Proceedings of Machine Learning Research 1–11

Triplanar Ensemble U-Net Model for White Matter Hyperintensities Segmentation on MR Images

Siwar Mhadhbi SIWAR.MHADHBI@TELECOM-PARIS.FR
Eya Ghamgui EYA.GHAMGUI@TELECOM-PARIS.FR

Background

White Matter Hyperintensities (WMHs) are lesions observed in the brain as bright regions in Fluid Attenuated Inversion Recovery (FLAIR) images from Magnetic Resonance Imaging (MRI). Their presence is related to conditions such as aging, small vessel disease, stroke, depression, and neurodegenerative disease. Studying WMH lesions on these types of images, through a correct and accurate segmentation process, would improve the understanding of brain damage and associated cognitive and physical problems, and provide early diagnosis. Nevertheless, robust segmentation of WMHs is challenging due to differences in acquisition, population, and limited availability of manual segmentations. Therefore, many methods have been proposed in this context. Triplanar ensemble U-Net model for White Matter Hyperintensity segmentation on MR images is one of the solutions proposed by (Sundaresan et al., 2021). It consists of three 2D U-Nets each applied on different planes (Axial, Sagittal, and Coronal) of structural MR images of the brain. The models are combined and processed in parallel in an ensemble model. Concerning the loss function, the proposed loss uses anatomical maps regarding the spatial distribution of WMHs to balance the contrast between deep and periventricular WMHs. Finally, this model is evaluated on 5 datasets. Three of these datasets are publicly available for the 2017 MICCAI WMH Segmentation Challenge. This model is also evaluated on the 2017 MWSC and other unseen test datasets.

In addition to this method, there are many other methods used for WMHs segmentation. (Oktay et al., 2018) uses an adaptive outlier detection approach to detect outliers in the pixel intensity distribution. (Samaille et al., 2012) selects WMHs according to an automatically calculated threshold based on the subject and its anatomical information using nonlinear diffusion filtering alternated with watershed segmentation to obtain piecewise constant images with increased contrast between WMHs and surrounding tissue. These two methods deal with 2D images. There are others that use 3D images. (Griffanti et al., 2018) classifies image voxels based on their intensity and spatial features. Other approaches are statistically based methods. (Jack Jr et al., 2001) segment the image histogram into regions that are classified as lesion, CSF, and normal brain. As limitations, all these methods require high image quality, which is not always the case. In addition, the detection of the region of interest is influenced by demographic factors, clinical conditions, artifacts, and noise. Thus, preprocessing of the data (e.g., morphological operations) is mandatory. Furthermore, these methods are only applied to a small number of populations due to the limitation of computational time and memory load.

On the other hand, deep learning models are widely used in WMHs segmentation, offering improved performance over traditional models, as they use larger datasets that contain a variety of samples.

© S. Mhadhbi & E. Ghamgui.

(Ronneberger et al., 2015) has implemented the U-Net architecture which is an encoder-decoder architecture based on fully convolutional neural networks. However, using this method can lead to segmentation discontinuities between different planes of 3D images since this method deals only with 2D images. Thus, the use of 3D models can solve this issue. (Anbeek et al., 2004) get benefits from multi-dimensional gated recurrent units to generate an accurate method in terms of speed, accuracy and memory efficiency applied to 3D data. (Li et al., 2018) proposed a fully convolutional deep network using 3D CNNs within an ensemble model to automatically detect WMHs. Nevertheless, processing 3D data requires a higher number of parameters compared to 2D models which demand more computational time and resources to train the parameters. Added to that, small lesions are not present on a large number of slices, which limits the interest of 3D representations. One of the solutions proposed by (Xu et al., 2017) is therefore to use a VGG neural network pre-trained on the ImageNet dataset and refined on the MICCAI WMH dataset. It uses 2D segmentation of three modalities and then concatenates all results to form a 3D brain volume.

In our course, we discussed the 3D U-Net architecture in more detail. Thus, instead of using the ensemble of 2D U-Nets, we can use a 3D U-Net one at a time. (Li et al., 2017) also proposed HighRes3DNet which is a compact high-resolution convolutional network for volumetric image segmentation. The network mainly uses dilated convolutions and residual connections to build an end-to-end mapping of the image volume to dense voxel-level segmentation. We have also seen Multi-Scale processing for accurate lesion segmentation proposed by (Kamnitsas et al., 2017). It uses small convolutional kernels in 3D CNNs, which allows for the development of a deeper and more discriminative network without increasing the computational cost and the number of trainable parameters. Furthermore, it has been shown to be able to segment even small and diffuse pathologies.

