medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

WHITE-Net : White matter HyperIntensities Tissue Extraction using deep learning Network

Camille Cathala [1,2*], Ferath Kherif [1], Jean-Philippe Thiran [2], Aurélie Bussy [1,3†], Bogdan Draganski [1,3*†]

[1]Laboratoire de Recherche en Neuroimagerie, CRN-DNC, CHUV, University of Lausanne, Switzerland

[2]Signal Processing Laboratory 5, EPFL-STI-IEL-LTS5, Station 11,CH-1015 Lausanne, Switzerland

[3]Neurology and University Institute for Diagnostic and Interventional Neuroradiology - Inselspital, University Hospital Bern, University Bern, Rosenbuehlgasse 25, CH-3010 Bern, Switzerland

*Correspondence: camille.cathala@alumni.epfl.ch, bogdan.draganski@unibe.ch

† Equal contribution

January 7, 2025

Keywords: MRI, Deep Learning, ResUNet, Neural Network, White Matter Hyperintensities, Aging, Image Segmentation

Abstract: Given the high prevalence of aging-associated cerebral small vessel disease in the general population, accurate detection of the related white matter hyperintensities (WMH) in large-scale magnetic resonance imaging (MRI) studies is of critical importance. The performance of currently available semi-automated and automated methods for WMH classification is hampered by their inherent dependence on MRI contrast parameters and long computational processing time. We sought to improve the accuracy and computational cost of automated WMH detection by creating a whole-brain deep learning-based framework: WHITE-Net. We use a 3D ResUNet architecture trained on manually segmented WMHs from fluid-attenuated inversion recovery MRI (n=141) and test its accuracy in a large-scale dataset (n=192). We demonstrate a good generalizability across WMH lesion loads, different MRI scanner vendors, field strengths, imaging protocols, and MR contrasts. The comparison

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

to existing WMH segmentation tools shows a similar to superior accuracy performance at significantly lower computational cost. WHITE-Net tool performance makes it well-suited for application to large-scale MRI datasets, enabling the study of the aging brain while offering the advantage of detecting early or subtle WMH changes often missed by other methods.

1 Introduction

White Matter Hyperintensities (WMH) are areas of abnormal signal intensity on diagnostic magnetic resonance imaging (MRI) protocols, such as fluid attenuated inversion recovery (FLAIR) (Figure 1). Their etiology varies depending on the context: WMHs of presumed vascular origin are strongly associated with cardiovascular risk factors (Aljondi et al., 2018; de Bresser et al., 2018; Fuhrmann et al., 2019; Ghaznawi et al., 2018; Kuller et al., 2004; Trofimova et al., 2023), while in migraine (Kruit et al., 2009; Seneviratne et al., 2013), and multiple sclerosis (Fazekas et al., 1999; Filippi et al., 2019; Rovira & León, 2008) the underlying processes are inflammatory or unknown. Vascular WMHs are strongly associated with aging and represent a common incidental finding in the general population above the age of 60 years. They represent one of the hallmarks of small vessel disease (SVD), which is established as a major factor for vascular cognitive impairment (Coenen et al., 2022; Jiménez-Balado et al., 2022), and vascular dementia (Lee et al., 2016; Mortamais et al., 2013).

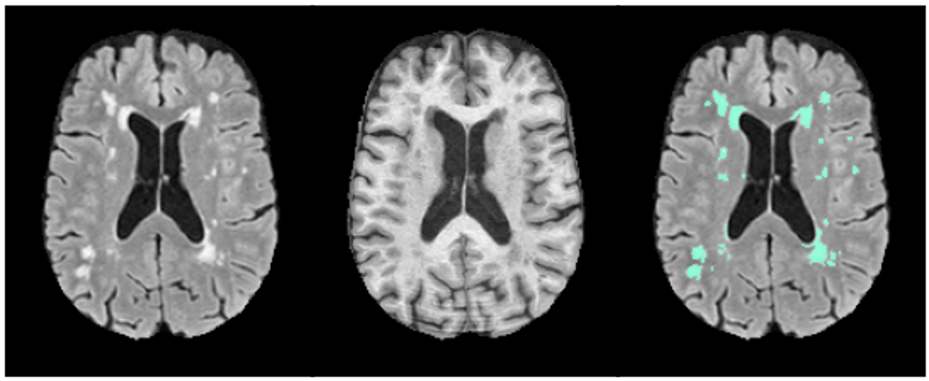

Figure 1: Example of white matter hyperintensities (WMH) on magnetic resonance imaging (MRI). From left to right : fluid attenuated inversion recovery (FLAIR), T1-weighted MRI and WMH mask overlaid in blue on FLAIR MRI.

The importance of an automatic and accurate detection of WMHs complementary to the neuroradiologist' diagnostic expertise, is recognized in both the clinical and research setting. The vast amount of data in large-scale computational anatomy studies using MRI that preclude a manual delineation, make the case of automated WMH detection indispensable. Despite its shortcomings in terms of time-intensive labor and

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

individual rater dependency (Commowick et al., 2018; Grimaud et al., 1996), the manual delineation of WMHs is still considered the gold standard in neuroimaging. The currently existing WMH segmentation algorithms span from intensity-based methods (Gaser et al., 2022; Schmidt et al., 2012), to machine learning (Griffanti et al., 2016) and deep learning algorithms (Li et al., 2018; Park et al., 2021).

The intensity-based methods rely on predefined or automatically determined thresholds that identify WMHs based on voxel intensity and morphology. While straightforward, these approaches are often limited by their inherent sensitivity to noise. Machine learning approaches overcome this by using trained classifiers on features such as intensity and spatial location, to better adapt to lesion variability, but they may lack generalizability across datasets. Deep learning models, particularly convolutional neural networks (CNNs), automatically learn complex features from the data, improving segmentation of WMH lesions, but require large datasets and higher computational power for efficient training.

The existing methods offer clear advantages in terms of accuracy and time efficiency compared to standard manual detection. However, most of these algorithms are trained on data comprising well-defined clinical disorders, such as multiple sclerosis (Schmidt, 2017; Schmidt et al., 2012), stroke or neurodegeneration (Griffanti et al., 2016; Li et al., 2018; Park et al., 2021), reducing their generalizability to epidemiological MRI datasets. The emergence of large-scale studies in the community-dwelling population that include brain MRI acquisition like the UK Biobank (Sudlow et al., 2015) aiming at 100 '000 MRIs demonstrate the need for accurate and automatic WMH segmentation tools. Robust and accurate WMH segmentation in a cohort characterized by a wide age range, image quality variability and prone to aging-associated pathologies is necessary to ensure the validity of any computational anatomy study using MRI data.

In this context, we present WHITE-Net (White matter HyperIntensities Tissue Extraction using deep learning Network), a user-friendly, fast, and generalizable deep learning algorithm for WMH segmentation based on a 3D ResUNet architecture. We train the model on epidemiological MRI data, encompassing individuals with a wide range of lesion topographies and extents. Additionally, to assess the external validity and generalizability of the model, we test WHITE-Net on open-access MRI data including protocols with varying spatial resolution acquired using different MRI scanners. The performance of WHITE-Net is evaluated by comparing the results with manual delineation and established automated lesion segmentation tools.

Overall, the results suggest that WHITE-Net is a reliable method to rapidly and accurately detect WMHs on different datasets. The assessed performance of WHITE-Net demonstrates its efficient application to data stemming from different scanners, including images with anisotropic voxel size, as well as to a large spectrum of participants ranging from healthy individuals to those with neurodegenerative and vascular pathologies.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

2 Methods

2.1 Datasets

For this study, we use two MRI datasets with distinct imaging protocols and participants' characteristics. The BrainLaus dataset comprises community-dwelling participants from a large-scale epidemiological cohort, while the second, the WMH challenge dataset, includes individuals with cardiovascular and neurodegenerative pathologies. These datasets are acquired using different MRI imaging protocols and scanners.

2.1.1 BrainLaus dataset

The first set of MRI data used in this study comes from the BrainLaus study, a nested project within the CoLaus|PsycoLaus cohort, which investigates the link between cardiovascular risk factors and mental health in the general population (Firmann et al., 2008; Preisig et al., 2009). The study was approved by the Ethics Commission of Canton de Vaud and individuals were included after signing an informed consent.

