



---

# Automated detection of progression of white matter hyperintensities in cerebral small vessel disease using machine learning

---

MSC THESIS ARTIFICIAL INTELLIGENCE

*Daily supervisor:*  
Gemma Solé Guardia <sup>1</sup>

*2nd Supervisor:*  
Dr. ir. Geert Litjens <sup>2</sup>

*3rd Supervisor:*  
Prof. dr. Amanda Kiliaan <sup>1</sup>

*Internal supervisor:*  
Dr. Olympia Colizoli <sup>3</sup>

*Second reader:*  
Prof. dr. Tom Heskes <sup>4</sup>

*Author:*

Matthijs LUIJTEN <sup>1, 2</sup>  
s4787757

<sup>1</sup> Department of Medical Imaging, Anatomy

<sup>2</sup> Department of Computational Pathology

<sup>3</sup> AI Department at RU, Donders Centre for Cognition

<sup>4</sup> Data science Department at RU

## Abstract

White matter hyperintensities (WMH) are often seen in the brain of healthy elderly people. However, WMH are often used as biomarkers for prognosis or disease progression of cerebral small vessel disease (CSVD). WMH, seen as hyperintense on fluid-attenuated inversion recovery (FLAIR)- and T2-weighted MRI sequences, are amongst the foremost recurrent CSVD features observed on MRI scans and represent different degrees of axonal loss, demyelination, and gliosis. This research paper investigates the progression of WMH and NAWM in human brains through in vivo and post-mortem MRI segmentation and clustering of different regions. The proposed method achieves high segmentation performance with Dice coefficients of 0.76 for in vivo (compared to 0.805 for an independent human observer) and up to 0.94 for post-mortem MRI data. Clustering experiments successfully distinguish different levels of severity for WMH and NAWM using FLAIR intensity. Additionally, the study provides new insights into the mechanisms of WMH and their relationship with myelin loss. These findings have important implications for the development of new therapies to slow or halt the progression of WMH and related disorders. Future research should explore the relationship between WMH and other markers of white matter damage, such as axonal injury and inflammation, and perform larger-scale studies to further understand the pathogenesis of WMH.

# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>  | <b>3</b>  |
| 1.1      | Theoretical background . . . . .   | 3         |
| 1.2      | Aim of the project . . . . .   | 4         |
| 1.3      | Impact of the research . . . . .   | 4         |
| <b>2</b> | <b>Materials and methods</b>   | <b>5</b>  |
| 2.1      | In vivo MRI . . . . .  | 5         |
| 2.2      | Post-mortem MRI . . . . .  | 5         |
| 2.3      | (Immuno-)histochemistry stainings . . . . .  | 6         |
| 2.3.1    | MRI-pathology co-registration . . . . .  | 6         |
| 2.4      | Preprocessing . . . . .  | 7         |
| 2.4.1    | In vivo MRI . . . . .  | 7         |
| 2.4.2    | Post-mortem MRI . . . . .  | 7         |
| 2.4.3    | IHC stainings . . . . .  | 8         |
| 2.5      | Data augmentation . . . . .  | 8         |
| 2.5.1    | In vivo MRI . . . . .  | 8         |
| 2.5.2    | Post-mortem MRI . . . . .  | 8         |
| 2.6      | Fully Convolutional Network (FCN) U-Nets . . . . .   | 8         |
| 2.6.1    | In vivo MRI . . . . .  | 9         |
| 2.6.2    | Post-mortem MRI . . . . .  | 9         |
| 2.6.3    | (Categorical) Dice loss . . . . .  | 9         |
| 2.7      | Ensemble networks . . . . .  | 10        |
| 2.8      | Post Processing . . . . .  | 11        |
| 2.8.1    | In vivo MRI . . . . .  | 11        |
| 2.8.2    | Post-mortem MRI . . . . .  | 11        |
| 2.9      | Transition zones . . . . .   | 11        |
| 2.10     | MRI-pathology correlation . . . . .  | 11        |
| <b>3</b> | <b>Results</b>   | <b>11</b> |
| 3.1      | In vivo segmentation . . . . .   | 11        |
| 3.2      | Post-mortem segmentation . . . . .   | 13        |
| 3.3      | Transition zones . . . . .   | 13        |
| 3.4      | MRI-pathology correlation . . . . .  | 14        |
| <b>4</b> | <b>Discussion</b>  | <b>14</b> |
| 4.1      | Limitations of Dice score in in vivo MRI segmentation model and the need for visual inspection . . . . . | 15        |
| 4.2      | Applying in vivo method to post-mortem data . . . . .  | 16        |
| 4.3      | Different approaches to multi-class segmentation . . . . .   | 16        |
| 4.4      | Transition zones . . . . .   | 17        |
| <b>5</b> | <b>Conclusions</b>   | <b>17</b> |