1. Objective & Main Contributions

1.1. Objective

The objective of this paper is to provide highly accurate segmentation of WMHs with respect to both periventricular and deep regions.

1.2. Main Contributions

- This paper focuses on the potential contribution of an encoder-decoder method (specifically, U-Net) that combines the three different medical imaging planes (Axial, Sagittal and Coronal) of brain MR images to provide accurate WMH segmentation. Indeed, the 3D models retain more contextual information and guarantee the continuity of the segmentation in the z-dimension since the individual slices are not considered separately.

- Two main contributions of this paper lie in the loss function. First, the method provides an accurate segmentation, independent of the location and load of the lesions. Indeed, the loss function takes into account information about the anatomical location and spatial distribution of WMHs. Second, the method ensures a robust segmentation of deep and periventricular WMHs thanks to the information incorporated in the loss function that helps overcome the contrast variations between these two regions.

- Another valuable contribution is that the method overcomes the high variability of the lesion characteristics depending on their location by providing an exact voxel-level localisation thanks to the voxel-wise terms of the loss function.
- One other interesting contribution of the paper is its validation/evaluation procedure. In fact, the method is evaluated on five different data sets and also has been submitted to MICCAI Challenge (MWSC).

1.3. Datasets used for training and validation

At first, the underlying method was trained on three different datasets with different acquisition and lesion characteristics. The first dataset is the NDGEN, a neurodegenerative cohort data that includes T1-weighted and FLAIR images. The second dataset is the MICCAI Challenge training set that includes images with the same acquisition protocol as NDGEN but from three different sources. And the third dataset is the OXVASC, a vascular cohort data that comprises the same acquisition as the previous datasets but adding diffusion-weighted images. Then, the method was evaluated on five different datasets also with different acquisition and lesion characteristics. Three of these datasets are NDGEN, MICCAI Challenge training set, and OXVASC. Additionally, the method was evaluated on two unseen datasets which are the MWSC which is an in-house dataset used by the organizers, and the test set of the MICCAI challenge which includes, in addition to the previously mentioned acquisition, 1.5T and PET-MR images.

2. Methodology

2.1. Pre-processing step

The underlying model takes T1-weighted images and FLAIR images as input. Thus, in order to standardize these two types of images, the authors first proceed by reorienting the images to the standard MNI space. This is legitimate because we should make images from different studies comparable and uniform by aligning them to a standard space so that they can be fed to the same model. Next, they perform skull-stripping and correct bias field using FSL FAST BET. After that, a linear rigid-body registration to FLAIR images is applied on T1-weighted images. Finally, a Gaussian kernel is applied to normalize the intensity values. For model input, slices are extracted and cropped to a specific size for each of the three planar objects.

2.2. Architecture

The Triplanar U-Net ensemble network (TrUE-Net) is a Deep Learning method for WMH segmentation, consisting of an ensemble of three parallel U-Nets, each applied to one of the three planes (Axial, Sagittal and Coronal) of structural MRI images of the brain. Figure 1 shows the overall architecture. For each plane, a deep 3-layer 2D U-Net model takes the extracted 2D FLAIR and T1-weighted slices as input channels and provides the probability map in the corresponding plane. The reason for choosing the selected depth is not only to reduce the computational burden but also to improve the sensitivity of the model towards small lesions. In the sagittal and coronal U-Nets, the first layer uses 5×5 convolutional kernels in order to reduce discontinuities across slices in these planes. In fact, a larger receptive field allows learning more generic lesion patterns. Whereas in all other layers, the architecture uses 3×3 convolutional kernels. Each convolution layer is followed by

a batch normalization layer and ReLU as the activation function. At the end, a 1×1 convolutional kernel is applied followed by a softmax layer in order to predict the probability maps. Each of the three 2D U-Nets associated with an individual plane is trained independently. During testing, for each of the three 2D U-Nets, the 2D probability maps are resized to their original dimensions and then assembled into a 3D probability map. Thereafter, in a final step, the three 3D probability maps from the three planes are averaged together, resulting in the final 3D probability map representing the segmentation along Axial, Sagittal, and Coronal planes.