2.1.2 WMH challenge dataset

The second source of MRI data comes from the WMH Segmentation Challenge (H. Kuijf et al., 2022), a recent initiative created to compare WMH segmentation algorithms. This dataset explicitly includes subjects with neurodegenerative and cardiovascular pathologies and different acquisition protocols.

2.2 MRI acquisition

2.2.1 BrainLaus dataset

MRI data are acquired on a 3T whole-body system (Magnetom Prisma - Siemens, Erlangen Germany), with a 64-channel radiofrequency receive head coil and a body coil for transmission. Details of the imaging protocol for both T1-weighted and FLAIR sequences are provided in Table 1.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

2.2.2 WMH challenge dataset

The MRI data are acquired from five different MR scanners. The characteristics of each scanner is presented in Table 1 (Boomsma et al., 2017; van Veluw et al., 2015).

Table 1: MRI Sequence Parameters.

Institution	Scanner	Sequence	Slices	Voxel Size (mm^3)	TR/TE/TI (ms)
BrainLaus Lausanne	3T Siemens Magnetom Prisma	3D T1-weighted 3D FLAIR	176 176 transversal	1.00 × 1.00 × 1.00 1.00 × 1.00 × 1.00	2000/2.39/920 5000/389/1800
UMC Utrecht	3T Philips Achieva	3D T1-weighted 2D FLAIR	192 48 transversal	1.00 × 1.00 × 1.00 0.96 × 0.95 × 3.00	7.9/4.5/- 11000/125/2800
NUHS Singapore	3T Siemens TrioTim	3D T1-weighted 2D FLAIR	- transversal	1.00 × 1.00 × 1.00 1.0 × 1.0 × 3.00	2300/1.9/900 9000/82/2500
VU Amsterdam	3T GE Signa HDxt	3D T1-weighted 3D FLAIR	176 132 sagittal	0.94 × 0.94 × 1.00 0.98 × 0.98 × 1.20	7.8/3.0/- 8000/126/2340
	3T Philips Ingenuity	3D T1-weighted 3D FLAIR	180 321 sagittal	0.87 × 0.87 × 1.00 1.04 × 1.04 × 0.56	9.9/4.6/- 4800/279/1650
	1.5T GE Signa HDxt	3D T1-weighted 3D FLAIR	172 128 sagittal	0.98 × 0.98 × 1.50 1.21 × 1.21 × 1.30	12.3/5.2/- 6500/117/1987

2.3 MRI preprocessing

2.3.1 BrainLaus dataset

For preprocessing of T1w and FLAIR images, we use the SPM12-based CAT12 toolbox (Gaser et al., 2022) using MATLAB_R2021 (MathWorks, Sherborn, MA USA). After denoising using a spatial adaptive non-local means denoising filter (Manjón et al., 2009), we apply a bias field correction followed by a rigid transformation to align FLAIR and T1w images in the same space. Data quality assessment includes visual quality control for motion artifacts (Wood & Henkelman, 1985). Motion is assessed with a scale from 1 to 4, corresponding to minimal, mild, moderate, and severe degree of motion, with 0.5 increments to provide more granularity. Only the images with a quality score below 2.5 are selected in the analyses. Additionally, we assess the FLAIR images for WMH load and severity using the Fazekas scale (Fazekas et al., 1987). This scale categorizes WMH into periventricular (PVWM) and deep white matter (DWM) regions, assigning grades from 0 to 3 based on lesion size and confluence (0 indicating no lesion and 3 indicating extended lesions). In our study, scores from both regions are summed to create an overall lesion severity score ranging from 0 to 6.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

2.3.2 WMH challenge dataset

The preprocessing steps follow the defaults described in the WMH segmentation challenge publication (H. J. Kuijf et al., 2019). FLAIR and T1w images are bias-field corrected using SPM12 (Ashburner & Friston, 2005), followed by image registration of the T1w and the resampled FLAIR using the Elastix toolbox (Klein et al., 2010).

2.3.3 Masks and inputs preparation

Many WMH algorithms require a WM mask to effectively constrain the search volume for lesions. However, creating a WM mask can be challenging since WMH voxels have hypointense intensities in T1w images, making them similar to gray matter voxel intensities (Karim et al., 2016; Levy-Cooperman et al., 2008). To create the WM mask, we use FreeSurfer's segmentation tool Synthseg (Billot et al., 2023), which empirically demonstrated the highest robustness against this potential bias.

After subject-specific SPM12's segmentation-based skullstripping, we apply min-max normalization to all images followed by removal of the background and matrix size standardization to $192 \times 176 \times 160$ voxels with zero-padding. If the size before padding does not match, the image is resampled and then padded. In such cases, WM and WMH mask outputs are recovered to their original dimensions, and all subsequent analyses are performed using these original dimensions. For further analysis, the WM mask obtained with FreeSurfer is divided in 3 regions: deep white matter (DWM), periventricular white matter (PVWM) and superficial white matter (SWM). The SWM mask is obtained by creating a 2mm thick outer contour of the WM mask. The PVWM mask is created by dilating the cerebro-spinal fluid (CSF) mask obtained with FreeSurfer (Billot et al., 2023) by 8 mm. The DWM is defined by subtracting the PVWM and SWM masks from the total WM mask.

2.4 Manual segmentation

2.4.1 BrainLaus dataset

For the WMHs manual labeling in 163 individuals FLAIR MRIs, we use multiple raters with one rater assigned per image. The labeling process is carried out using the Display MincTool for visualization. Participants are pseudo-randomly selected from the BrainLaus dataset using selection criteria for a balanced distribution across Fazekas categories (0: 23%, 1: 13%, 2: 13%, 3: 13%, 4: 13%, 5: 12%, and 6: 11%).

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

2.4.2 WMH challenge dataset

WMHs delineation is performed by four raters per scan (H. J. Kuijf et al., 2019). Contours are converted to binary masks such that all voxels that have 50 % of the volume within the manual delineation are considered as WMH.

2.5 Dataset description

2.5.1 BrainLaus dataset

The dataset comprises 163 participants from the community-dwelling population, pseudo-randomly selected based on image quality and Fazekas score (mean age: 59.65 years (min: 19.78, max: 92.83); 56.6 % females, 44.4 % males).

50 % of the individuals (n=81) are randomly selected for the training phase of the deep learning model and the other 50 % (n=82) remain unseen for testing.

2.5.2 WMH challenge dataset

The training data comprises 60 scans, acquired with 3 different MRI scanners: 3T Philips Achieva, 3T SiemensTrioTim, and 3T GE Signa HDx. For testing, we use the WMH challenge test set, which includes 110 individuals scanned on five different machines: 3T Philips Achieva, 3T Siemens TrioTim, 3T GE Signa HDx, 3 T Philips Ingenuity and 1.5 T GE Signa HDxt. This allows for testing the model on different imaging protocols stemming from different MRI scanners.

2.6 Network architecture

The proposed architecture is a variant of the UNet (Ronneberger et al., 2015), called ResUNet (Diakogiannis et al., 2020), which integrates residual learning with the traditional U-Net structure to enhance image segmentation performance. The model takes FLAIR images as inputs and is trained to output both WMH and WM masks.

The architecture of the model features an encoder and a decoder path (Figure 2). The encoder reduces the spatial information while feature information is increased. It is composed of a series of convolutions layers, followed by the application of Rectified Linear Unit (ReLU) and max-pooling. A skip connection

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

bypasses these layers and directly adds the input of the block to its output, which helps to mitigate the "vanishing gradient" problem and improves the network's ability to learn complex features.

The decoder re-combines the feature and spatial information through deconvolutions and concatenations with high-resolution features from the encoding path. Dropout layers are added to prevent overfitting. The initial layers in the encoder start with 16 kernels and increase by a factor of 2 at each downsampling step. The decoder follows a similar pattern in reverse, decreasing the number of kernels as the spatial resolution is restored.