# 1 Introduction

## 1.1 Theoretical background

Cerebral small vessel disease (CSVD) is a disorder of the cerebral small vasculature that leads to white matter hyperintensities (WMH), among several other MRI indicators, such as lacunes, enlarged perivascular spaces, cerebral microbleeds, recent small subcortical infarcts and brain atrophy [1, 2]. CSVD, and thus some extent of WMH, are commonly seen in elderly subjects (60 years or older) and patients with several neurological and vascular disorders [3]. CSVD causes about a fifth of all stroke cases worldwide and is the major vascular cause of cognitive decline and dementia [4, 5]. While WMH can vary greatly in their level of clinical significance, it is known that those with moderate WMH will most likely develop severe WMH, which hold a strong association with stroke and cognitive impairment [6–8]. Studying the transition zones from WMH to normal-appearing white matter (NAWM), areas at risk for WMH, can help understand the main contributors to increased WMH burden and its clinical outcome.

WMH, seen as hyperintense (bright) on fluid-attenuated inversion recovery (FLAIR) and T2-weighted MRI sequences, are amongst the foremost recurrent features of CSVD observed on MRI scans [1, 9, 10]. These white matter lesions represent different degrees of axonal loss, demyelination, and gliosis in the brain, which are thought to be caused by a variety of pathological processes, such as inflammation, ischemia, and oxidative stress [11, 12]. WMH have been found to have a strong correlation with common cardiovascular risk factors, specifically hypertension [13–16].

WMH appear bright on FLAIR images and dark on T1-weighted images because these sequences are designed to highlight different tissue characteristics in the brain, specifically the water content of the tissue. FLAIR images use an inversion pulse and a long inversion time to suppress the signal from cerebrospinal fluid (CSF) and reduce the signal from tissues with a long T1 relaxation time, such as white matter [17], which results in an image where WMH appear bright and the surrounding NAWM appears dark [18]. On the other hand, WMH can appear as isointense or hypointense (although not as hypointense as cerebrospinal fluid) on T1-weighted sequences, depending on the sequence parameters and severity of pathological change [1]. If these images are properly aligned, the high contrast between WMH and NAWM on these sequences makes them well suited for the automatic segmentation of WMH using techniques such as deep learning [17, 18], which can learn to identify and segment WMH based on the visual features present in the images. T1 and FLAIR images are often used together for the automatic segmentation of WMH in MRI scans because these sequences provide high contrast between WMH and the surrounding NAWM [17–20].

While there is evidence linking WMH to gliosis, myelin

loss, and axonal loss [21], the specific mechanisms and pathways involved are still not well understood. It is unclear whether these changes are exclusively related to myelin loss, as WMH can also represent edema [21, 22]. Further research is needed to elucidate the complex and heterogeneous processes underlying WMH development and progression. This uncertainty makes it difficult to fully understand the mechanisms behind WMH and their progression. This is partly why investigating the transition from lesion core to normal tissue and combining this with pathology can help elucidate the main mechanisms behind these changes observed in MRI. Despite the fact that WMH hold a strong association with an increased risk of long-term cognitive impairment [23], little is known about what causes these WMH. Studies focusing on pathological examination on WMH until now have resulted in semi-quantitative quantification methodologies, which might lead to subjective data analysis and mischaracterization of the pathological processes behind these lesions. Despite the fact that segmentation for WMH is done [19, 24], the pathology leading the progression of NAWM to WMH is unknown. Therefore, there is a need for an automatic analysis of WMH progression across images and its detailed correlation to pathological markers. In particular, current methods tend to focus on the binary classification of WMH (present or not) and possibly segmenting these WMH, but do not address the issue of WMH progression.