2.3. Loss function

WMH segmentation is known as an unbalanced class problem. If training is performed as usual, an unbalanced class leads to an unreliable training that yields only the majority class. The authors of the underlying paper used the sum of a voxel-weighted cross-entropy (CE) and the Dice loss (DcL). A well-known drawback of the per-voxel CE loss is that it is biased toward the detection of larger periventricular WMH. To address this problem, the authors used a weight map based on distances from ventricles and Gray Matter (GM). The latter penalizes more the deep areas that contain less WMH compared to the background. This term helps in the segmentation of periventricular regions. As for the deep regions, a second term is added which is the Dice loss. This in turn helps to detect small lesions especially in areas of high class imbalance.

3. Validation & Results

3.1. Validation procedure

For evaluation, the authors used several evaluation metrics. They considered the Dice Similarity Index (SI), True Positive Rate (TPR) per voxel, False Positive Rate (FPR) per voxel, TPR per cluster, Absolute Volume Difference (ADV), F1 score per cluster, and the 95th percentile of the Hausdorff distance (H95). Due to noise in manual and predicted segmentation, the authors used cluster-based evaluation methods based on 26 connected neighboring areas.

The model is first trained using the NDGEN dataset to tune the parameters. The authors studied the impact of varying the number of epochs, the batch size, the learning rate, and the epsilon value of the adam optimizer on the convergence of the model. To make their study more concise, they added an ablation study of the loss function, removing one component of the loss function at a time and then training the model on the NDGEN dataset to compare the cases: cross-entropy loss, weighted cross-entropy loss, weighted cross-entropy loss + Dice loss and Dice loss. On the other hand, they investigated the impact of model size on segmentation performance using the previous dataset. They compared 3D U-Net, 2D U-Net (on the Axial plane only) and TrUE-Net. In both studies, they determined SI values in the periventricular and deep regions separately and performed a Wilcoxon signed rank test. This test is conducted mainly to examine the difference between models.

After tuning the parameters and choosing the best combination of losses, the authors proposed to evaluate the model using the Leave One Out method, as it uses larger training data and provides more reliable estimates in smaller data sets. To do this, they computed metrics for three-fold cross-validation on the MWSC dataset. Then, they evaluated it on a new unseen dataset (MWSC 2017).

After that, they compared the proposed model with the top ranked method from MWSC 2017 (Kuijf et al., 2019) using the MWSC dataset with Leave One Out validation method. In addition, they compared the performance of their model with the model proposed by (Li et al., 2018) by training and evaluating the models on the OXVASC and NDGEN datasets using the same evaluation method. They kept the same parameters proposed in (Li et al., 2018) for the model. As before, the Wilcoxon signed rank test is performed to compare the two models. Finally, they completed their study with the other existing methods found in the literature (until 2019) and included only the methods with promising performance.

3.2. Main results

- The weighted cross entropy term in loss indeed helped overcome the class imbalance between WMH and non-WMH voxels. As we can observe in Figure 2, the segmentation using the weighted cross entropy doesn't extend beyond the ventricle lining and also is more accurate compared to the manual segmentation. Also, adding the Dice term in the loss function led to an even more accurate segmentation especially regarding PWMH borders (Figure 2.e).

- The method leads to continuous segmentation of WMHs. This is thanks to the contextual information used from the ensemble on the three different MRI planes. In turn, this leads to more accurate segmentation (Figure 3.e) compared to the results given with 2D U-Nets.

- A worth mentioned result is that the model, in specific cases, could even outperform the manual segmentation provided. For instance, in Figure 4, the manual segmentation considers the circled cerebrospinal fluid (CSF) as WMH region while it is successfully discarded by the TrUE-Net model.

- The model performs as well in deep regions as in periventricular regions according to the results given in Figure 5. Most of the metrics values represented as boxplots are not significantly different between PWMHs and DWMHs. Thus, we can say that the proposed approach provides good delineations of WMHs, with good sensitivity in both regions.

- Comparing the results of TrUE-Net and BIANCA (Figure 6), it is clear that BIANCA detects more false positives than TrUE-Net, in particular around the ventricles and especially in the low lesion load case. This means that the TrUE-Net method is significantly better especially regarding borders delimitation as also explained in the first point.

- Compared to the top-ranked methods in the MICCAI Challenge (MWSC), the method performs poorly regarding cluster-wise detection in spite of high SI values. This shows that even though TrUE-Net provides better segmentation of the true detected WMHs, it still misses small and subtle WMHs in the unseen test data sets.