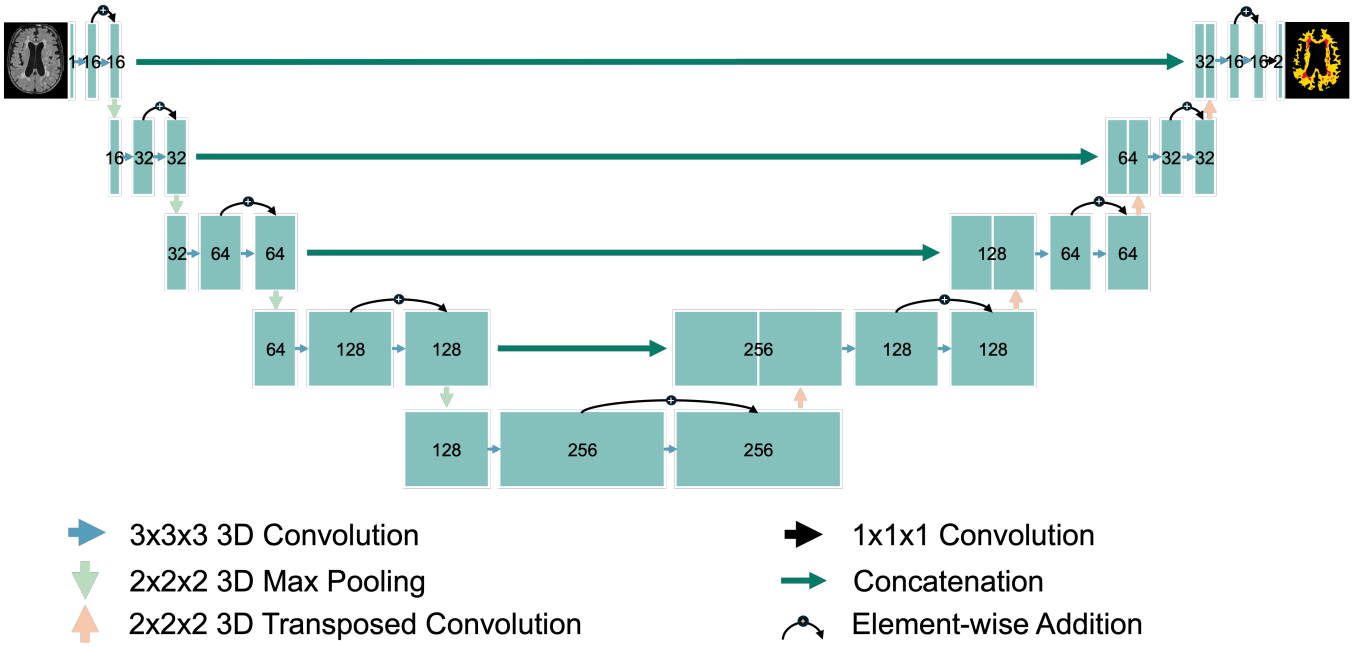

Figure 2: ResUNet architecture.

2.7 Implementation information

The network is implemented in Python3 using the PyTorch framework (Paszke et al., 2019). The model is trained on one NVIDIA GeForce RTX 2080 Ti GPU with 11GB RAM. The optimization process is performed with an Adam optimizer, batch size = 2, learning rate = $1e^{-4}$ for 300 epochs. 20 % of the training set, randomly selected, is used as the validation set. The model weights are saved each time the validation metric increases. Early stopping is used if the validation loss did not decrease for 20 epochs. The model is optimized using a variant of the Dice loss (Milletari et al., 2016) as follows :

$$L_{P-DSC}(y_{true}, y_{pred}) = 1 - \frac{2y_{true} \cdot y_{pred} + \epsilon}{y_{true}^{\gamma} + y_{pred}^{\gamma} + \epsilon} \tag{1}$$

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

with $\gamma = 2$.The power γ increases the sensitivity to difference in prediction. It amplifies the differences, therefore the loss largely penalizes misclassification.

2.8 Performance evaluation

The model performance is evaluated using the following metrics.

- The Dice Similarity coefficient (DSC) is used to evaluate the spatial overlap between the manually labeled WMH and the WMH mask generated using the algorithm.

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{2}$$

  where TP, FP, FN correspond respectively to the number of true positives, false positives and false negatives.

- The Recall evaluates the sensitivity of the model, the probability that an actual positive (the voxel belongs to WMH category) will be predicted as positive.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

- The absolute volume difference (AVD) between manual WMH segmentation and algorithm output.

$$AVD = \frac{\mid V_{true} - V_{pred} \mid}{V_{true}} \tag{4}$$

  If the manual segmentation contains no lesions, the AVD is set to 0 if the predicted mask is empty and 1 if the reverse is true.

- The Hausdorff distance (95_{th} percentile) computes the maximum distance between manually segmented lesion voxels and the nearest voxel for predicted lesions.

$$\begin{aligned} H(S_{true}, S_{pred}) &= \max(h(S_{true}, S_{pred}), h(S_{pred}, S_{true})) \\ h(S_{true}, S_{pred}) &= \underset{s_1 \in S_{true}}{95^{th}} \min_{s_2 \in S_{pred}} \|s_1 - s_2\| \end{aligned} \tag{5}$$

  where S_{true} is the set of "1" voxels in the manual segmentation mask and S_{pred} the set of "1" voxels in the predicted mask. Hausdorff distance is defined only if the manual segmentation and the predicted mask contain lesions.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

2.9 Performance comparison with other WMH segmentation tools

The performance of WHITE-Net is compared with other automatic segmentation tools widely used in the community. The selected tools for this comparison are the following :

- LPA: Lesion Prediction Algorithm (Schmidt, 2017) from the lesion segmentation toolbox (LST) on SPM12, based on logistic regression.
- LGA: Lesion Growth Algorithm (Schmidt et al., 2012) from the lesion segmentation toolbox (LST) on SPM12. This tool is based on probabilistic modeling and region growing.
- CAT12 (Gaser et al., 2022): The Computational Anatomy Toolbox on SPM12. It is based on tissue probability maps and intensity. Isolated GM clusters within WM and voxels around ventricles with a GM intensity but a high WM probability are selected as WMH.
- BIANCA (Griffanti et al., 2016): The Brain Intensity AbNormality Classification Algorithm in FSL, performing k-NN.
- PGS (Park et al., 2021): Deep Learning tool from the WMH challenge (H. J. Kuijf et al., 2019). It is the winner of this challenge. It uses ensemble U-Net with multi-scale highlighting foregrounds.
- SYSU_MEDIA2 (Li et al., 2018): Deep Learning tool from the WMH challenge (H. J. Kuijf et al., 2019), based on a two-channels U-Net.

Each of the aforementioned algorithms is applied to the BrainLaus test set. BIANCA is evaluated using two different input configurations: first, using only FLAIR images and second, using both FLAIR and T1w images. The training phase of BIANCA uses the same BrainLaus training set as WHITE-Net. BIANCA faces challenges in determining a suitable threshold that can accurately segment both small and large lesions. After testing various thresholds, a value of 0.1 is selected as the optimal threshold for segmentation.

3 Results

3.1 Performance of WHITE-Net on BrainLaus test set

Across all participants in the test set, the overall Dice score for the WMH segmentation is $DSC = 0.55 \pm 0.32$. The Dice score for the WM segmentation is 0.90 ± 0.02. When considering only participants

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

with high WMH severity (Fazekas 4-6), the WMH Dice increases to 0.76 ± 0.11. Examples of predicted WMH and WM masks are presented in Figure 3.