In this study we used the FLAIR and T1 modalities of MRI imaging together with immuno-histochemical (IHC) stainings, specifically Luxol fast blue (LFB), in order to identify the mechanisms preceding the progression of WMHs in cerebral small vessel disease. IHC is a widely used technique in pathology that allows for the visualization of specific proteins and cellular structures within tissue sections. LFB, in particular, is commonly used to stain myelin, which is a key component of white matter in the brain. The alignment of IHC stainings with MRI scans is crucial for understanding the relationship between pathological changes in the brain and imaging findings. This can provide important information on the underlying mechanisms of disease progression and potentially aid in the development of new treatment strategies.

Previous studies have demonstrated the utility of combining IHC and MRI in the analysis of cerebral small vessel disease. Many of them have found a positive correlation between LFB and (the etiology of) WMH [25–37]. By investigating changes in white matter and how these might differently impact pathology depending on the degree to which WMH is present, we expect to elucidate which mechanisms play a role in later stages of disease and ultimately lead to dementia. Our study aims to add to this body of literature by using advanced machine learning techniques to automate the detection of progression of WMHs.

The combination of post-mortem MRI and *in vivo* MRI

with IHC staining is a powerful research approach that enables the study of the relationship between radiological and pathological changes in the brain. In vivo MRI provides detailed information about the structure and function of the brain, while post-mortem MRI allows for the examination of pathological changes at a cellular level. The use of IHC staining in conjunction with MRI is particularly useful for the study of WMH, as it allows for the visualization of specific proteins and cells in the brain that can provide insight into the underlying pathological processes. For example, luxol fast blue staining can aid in identifying the myelin content in the brain tissue and help us understand the degree of myelin loss associated with WMH. By combining these techniques, we can gain a more complete understanding of the mechanisms behind WMH and their progression, as well as suggest possible treatment options.

The field of automatic detection and segmentation of WMH is still in its infancy, and most research in this area has focused on the use of in vivo MRI. However, combining post-mortem MRI with IHC stainings offers a unique opportunity to study the pathology behind WMH in a way that is not possible with in vivo MRI alone. The goal of this project is to develop an automatic segmentation pipeline for identifying and segmenting WMH and identifying the transition zones between NAWM and WMH in post-mortem brain MRI scans. By using high-field MRI and IHC stainings, we can study the relationship between radiological and pathological changes in the brain, and identify potential therapeutic targets for reducing the impact of WMH on brain function.

## 1.2 Aim of the project

The aim of this project was to develop a method for automatically identifying and segmenting WMH and identifying the transition zones between NAWM and WMH in post-mortem brain MRI scans. This will be accomplished through combining high-field MRI with IHC stainings on sections of the same brain regions. To date, limited research has been conducted on the relationship between radiological and pathological changes in CSVD, and there is currently no comprehensive understanding of the complete pathology behind WMH. The dataset used in this project offers the unique opportunity to accurately characterize the pathological changes associated with multiple neurodegenerative processes in the brain.

The two main research questions that will be explored in this project are:

- *How can we develop an automatic segmentation pipeline for identifying and segmenting WMH in MRI scans?*
- *How can we apply this method to post-mortem data to identify the transition zones between NAWM and WMH?*

The goal was to create a segmentation pipeline that can accurately identify and classify WMH and the transition zones between NAWM and WMH, in order to gain a better understanding of the underlying mechanisms behind WMH and their progression. By using high-field MRI and IHC stainings, we can study the relationship between radiological and pathological changes in the brain, and identify potential therapeutic targets for reducing the impact of WMH on brain function.

Overall, this research will investigate how changes in white matter impact pathology, and will help to uncover the mechanisms that drive the progression of WMH and their impact on brain health. Therefore, we believe that the possibility of developing an automatic classifier for white matter changes from our high-field MRI dataset can be beneficial for research and clinical practice, from speeding up diagnostic procedures to helping understand the progression of white matter lesions. An automatic classifier to assess WMH burden beyond the standard division of lesion or normal-appearing brain tissue will provide novel, clinically meaningful insights into the characterization of CSVD and its prognosis.

