U-Net architecture, compared via IoU and F-measure.

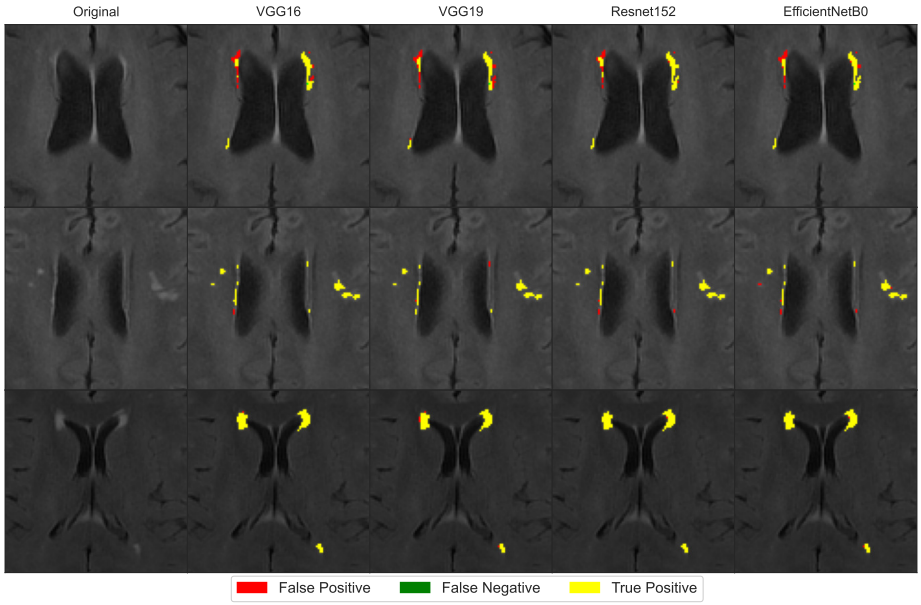
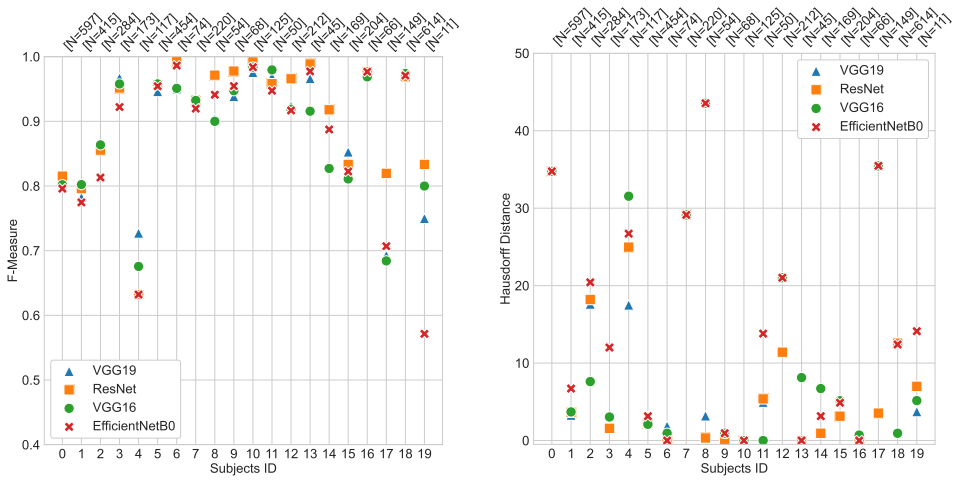


Figure 3: Segmentation outcome for training samples (zoomed-in for illustration) in each architecture.

5. Discussions

WMH segmentation is a challenging task because there is a high variability in size, location, and shape of the presumed lesions. Also, some WMH are highly localized and intense while others are diffuse. We proposed an approach to fulfill this task and to distinguish the affected areas using 3D U-Net CNNs. In total, four well-known architectures were adapted



(a) (b)

Figure 4: (a) F-measure by participant; and (b) F-measure evaluated by HD.