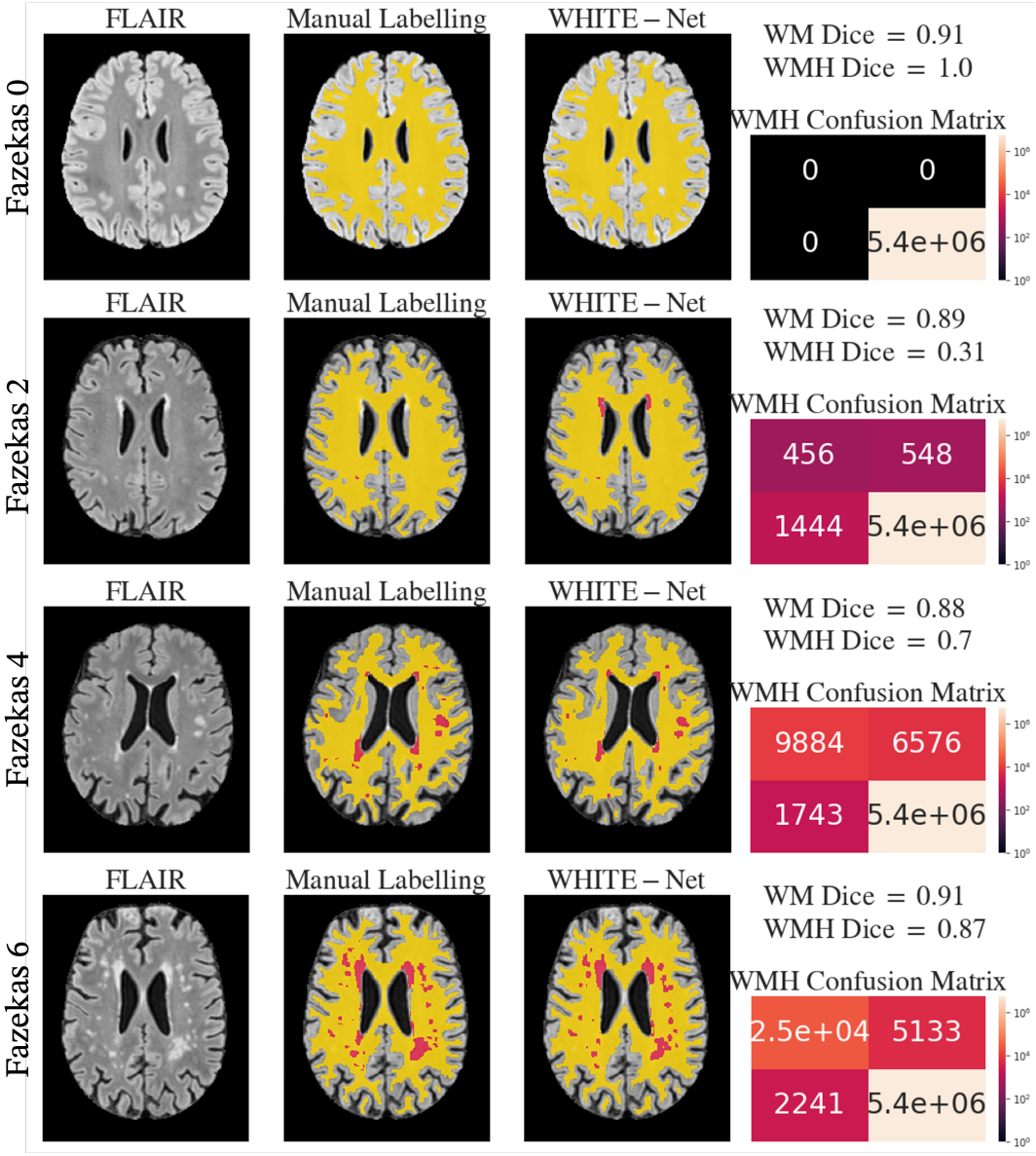

Figure 3: Examples of WHITE-Net outputs on four different participants with different lesion extent. FLAIR contrasts, manual labeling and WHITE-Net outputs can be seen from left to right. In the confusion matrix, false negatives are on the top right and false positives on the bottom left. Red color represents the WMHs, while the yellow color shows the rest of the WM voxels.

We also trained a multi-contrast model that includes both FLAIR and T1w MRIs, resulting in a WMH DSC of 0.54 ± 0.32. Given the similar performance of single and multiple MR contrast models, we use the single MR contrast model in the subsequent analyses.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

3.1.1 Performance across lesion severity

We assess the algorithm's performance in relation to lesion severity using the test set from the Brainlaus dataset. For these participants, the lesion extent is described by the Fazekas rating (Figure 4). WHITE-Net performs better as the severity and spatial extent of the lesion load increases.

3.1.2 Performance across white matter compartments

The performance of WHITE-Net is analyzed also in the three WM compartments - DWM, PVWM and SWM (Figure 5). DSC scores are higher in the DWM while PVWM's lesion Dice demonstrated larger interquartile range.

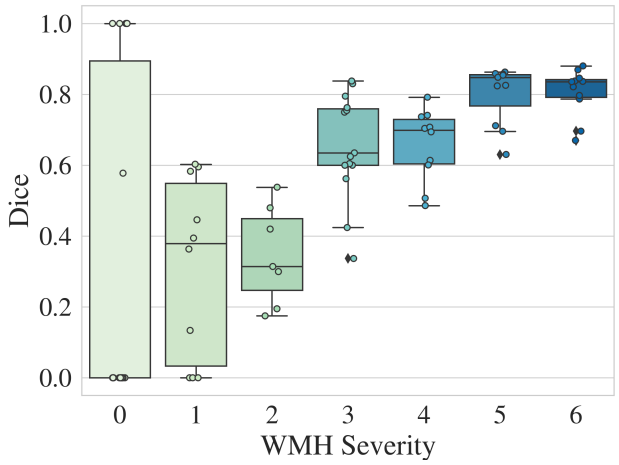

Figure 4: Dice Score (DSC) per Fazekas category, ranging from 0 (no WMH) to 6 (extensive PVWM and DWM lesions).

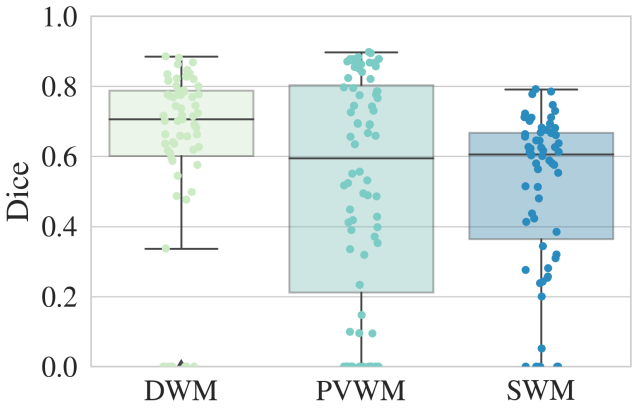

Figure 5: Dice score for different white matter compartments. DWM corresponds to Deep White Matter, PVWM to Periventricular WhiteMatter, and SWM to Superficial White Matter.

3.2 Comparison with other algorithms

3.2.1 Performance across entire dataset

The Dice index of the performances of each WMH detection algorithm on the BrainLaus test set is calculated based on the manual WMH labels and the algorithms' WMH detection results (Table 2). The LGA algorithm achieves the lowest absolute volume difference, while our model outperforms all other models across the other metrics. Despite multiple attempts, the deep learning algorithm SYSU_MEDIA fails to perform in the BrainLaus dataset.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

The execution time is computed for 1 subject with Intel(R) Core(TM) i9-10920X CPU @ 3.50GHz.WHITE-Net and is significantly faster compared to all other methods, with execution times ranging from 5 to over 200 times.

Table 2: Performance of each algorithm on the BrainLaus test set. Mean ± STD of the DSC are reported. Bold results outline the best performing algorithms.

Algorithm	Input	DSC	Recall	AVD	H_{95}	Execution Time
WHITE-Net	FLAIR	$\mathbf{0.55 \pm 0.32}$	$\mathbf{0.57 \pm 0.32}$	0.74 ± 1.41	$\mathbf{19 \pm 25}$	**11.434 s**
LST LPA	FLAIR	0.33 ± 0.28	0.26 ± 0.25	0.78 ± 1.39	27 ± 18	2.25 min
LST LGA	FLAIR+T1w	0.48 ± 0.33	0.41 ± 0.32	$\mathbf{0.71 \pm 2.12}$	27 ± 19	5.05 min
CAT12	T1w	0.30 ± 0.24	0.29 ± 0.21	2.52 ± 8.46	27 ± 26	35.2 min
BIANCA	FLAIR	0.17 ± 0.18	0.15 ± 0.19	4.21 ± 13.14	47 ± 29	>40 min
BIANCA	FLAIR+T1w	0.25 ± 0.23	0.25 ± 0.23	2.73 ± 9.83	39 ± 27	>40 min
PGS	FLAIR+T1w	0.43 ± 0.29	0.44 ± 0.27	1.17 ± 2.25	23 ± 25	55.951 s
SYSU_MEDIA	FLAIR+T1w	0.17 ± 0.38	0.17 ± 0.38	0.83 ± 0.38	139 ± 26	2.78 min

3.2.2 Performance on severe lesions

The same performance metrics are assessed but only considering the more severe cases of WMH (Table 3), keeping only scans with a Fazekas score between 4 and 6. In these cases, WHITE-Net outperforms all other methods across all metrics.

Table 3: Performance of each algorithm on BrainLaus test set for severe cases only (Fazekas 4-6).