Conclusion

In this work, a deep learning model using an ensemble of U-Nets on the three different MRI images (Axial, Sagittal and Coronal) has been proposed to achieve an accurate segmentation of WMHs better than BIANCA and on par with the top ranking methods of MWSC 2017. This architecture indeed combines the advantages and overcomes the disadvantages of both 3D U-Net and 2D Axial U-Net architectures in the sense that the approach detects most PWMHs lesions and smaller DWMHs

lesions while providing continuous boundary delineation. Therefore, the main result provided by the triplanar model is accurate, sensitive and precise detection. Moreover, the triplanar model not only outperforms the 3D model, but also requires far fewer parameters, allowing for both faster and lighter calculations. Another worth-mentioning point is that the proposed weighting procedure of cross-entropy that is based on distances from ventricles and GM has proven to be very effective in overcoming the over-segmentation problem and providing accurate segmentation in deep regions regardless of lesion load. Furthermore, being trained and evaluated on various modalities, the approach is more generalizable to other data sets, unlike other methods in the literature, which may require additional experiments to validate/improve their generalization.

However, even though the model was trained on data from different cohorts and evaluated on five different datasets, the model performance is still dependent on the dataset. In fact, the method detects more true positive lesions on certain datasets, while the lesion boundaries are better delineated in other datasets.

Although the approach performs very well, it has some limitations. First, as mentioned by the authors, the application of the white matter mask as a post-processing step may affect the segmentation of DWMHs near the WM-GM interface. Second, it was mentioned that among the five validation data sets, only MWSC includes samples with very low lesion load. Thus, it would be more interesting to vary the five data sets to incorporate samples with approximately equal proportions of high and low lesion load to further generalize the performance of the proposed method, which can explain the low cluster-wise performance when compared to the top ranking methods. Therefore, exploring various domain adaptation techniques to improve WMH detection in various unseen datasets would be an interesting ongoing work for the underlying article.