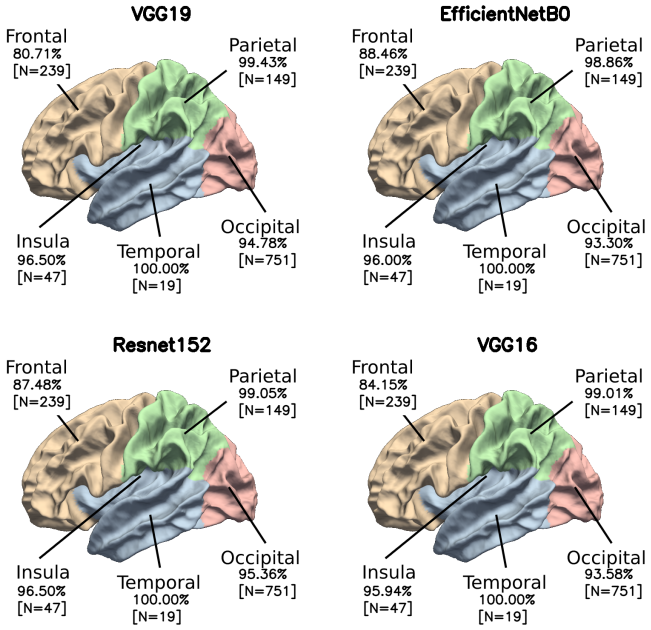


Figure 5: F-measure evaluation over all participants by brain lobes.

to perform the WMH segmentation. Surprisingly, the ResNet architecture was the least time-consuming, but also achieved highest performance in the early epochs of training. Both ResNet and EfficientNetB0 reached their performance plateau at earlier epochs than the other architectures, suggesting that the training step time could be potentially reduced if convergence was achieved, *i.e.*, in epoch #500 in this case. The EfficientNetB0 required

6× greater computational effort than ResNet per epoch, and therefore did not demonstrate greater efficiency than the other network architectures for this specific problem.

Because our classes (WMH versus non-WMH) were imbalanced and TN findings were disproportionately high, accuracy is not a suitable metric. Instead, we used IoU and F-measure. The HD provided a distinct insight about our results, which may lead to a better understanding of the predictions. Overall, the HD showed small values, which corresponds that ground-truth and predicted present small variation. The class imbalance is also the reason for using focal loss and Dice loss as the loss functions. The focal loss adds a dynamically scalar function to the binary cross-entropy, which reduces the importance of negative over positive classes. The scalar factor value is inversely proportional to the confidence of correct predictions, thus meaning that it reduces the true non-WMH segmentation importance over true WMH labels. Addressing the same issue, Dice loss also reduces the importance of background over foreground, even in complex scenarios.The brain lobe analysis provided some insights that requires further analysis: although the frontal lobe presented a lower F-measure than the other lobes and higher variance between networks, *e.g.*, 8% between VGG19 and EfficientNetB0, which might imply that WMH segmentation in the frontal lobe needs to be further studied. On the other hand, all other lobes presented only a small variation in F-measure. The temporal lobe achieved 100% on the F-measure, but we assumed that this was due to the small number of WMH voxels in the region. In addition, we highlighted the small quantity of WMH voxels per participant and how this finding might present more variety to the different architectures (Figure 5). Note that the participants in the study were presumed to be neurologically healthy.

We assumed that Resnet152 performed better segmentation than VGGs since it uses identity mapping, thus not increasing the complexity and the training error of the system. Unfortunately, we found no single architecture that provided the best overall WMH segmentation. We anticipate that increasing the number of training images might potentially improve the performance of the networks.

6. Conclusions

3D U-Net CNNs are powerful tools for segmentation of WMH in MR FLAIR images, thus enabling the identification of probable brain lesions. The challenge when using CNNs is the need to generalize the result, *i.e.*, correctly classifying data that was unseen during the training phase. The evaluation of the analyzed architectures demonstrates that the proposed approaches reliably segment WMH and identifies their location by brain lobe. However, only segmenting the WMH does not provide information about possible clinical indications. The location and increase with time of WMH burden is crucial to understand the impact of these changes on the patient. As suggested by our results, the location of the possible lesions were highlighted accurately and the segmentation was able to predict even subtle changes in size and shape. Our intention in identifying the white matter tissue hyperintensity is to have it serve as a predictor for vascular and related dementias. Future studies will increase the number of input images, in addition to classifying the input images according to the presence of dementia, stroke, or any other pathological disorder. Further work is needed to fully understand the difference in behavior of the networks in the frontal lobe. In addition, we aim to separate the para-ventricular WMH region from deep WMH .