Algorithm	Input	DSC	Recall	AVD	H_{95}
WHITE-Net	FLAIR	$\mathbf{0.76 \pm 0.11}$	$\mathbf{0.75 \pm 0.12}$	$\mathbf{0.23 \pm 0.21}$	$\mathbf{5.96 \pm 5.34}$
LST LPA	FLAIR	0.47 ± 0.19	0.36 ± 0.19	0.57 ± 0.20	16.83 ± 5.93
LST LGA	FLAIR+T1w	0.54 ± 0.17	0.42 ± 0.17	0.46 ± 0.19	17.42 ± 5.49
CAT12	T1w	0.53 ± 0.15	0.44 ± 0.13	0.36 ± 0.18	11.31 ± 7.57
BIANCA	FLAIR	0.28 ± 0.19	0.21 ± 0.17	0.68 ± 0.23	32.97 ± 21.22
BIANCA	FLAIR+T1w	0.43 ± 0.20	0.36 ± 0.22	0.60 ± 0.23	25.02 ± 16.98
PGS	FLAIR+T1w	0.68 ± 0.11	0.59 ± 0.12	0.31 ± 0.14	7.38 ± 6.93
SYSU_MEDIA	FLAIR+T1w	0 ± 0	0 ± 0	1 ± 0	145.89 ± 22.69

The statistical analysis using a t-test comparing the Dice coefficients of each tool against the Dice coefficients of WHITE-Net confirms the significance of performance differences (Table 4). The tests are performed on the entire BrainLaus test set and also on the subset of MRIs with the most severe SVD (Fazekas score of 4-6).

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Table 4: Statistical comparison of WHITE-Net and other tools dice scores using t-tests.

Algorithm	Fazekas 0-6		Fazekas 4-6	
	t_stat	p_val	t_stat	p_val
LST LPA	-9.76	0.0	-11.21	0.0
LST LGA	-1.89	0.06	-11.13	0.0
CAT12	-9.24	0.0	-11.21	0.0
BIANCA FLAIR	-11.37	0.0	-11.80	0.0
BIANCA FLAIR+T1w	-9.18	0.0	-8.21	0.0
PGS	-4.20	0.0	-4.26	0.0
SYSU_MEDIA	-5.93	0.0	-40.13	0.0

3.2.3 Performance in absence of lesions

The number of false positives on scans classified as Fazekas 0 (without lesions) is studied. Compared to other WMH segmentation algorithms, WHITE-Net shows optimal performance on Fazekas 0 scans (Figure 6). SYSU_MEDIA has the lowest number of false positives due to its failure to detect WMHs in the BrainLaus dataset. Excluding SYSU_MEDIA, LGA and LPA have the best performance in the absence of lesions. WHITE-Net also has a much lower number of false positive predictions than BIANCA, CAT12 and PGS.

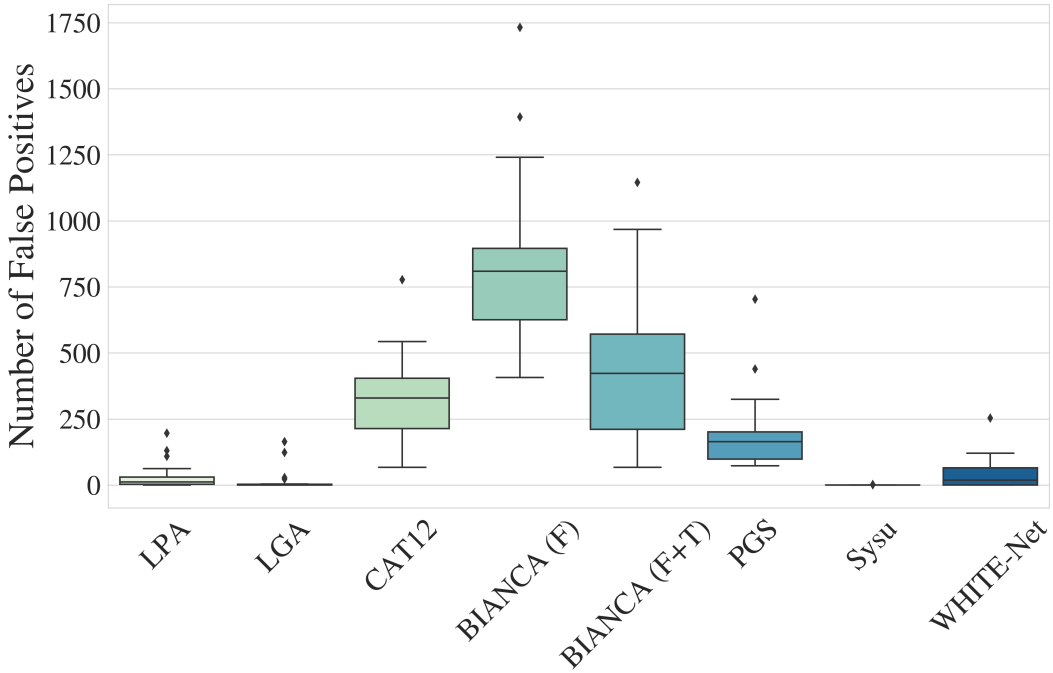

Figure 6: Number of False Positive voxels for Fazekas 0 scans using each segmentation tool. For BIANCA, F stands for FLAIR and F+T for FLAIR+T1w.

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

3.2.4 Performance according lesion size

The performance of each algorithm according to the lesion size shows significant differences in DSC scores (Figure 7). The PGS algorithm outperforms the other algorithms for reduced WMH lesion extent, followed by WHITE-Net. However, WHITE-Net becomes the most accurate with WHM lesion sizes of approximately 50 voxels or larger.

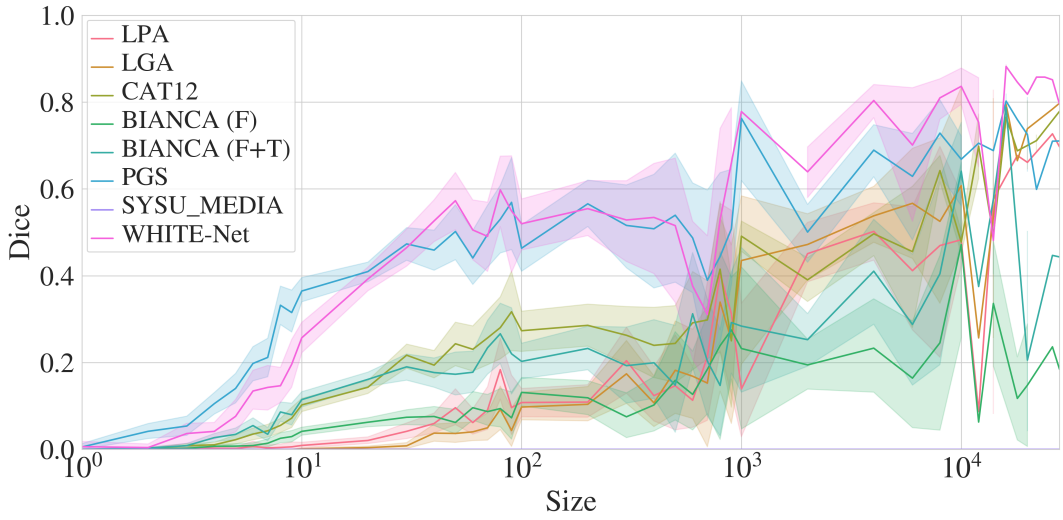

Figure 7: Dice score and 95 % confidence interval per lesion size for each segmentation tool. The size is equal to the number of voxels in a lesion.

3.3 Performance in the WMH challenge test set

We test the WHITE-Net model using the WMH challenge test set data. The WMH Dice is 0.72 ± 0.13, while the overall WM Dice is 0.83 ± 0.01. Figure 8 presents the Dice scores for each scanner type. The performance of WHITE-Net on WMH segmentation, are in the same range for all scanners.

3.4 Manual segmentation analysis

Finally, we ensure the test-retest accuracy of the manual labeling in 7 MRIs segmented twice by the same rater, one per Fazekas category (Table 5).