Appendix

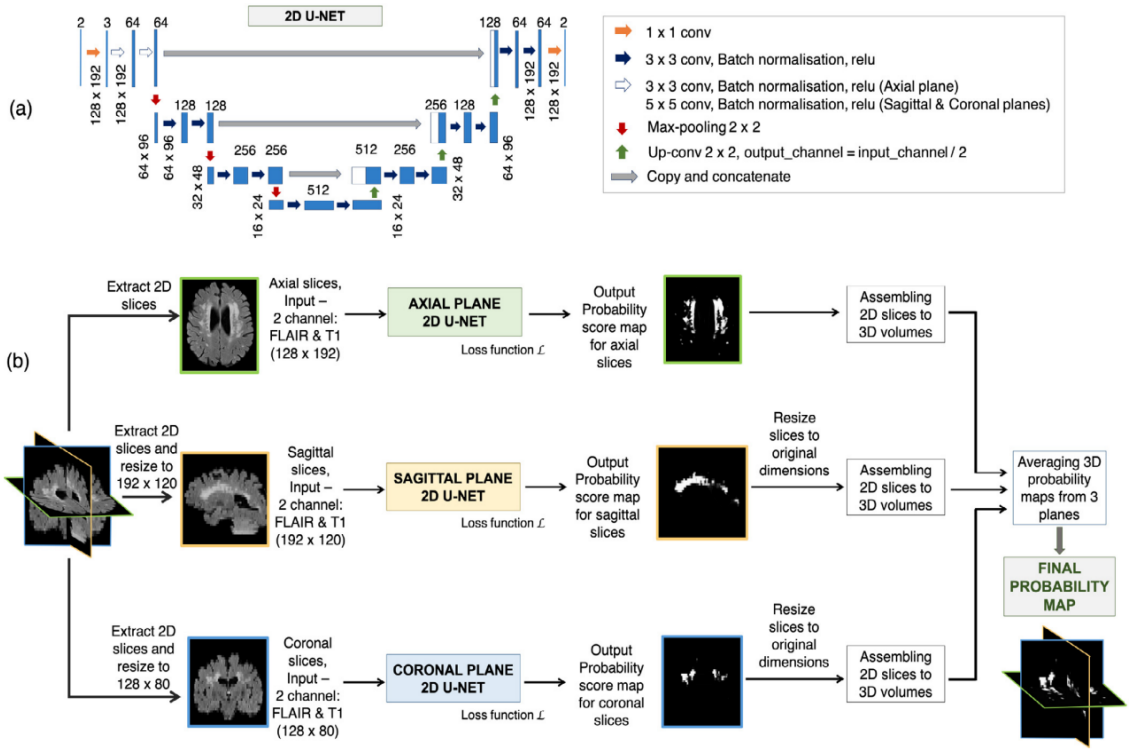

Figure 1: TrUE-Net Architecture

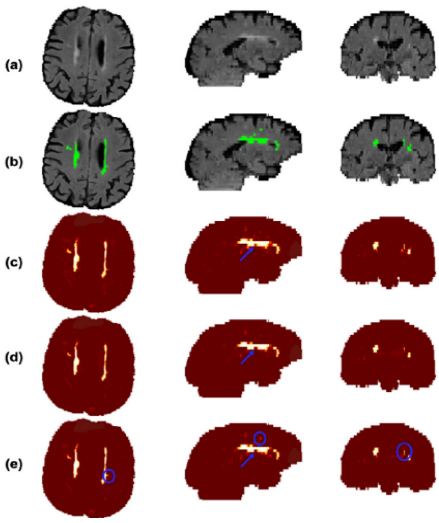

Figure 2: (a) FLAIR images. (b) Manual segmentation. (c) Result with CE loss. (d) Result with weighted CE loss. (e) Result with weighted CE loss + Dice loss.

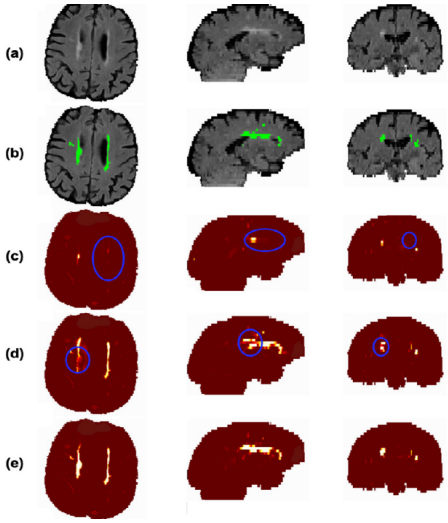

Figure 3: (a) FLAIR images. (b) Manual segmentation. (c) 3D U-Net. (d) 2D Axial U-Net. (e) TrUE-Net.

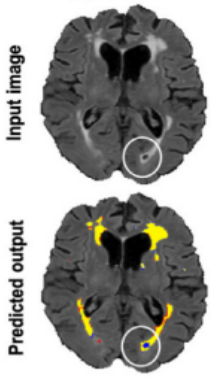

Figure 4: Example TrUE-Net discards CSF region labeled as WMH region in manual segmentation.

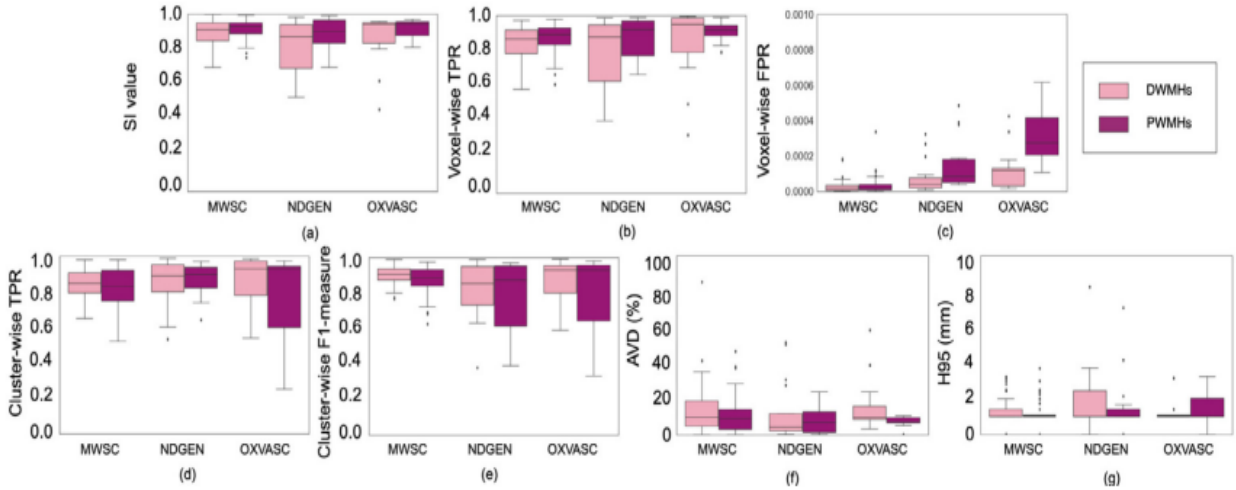

Figure 5: Boxplot performance metrics from LOO evaluation method.