Acknowledgments

KTND holds an Eyes High Postdoctoral Fellowship from the University of Calgary. The university's Evolve2Innovate (E2I) program supported some of these experiments. Other operational and infrastructure funding was provided by the Canadian Institutes for Health Research and the Canada Foundation for Innovation, respectively.

References

Jessica Alber, Suvarna Alladi, Hee-Joon Bae, and ... Atticus H. Hainsworth. White matter hyperintensities in vascular contributions to cognitive impairment and dementia (VCID): Knowledge gaps and opportunities. *Alzheimer's and Dementia: Translational Research and Clinical Interventions*, 5(1):107–117, January 2019. doi: 10.1016/j.trci.2019.02.001.

E. Belogay, C. Cabrelli, U. Molter, and R. Shonkwiler. Calculating the hausdorff distance between curves, 1997.

Michael T Duong, Jeffrey D Rudie, Jiancong Wang, Long Xie, Suyash Mohan, James C Gee, and Andreas M Rauschecker. Convolutional neural network for automated FLAIR lesion segmentation on clinical brain MR imaging. *American Journal of Neuroradiology*, 40(8):1282–1290, 2019. ISSN 0195-6108. doi: 10.3174/ajnr.A6138.

Bruce Fischl, David H. Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre van der Kouwe, Ron Killiany, David Kennedy, Shuna Klaveness, Albert Montillo, Nikos Makris, Bruce Rosen, and Anders M. Dale. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3): 341–355, 2002. ISSN 0896-6273. doi: 10.1016/S0896-6273(02)00569-X.

David Gobbi, Qian Lu, Richard Frayne, and Marina Salluzzi. Cerebra-WML: a rapid workflow for quantification of white matter hyperintensities. pages E128–E129, 11 2012.

Ricardo Guerrero, Chen Qin, Ozan Oktay, Christopher Bowles, Liang Chen, Richard Joules, Robin Wolz, Maria del C Valdés-Hernández, David A Dickie, Joanna M Wardlaw, and Daniel Rueckert. White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks. *NeuroImage: Clinical*, 17:918–934, 2018. ISSN 2213-1582. doi: 10.1016/j.nicl.2017.12.022.

Xinxin Li, Yu Zhao, Jiyang Jiang, Jian Cheng, Wanlin Zhu, Zhenzhou Wu, Jing Jing, Zhe Zhang, Wei Wen, Perminder S. Sachdev, Yongjun Wang, Tao Liu, and Zixiao Li. White matter hyperintensities segmentation using an ensemble of neural networks. *Human Brain Mapping*, Oct 2021. doi: 10.1002/hbm.25695.

Cheryl R McCreary, Marina Salluzzi, Linda B Andersen, David Gobbi, Louis Lauzon, Feryal Saad, Eric E Smith, and Richard Frayne. Calgary normative study: design of a prospective longitudinal study to characterise potential quantitative MR biomarkers of neurodegeneration over the adult lifespan. *BMJ Open*, 10(8), 2020. ISSN 2044-6055. doi: 10.1136/bmjopen-2020-038120.

Jeffrey D. Rudie, David A. Weiss, Rachit Saluja, Andreas M. Rauschecker, Jiancong Wang, Leo Sugrue, Spyridon Bakas, and John B. Colby. Multi-disease segmentation of gliomas and white matter hyperintensities in the BraTS data using a 3D convolutional neural network. *Frontiers in Computational Neuroscience*, 13, Dec 2019. doi: 10.3389/fncom.2019.00084.

Vaanathi Sundaresan, Giovanna Zamboni, Peter M. Rothwell, Mark Jenkinson, and Ludovica Griffanti. Triplanar ensemble U-Net model for white matter hyperintensities segmentation on MR images. *Medical Image Analysis*, 73, Oct 2021. ISSN 1361-8415. doi: 10.1016/j.media.2021.102184.

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.