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

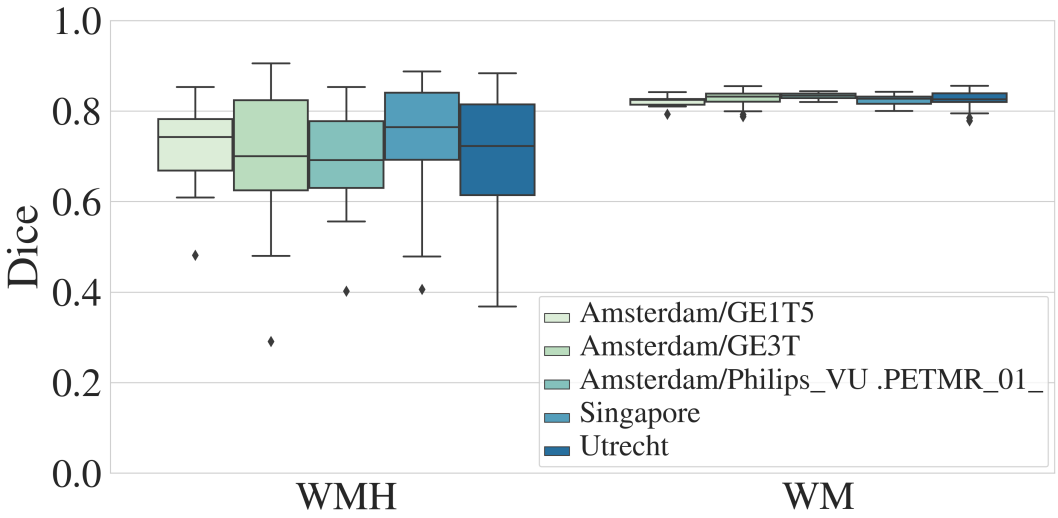

Figure 8: WMH and WM Dice scores for each scanner in the WMH challenge test set.

Table 5: Dice score between first and second manual labeling by the same rater.

Fazekas	0	1	2	3	4	5	6
DSC	0	0.57	0.41	0.76	0.76	0.79	0.83
Mean	0.59						

4 Discussion

We introduce WHITE-Net, a novel deep learning-based tool for WMH segmentation trained on MRI datasets representative of aging-associated disorders. These datasets are acquired from MRI machines of different vendors and operating at different field strengths. The results obtained with WHITE-Net demonstrate high accuracy and fast execution times. This open-source tool is user friendly, requiring no parameter adjustment. According to its accuracy and computational efficiency, WHITE-Net ranks among the top existing algorithms for automated WMH lesion detection in the non-clinical setting of computational anatomy studies.

4.1 Performance for WMH detection

WHITE-Net's segmentation accuracy is notably influenced by lesion size: larger lesions are segmented with higher precision, while small lesions (under 10 voxels) remain more challenging (Figure 7). This limitation is particularly noticeable in cases with low Fazekas scores, where small lesions are prevalent and harder to detect accurately (Figure 4, Table 3). Specifically, scans classified as Fazekas 0 (no lesions) are

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

problematic, as even a single small lesion detection error can drive the Dice coefficient to zero, leading to high sensitivity to false positives in these cases (Figure 6). Despite these challenges, WHITE-Net maintains a low false positive rate across all Fazekas levels, which makes it a reliable tool even for scans without visible WMHs.

WHITE-Net's performance also varies by white matter compartment, as indicated in Figure 5. In particular, the PVWM is challenging, with a wide interquartile range of Dice scores. This difficulty likely stems from bright areas around the ventricles (due to the orientation of the ventricles in the slice) that may be misinterpreted as lesions, especially when lesion load is low. This challenge also potentially contributes to the lower performance observed in less severe scans (Figure 4), where the low lesion load around ventricles can lead to segmentation difficulties. However, WHITE-Net can be a good approach for addressing this issue due to its high level of objectivity and consistency, unlike manual segmentation that can be highly subjective. Conversely, lesions in the DWM are segmented with higher accuracy, as they typically have a well-defined and bright appearance on FLAIR images. Superficial white matter (SWM) also presents some challenges due to its proximity to gray matter, where similar intensity can complicate segmentation.

In comparison to other segmentation tools, WHITE-Net generally outperforms them on most metrics. The exception is the absolute volume difference (AVD) metric, where LGA achieves better results in non-severe cases due to a lower false positive rate, which significantly affects AVD in scans without lesions. Table 4 illustrates that WHITE-Net's Dice coefficient differences with other tools are statistically significant ($p < 0.05$) in favor of WHITE-Net, except when compared to LGA, where the significance is observed only in severe cases. Other tools also struggle with accurately segmenting small lesions and with scans having low Fazekas scores. This limitation is common across WMH segmentation models due to the lower contrast in these cases and the bigger impact of a wrong prediction.

To provide context for the Dice score results, we also evaluate the performance of human raters in manual WMH segmentation. This is done by computing the intra-rater Dice score between two manual segmentations done by the same rater on the same scans (Table 5). The resulting Dice score of 0.59 falls within the same range as the Dice score achieved by WHITE-Net (0.55), indicating comparable performance between the model and gold standard manual segmentation reliability.

4.2 Performance across vendors and field strengths

WHITE-Net also demonstrates robust performance across various MR scanner types, including two new scanners, as highlighted in Figure 8, even with anisotropic voxel scans. Anisotropic voxel dimensions typically lead to spatial resolution differences between planes and could cause algorithms to segment

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

accurately the lesions in one plane but not in the other directions. However, WHITE-Net effectively handles these differences, as demonstrated in UMC Utrecht and NUHS Singapore data, where it accurately segments WMHs despite anisotropic voxel dimensions. WHITE-Net demonstrates also its applicability to different magnetic field strength. Indeed, lower field strength results in a reduced signal-to-noise ratio and contrast-to-noise ratio (Biswas et al., 2005), yet WHITE-Net demonstrates strong performance on 1.5T GE data. This versatility makes it suitable for studies using diverse datasets, supporting its reliability in heterogeneous clinical data.

4.3 Computational efficiency and accessibility

WHITE-Net demonstrates high computational efficiency, with an execution time of only 11.434 seconds (Table 2). This fast processing speed is especially valuable for large cohort studies and clinical environments, where fast data processing is essential. Additionally, WHITE-Net requires no coding skills or parameter adjustments, making it highly accessible to users without technical expertise.

Compared to other algorithms, WHITE-Net's computational efficiency and simplicity make it a user-friendly tool. The execution time for BIANCA in this study was defined by a lower bound because this tool required model training. Therefore, the time depends on the size of the training set. Moreover, several steps are needed before running BIANCA and were not taken into account in the execution time (CSF pve map, non-linear transformation warp file from standard space to structural image, transformation matrix from subject space to standard space).

4.4 Improvements and conclusion

While WHITE-Net demonstrates excellent performance, it faces challenges in accurately detecting very small false positive lesions, as shown in Figure 7. Improvements could be made by adopting strategies like ensemble models or multi-scale approaches. An ensemble model, such as PGS, could leverage majority voting across multiple trained models, while a multi-scale model could process data at different resolutions to better detect features of varying sizes. Future work could also explore generalizing WHITE-Net to other imaging contrasts. For example, in the BrainLaus dataset, participants are not scanned with T1w and FLAIR sequences at each follow-up visit. Expanding WHITE-Net to work with alternative contrasts like Multi-Parametric Mapping (Weiskopf et al., 2013) would allow for longitudinal WMH studies even when standard imaging is unavailable.

To conclude, WHITE-Net is an efficient and user-friendly segmentation algorithm designed for fast

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

processing and ease of use. It processes a single image in a few seconds and the tool does not require any coding expertise. The instructions are clearly detailed on its GitHub repository (https://github.com/cathalacamille/WHITE-Net). Furthermore, WHITE-Net offers flexibility with two types of input options: FLAIR or FLAIR+T1w MRI, making it adaptable for a wide range of research and clinical applications.

WHITE-Net also holds significant potential for advancing lesion research, particularly in the context of aging and neurodegenerative diseases. Accurate lesion detection in healthy individuals is crucial for tracking lesion evolution and understanding the onset and progression of lesions. WHITE-Net can reliably segment small lesions and scans with low lesion load, making it a valuable tool for longitudinal studies aimed at early lesion detection. This capability is essential for advancing our understanding of lesion development and progression and developing personalized treatments tailored to individual patients.

5 Data and Code Availability

Code for the segmentation tool described in this article, WHITE-Net, is available on GitHub (https://github.com/cathalacamille/WHITE-Net).