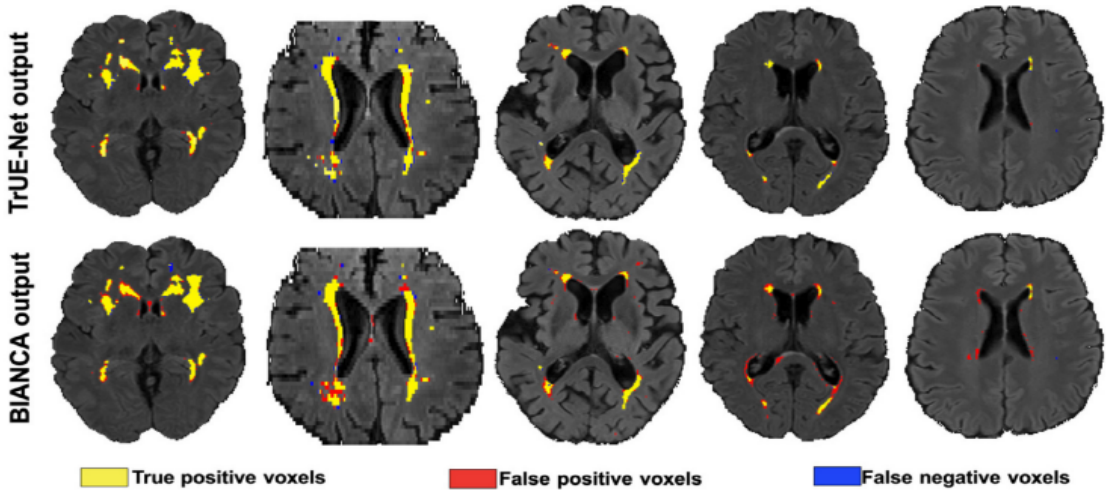

Figure 6: Comparison between TrUE-Net and BIANCA segmentation results.

References

Petronella Anbeek, Koen L Vincken, Matthias JP Van Osch, Robertus HC Bisschops, and Jeroen Van Der Grond. Probabilistic segmentation of white matter lesions in mr imaging. *NeuroImage*, 21(3):1037–1044, 2004.

Ludovica Griffanti, Mark Jenkinson, Sana Suri, Enikő Zsoldos, Abda Mahmood, Nicola Filippini, Claire E Sexton, Anya Topiwala, Charlotte Allan, Mika Kivimäki, et al. Classification and characterization of periventricular and deep white matter hyperintensities on mri: a study in older adults. *Neuroimage*, 170:174–181, 2018.

Clifford R Jack Jr, Peter C O'Brien, Daniel W Rettman, Maria M Shiung, Yuecheng Xu, Raja Muthupillai, Armando Manduca, Ramesh Avula, and Bradley J Erickson. Flair histogram segmentation for measurement of leukoaraiosis volume. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 14(6):668–676, 2001.

Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, 36:61–78, 2017.

Hugo J Kuijf, J Matthijs Biesbroek, Jeroen De Bresser, Rutger Heinen, Simon Andermatt, Mariana Bento, Matt Berseth, Mikhail Belyaev, M Jorge Cardoso, Adria Casamitjana, et al. Standardized assessment of automatic segmentation of white matter hyperintensities and results of the wmh segmentation challenge. *IEEE transactions on medical imaging*, 38(11):2556–2568, 2019.

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, 2018.

Wenqi Li, Guotai Wang, Lucas Fidon, Sebastien Ourselin, M Jorge Cardoso, and Tom Vercauteren. On the compactness, efficiency, and representation of 3d convolutional networks: brain parcellation as a pretext task. In *International conference on information processing in medical imaging*, pages 348–360. Springer, 2017.

Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

Thomas Samaille, Ludovic Fillon, Rémi Cuingnet, Eric Jouvent, Hugues Chabriat, Didier Dormont, Olivier Colliot, and Marie Chupin. Contrast-based fully automatic segmentation of white matter hyperintensities: method and validation. *PloS one*, 7(11):e48953, 2012.

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*, 73:102184, 2021.

Yongchao Xu, Thierry Géraud, Élodie Puybareau, Isabelle Bloch, and Joseph Chazalon. White matter hyperintensities segmentation in a few seconds using fully convolutional network and transfer learning. In *International MICCAI Brainlesion Workshop*, pages 501–514. Springer, 2017.