Data of CoLaus|PsyCoLaus study used in this article cannot be fully shared as they contain potentially sensitive personal information on participants. According to the Ethics Committee for Research of the Canton of Vaud, sharing these data would be a violation of the Swiss legislation with respect to privacy protection. However, coded individual-level data that do not allow researchers to identify participants are available upon request to researchers who meet the criteria for data sharing of the CoLaus|PsyCoLaus Datacenter (CHUV, Lausanne, Switzerland). Any researcher affiliated to a public or private research institution who complies with the CoLaus|PsyCoLaus standards can submit a research application to research.colaus@chuv.ch or research.psycolaus@chuv.ch. Proposals will be evaluated by the Scientific Committee (SC) of the CoLaus|PsyCoLaus studies. Detailed instructions for gaining access to the CoLaus|PsyCoLaus data used in this study are available at www.colaus-psycolaus.ch/professionals/how-to-collaborate/.

Data from WMH challenge are freely available at https://doi.org/10.34894/aecrsd (H. Kuijf et al., 2022).

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

6 Author Contribution

Data preprocessing was conducted by A.B. and C.C.. Neural network design, model optimization and results analysis were performed by C.C.. C.C. wrote the manuscript with inputs from all authors. A.B., B.D. and J-P.T. supervised the project.

7 Declaration of Competing Interests

The authors declare that they have no competing interests.

8 Acknowledgments

Funding is supported by the Swiss National Science Foundation (project grants Nr. 213595, 32003B_135679, 32003B_159780, 324730_192755 and CRSK-3_190185), ERA_NET iSEE and BrainTree projects. LREN is very grateful to the Roger De Spoelberch and Partridge Foundations for their generous financial support.

References

Aljondi, R., Szoeke, C., Steward, C., Gorelik, A., & Desmond, P. (2018). The effect of midlife cardiovascular risk factors on white matter hyperintensity volume and cognition two decades later in normal ageing women. *Brain Imaging and Behavior*, *14*(1), 51–61. https://doi.org/10.1007/s11682-018-9970-5

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018

Billot, B., Greve, D. N., Puonti, O., Thielscher, A., Van Leemput, K., Fischl, B., Dalca, A. V., & Iglesias, J. E. (2023). Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining. *Medical Image Analysis*, *86*, 102789. https://doi.org/10.1016/j.media.2023.102789

Biswas, J., Nelson, C. B., Runge, V. M., Wintersperger, B. J., Baumann, S. S., Jackson, C. B., & Patel, T. (2005). Brain tumor enhancement in magnetic resonance imaging: Comparison of signal-to-noise ratio (snr) and contrast-to-noise ratio (cnr) at 1.5 versus 3 tesla. *Investigative Radiology*, *40*(12), 792–797. https://doi.org/10.1097/01.rli.0000187609.78338.dc

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Boomsma, J. M. F., Exalto, L. G., Barkhof, F., van den Berg, E., de Bresser, J., Heinen, R., Koek, H. L., Prins, N. D., Scheltens, P., Weinstein, H. C., van der Flier, W. M., & Biessels, G. J. (2017). Vascular cognitive impairment in a memory clinic population: Rationale and design of the "utrecht-amsterdam clinical features and prognosis in vascular cognitive impairment" (trace-vci) study. *JMIR Research Protocols*, *6*(4), e60. https://doi.org/10.2196/resprot.6864

Coenen, M., Kuijf, H. J., Huenges Wajer, I. M. C., Duering, M., Wolters, F. J., Fletcher, E. F., Maillard, P. M., Barkhof, F., Barnes, J., Benke, T., Boomsma, J. M. F., Chen, C. P. L. H., Dal-Bianco, P., Dewenter, A., Enzinger, C., Ewers, M., Exalto, L. G., Franzmeier, N., Groeneveld, O., ... Biesbroek, J. M. (2022). Strategic white matter hyperintensity locations for cognitive impairment: A multicenter lesion-symptom mapping study in 3525 memory clinic patients. *Alzheimer's amp; Dementia*, *19*(6), 2420–2432. https://doi.org/10.1002/alz.12827

Commowick, O., Istace, A., Kain, M., Laurent, B., Leray, F., Simon, M., Pop, S. C., Girard, P., Améli, R., Ferré, J.-C., Kerbrat, A., Tourdias, T., Cervenansky, F., Glatard, T., Beaumont, J., Doyle, S., Forbes, F., Knight, J., Khademi, A., ... Barillot, C. (2018). Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-31911-7

de Bresser, J., Kuijf, H. J., Zaanen, K., Viergever, M. A., Hendrikse, J., Biessels, G. J., Algra, A., van den Berg, E., Bouvy, W., Brundel, M., Heringa, S., Kappelle, L. J., Leemans, A., Luijten, P. R., Mali, W. P. T. M., Rutten, G. E. H. M., Vincken, K. L., & Zwanenburg, J. (2018). White matter hyperintensity shape and location feature analysis on brain mri; proof of principle study in patients with diabetes. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-20084-y

Diakogiannis, F. I., Waldner, F., Caccetta, P., & Wu, C. (2020). Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *162*, 94–114. https://doi.org/10.1016/j.isprsjprs.2020.01.013

Fazekas, F., Barkhof, F., Filippi, M., Grossman, R., Li, D. B., McDonald, W., McFarland, H., Paty, D., Simon, J., Wolinsky, J., & Miller, D. (1999). The contribution of magnetic resonance imaging to the diagnosis of multiple sclerosis. *Neurology*, *53*(3), 448–448. https://doi.org/10.1212/wnl.53.3.448

Fazekas, F., Chawluk, J., Alavi, A., Hurtig, H., & Zimmerman, R. (1987). Mr signal abnormalities at 1.5 t in alzheimer's dementia and normal aging. *American Journal of Roentgenology*, *149*(2), 351–356. https://doi.org/10.2214/ajr.149.2.351

Filippi, M., Preziosa, P., Banwell, B. L., Barkhof, F., Ciccarelli, O., De Stefano, N., Geurts, J. J. G., Paul, F., Reich, D. S., Toosy, A. T., Traboulsee, A., Wattjes, M. P., Yousry, T. A., Gass, A., Lubetzki, C., Weinshenker, B. G., & Rocca, M. A. (2019). Assessment of lesions on magnetic resonance imaging in multiple sclerosis: Practical guidelines. *Brain*, *142*(7), 1858–1875. https://doi.org/10.1093/brain/awz144

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Firmann, M., Mayor, V., Vidal, P. M., Bochud, M., Pécoud, A., Hayoz, D., Paccaud, F., Preisig, M., Song, K. S., Yuan, X., Danoff, T. M., Stirnadel, H. A., Waterworth, D., Mooser, V., Waeber, G., & Vollenweider, P. (2008). The colaus study: A population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovascular Disorders*, *8*(1). https://doi.org/10.1186/1471-2261-8-6

Fuhrmann, D., Nesbitt, D., Shafto, M., Rowe, J. B., Price, D., Gadie, A., Tyler, L. K., Brayne, C., Bullmore, E. T., Calder, A. C., Cusack, R., Dalgleish, T., Duncan, J., Henson, R. N., Matthews, F. E., Marslen-Wilson, W. D., Rowe, J. B., Shafto, M. A., Campbell, K., ... Kievit, R. A. (2019). Strong and specific associations between cardiovascular risk factors and white matter micro- and macrostructure in healthy aging. *Neurobiology of Aging*, *74*, 46–55. https://doi.org/10.1016/j.neurobiolaging.2018.10.005

Gaser, C., Dahnke, R., Thompson, P. M., Kurth, F., & Luders, E. (2022). Cat – a computational anatomy toolbox for the analysis of structural mri data. https://doi.org/10.1101/2022.06.11.495736

Ghaznawi, R., Geerlings, M. I., Jaarsma-Coes, M. G., Zwartbol, M. H., Kuijf, H. J., van der Graaf, Y., Witkamp, T. D., Hendrikse, J., & de Bresser, J. (2018). The association between lacunes and white matter hyperintensity features on mri: The smart-mr study. *Journal of Cerebral Blood Flow amp; Metabolism*, *39*(12), 2486–2496. https://doi.org/10.1177/0271678x18800463

Griffanti, L., Zamboni, G., Khan, A., Li, L., Bonifacio, G., Sundaresan, V., Schulz, U. G., Kuker, W., Battaglini, M., Rothwell, P. M., & Jenkinson, M. (2016). Bianca (brain intensity abnormality classification algorithm): A new tool for automated segmentation of white matter hyperintensities. *NeuroImage*, *141*, 191–205. https://doi.org/10.1016/j.neuroimage.2016.07.018

Grimaud, J., Lai, M., Thorpe, J., Adeleine, P., Wang, L., Barker, G., Plummer, D., Tofts, P., McDonald, W., & Miller, D. (1996). Quantification of mri lesion load in multiple sclerosis: A comparison of three computer-assisted techniques. *Magnetic Resonance Imaging*, *14*(5), 495–505. https://doi.org/10.1016/0730-725x(96)00018-5

Jiménez-Balado, J., Corlier, F., Habeck, C., Stern, Y., & Eich, T. (2022). Effects of white matter hyperintensities distribution and clustering on late-life cognitive impairment. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-06019-8

Karim, H. T., Andreescu, C., MacCloud, R. L., Butters, M. A., Reynolds, C. F., Aizenstein, H. J., & Tudorascu, D. L. (2016). The effects of white matter disease on the accuracy of automated segmentation. *Psychiatry Research: Neuroimaging*, *253*, 7–14. https://doi.org/10.1016/j.pscychresns.2016.05.003

Klein, S., Staring, M., Murphy, K., Viergever, M., & Pluim, J. (2010). Elastix: A toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, *29*(1), 196–205. https://doi.org/10.1109/tmi.2009.2035616

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Kruit, M., van Buchem, M., Launer, L., Terwindt, G., & Ferrari, M. (2009). Migraine is associated with an increased risk of deep white matter lesions, subclinical posterior circulation infarcts and brain iron accumulation: The population-based mri camera study. *Cephalalgia*, *30*(2), 129–136. https://doi.org/10.1111/j.1468-2982.2009.01904.x

Kuijf, H., Biesbroek, M., de Bresser, J., Heinen, R., Chen, C., van der Flier, W., Barkhof, Viergever, M., & Biessels, G. J. (2022). *Data of the White Matter Hyperintensity (WMH) Segmentation Challenge*. https://doi.org/10.34894/AECRSD

Kuijf, H. J., Casamitjana, A., Collins, D. L., Dadar, M., Georgiou, A., Ghafoorian, M., Jin, D., Khademi, A., Knight, J., Li, H., Llado, X., Biesbroek, J. M., Luna, M., Mahmood, Q., McKinley, R., Mehrtash, A., Ourselin, S., Park, B.-Y., Park, H., ... Cardoso, M. J. (2019). 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. https://doi.org/10.1109/tmi.2019.2905770

Kuller, L. H., Longstreth, W., Arnold, A. M., Bernick, C., Bryan, R. N., & Beauchamp, N. J. (2004). White matter hyperintensity on cranial magnetic resonance imaging: A predictor of stroke. *Stroke*, *35*(8), 1821–1825. https://doi.org/10.1161/01.str.0000132193.35955.69

Lee, S., Viqar, F., Zimmerman, M. E., Narkhede, A., Tosto, G., Benzinger, T. L., Marcus, D. S., Fagan, A. M., Goate, A., Fox, N. C., Cairns, N. J., Holtzman, D. M., Buckles, V., Ghetti, B., McDade, E., Martins, R. N., Saykin, A. J., Masters, C. L., Ringman, J. M., ... Brickman, A. M. (2016). White matter hyperintensities are a core feature of alzheimer's disease: Evidence from the dominantly inherited alzheimer network. *Annals of Neurology*, *79*(6), 929–939. https://doi.org/10.1002/ana.24647

Levy-Cooperman, N., Ramirez, J., Lobaugh, N. J., & Black, S. E. (2008). Misclassified tissue volumes in alzheimer disease patients with white matter hyperintensities: Importance of lesion segmentation procedures for volumetric analysis. *Stroke*, *39*(4), 1134–1141. https://doi.org/10.1161/strokeaha.107.498196

Li, H., Jiang, G., Zhang, J., Wang, R., Wang, Z., Zheng, W.-S., & Menze, B. (2018). Fully convolutional network ensembles for white matter hyperintensities segmentation in mr images. *NeuroImage*, *183*, 650–665. https://doi.org/10.1016/j.neuroimage.2018.07.005

Manjón, J. V., Coupé, P., Martí-Bonmatí, L., Collins, D. L., & Robles, M. (2009). Adaptive non-local means denoising of mr images with spatially varying noise levels. *Journal of Magnetic Resonance Imaging*, *31*(1), 192–203. https://doi.org/10.1002/jmri.22003

Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation. *2016 Fourth International Conference on 3D Vision (3DV)*. https://doi.org/10.1109/3dv.2016.79

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Mortamais, M., Artero, S., & Ritchie, K. (2013). Cerebral white matter hyperintensities in the prediction of cognitive decline and incident dementia. *International Review of Psychiatry*, *25*(6), 686–698. https://doi.org/10.3109/09540261.2013.838151

Park, G., Hong, J., Duffy, B. A., Lee, J.-M., & Kim, H. (2021, August). *White matter hyperintensities segmentation using the ensemble u-net with multi-scale highlighting foregrounds*. https://doi.org/10.1016/j.neuroimage.2021.118140

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. https://doi.org/10.48550/ARXIV.1912.01703

Preisig, M., Waeber, G., Vollenweider, P., Bovet, P., Rothen, S., Vandeleur, C., Guex, P., Middleton, L., Waterworth, D., Mooser, V., Tozzi, F., & Muglia, P. (2009). The psycolaus study: Methodology and characteristics of the sample of a population-based survey on psychiatric disorders and their association with genetic and cardiovascular risk factors. *BMC Psychiatry*, *9*(1). https://doi.org/10.1186/1471-244x-9-9

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention – miccai 2015* (pp. 234–241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Rovira, À., & León, A. (2008). Mr in the diagnosis and monitoring of multiple sclerosis: An overview. *European Journal of Radiology*, *67*(3), 409–414. https://doi.org/10.1016/j.ejrad.2008.02.044

Schmidt, P. (2017, January). *Bayesian inference for structured additive regression models for large-scale problems with applications to medical imaging*. Ludwig-Maximilians-Universität München. http://nbn-resolving.de/urn:nbn:de:bvb:19-203731

Schmidt, P., Gaser, C., Arsic, M., Buck, D., Förschler, A., Berthele, A., Hoshi, M., Ilg, R., Schmid, V. J., Zimmer, C., Hemmer, B., & Mühlau, M. (2012). An automated tool for detection of flair-hyperintense white-matter lesions in multiple sclerosis. *NeuroImage*, *59*(4), 3774–3783. https://doi.org/10.1016/j.neuroimage.2011.11.032

Seneviratne, U., Chong, W., & Billimoria, P. (2013). Brain white matter hyperintensities in migraine: Clinical and radiological correlates. *Clinical Neurology and Neurosurgery*, *115*(7), 1040–1043. https://doi.org/10.1016/j.clineuro.2012.10.033

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). Uk biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

medRxiv preprint doi: https://doi.org/10.1101/2025.01.09.25320242; this version posted January 9, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Trofimova, O., Latypova, A., DiDomenicantonio, G., Lutti, A., de Lange, A.-M. G., Kliegel, M., Stringhini, S., Marques-Vidal, P., Vaucher, J., Vollenweider, P., Strippoli, M.-P. F., Preisig, M., Kherif, F., & Draganski, B. (2023). Topography of associations between cardiovascular risk factors and myelin loss in the ageing human brain. *Communications Biology*, *6*(1). https://doi.org/10.1038/s42003-023-04741-1

van Veluw, S. J., Hilal, S., Kuijf, H. J., Ikram, M. K., Xin, X., Yeow, T. B., Venketasubramanian, N., Biessels, G. J., & Chen, C. (2015). Cortical microinfarcts on 3t mri: Clinical correlates in memory-clinic patients. *Alzheimer's amp; Dementia*, *11*(12), 1500–1509. https://doi.org/10.1016/j.jalz.2014.12.010

Weiskopf, N., Suckling, J., Williams, G., Correia, M. M., Inkster, B., Tait, R., Ooi, C., Bullmore, E. T., & Lutti, A. (2013). Quantitative multi-parameter mapping of r1, pd*, mt, and r2* at 3t: A multi-center validation. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00095

Wood, M. L., & Henkelman, R. M. (1985). Mr image artifacts from periodic motion. *Medical Physics*, *12*(2), 143–151. https://doi.org/10.1118/1.595782