## 1.3 Impact of the research

The potential impact of the research in this field could be significant in several ways. Firstly, developing an automatic method for detecting progression of WMH using machine learning could improve the accuracy and consistency of WMH assessments, which could lead to more accurate diagnosis and treatment of CSVD. This could help to identify individuals at higher risk of stroke and cognitive decline earlier, allowing for earlier interventions to prevent or mitigate these outcomes.

Secondly, by understanding the main contributors to increased WMH burden, the research could help to identify new targets for treatment and prevention of CSVD. This could include identifying new risk factors or biomarkers for the disease, as well as new strategies for managing existing risk factors, such as hypertension.

Finally, the research could also contribute to the understanding of the underlying mechanisms of WMH and CSVD. By studying the transition zones from WMH to NAWM in individuals with diverse WMH severity and their pathology, the research could help to uncover new insights into the causes of WMH and how they progress. This could ultimately lead to the development of new therapies and interventions to slow or halt the progression of WMH and CSVD. One example of how the research could lead to new insights into the causes of WMH and potential treatments is by identifying new biomarkers associated with the disease. Additionally, identifying biomarkers such as luxol fast blue staining could also help to identify individuals at higher risk of WMH and CSVD, allowing for earlier interventions to prevent or slow the progression of the disease.

Overall, this research could lead to a better understanding of the underlying mechanisms of WMH and

CSVD, as well as new strategies for detecting and managing the disease, which could ultimately improve the lives of individuals affected by this disorder.

## 2 Materials and methods

Diffusion Tensor Imaging (DTI) has shown that WMH are surrounded by penumbra regions with different diffusion metrics which appear normal on conventional MRI [38,39]. These penumbra regions are thought to be closely related to WMH progression, but DTI is not always available for studying WMH. Although DTI could help to identify WMH, this post-mortem study does unfortunately not include DTI data. Furthermore, WMH are mainly diagnosed based on FLAIR and T1-weighted MRI, so this project aims to focus on these imaging modalities and how they relate to pathological changes.

### 2.1 In vivo MRI

In order to answer the first research question “*How can we develop an automatic segmentation pipeline for identifying and segmenting WMH in MRI scans?*”, we used the Radboud University Nijmegen Diffusion Tensor and Magnetic Resonance Imaging Cohort (RUN DMC) data [8, 40]. The MRI images were acquired at three different time points (2006: Siemens Magnetom Sonata from Munich, Germany; 2011 and 2015: Siemens Magnetom Avanto). All imaging sessions were conducted in the same center and included whole brain scans using 3D T1 magnetization-prepared rapid gradient echo (MPRAGE) imaging (voxel size 1.0 x 1.0 x 1.0 mm) and fluid-attenuated inversion recovery (FLAIR) pulse sequences. The baseline FLAIR had a voxel size of 1.2 x 1.0 x 5.0 mm and an interslice gap of 1.0 mm, while the follow-up had a voxel size of 1.2 x 1.0 x 2.5 mm and an interslice gap of 0.5 mm. The same head coil was used for all three sessions. Full acquisition details, demographic and imaging characteristics, can be found in [40].

Particularly, semi-supervised labels are used to train a model for WMH segmentation on this dataset. These labels are the result of a previous study on automated detection of WMH on this dataset [41]. Ghafoorian et al., proposed a method to accurately detect all WMH independently of their size. Briefly, this method used the RUN DMC data from the first time point (2006) to train a model, where reference annotations were manually created in a slice by slice manner by two trained readers (further details can be found in [41]).

The resulting dataset consisted of over 1000 cases of 3D T1-weighted, FLAIR MRI and semi-supervised labels of WMH, see figure 1. Due to changes in the MRI protocol after the first acquisition time point, resulting in higher resolution images for the two follow-up moments, only data from the two follow-up moments are used in this study. After removing unusable cases with bad pre-processing results, 642 cases remained.

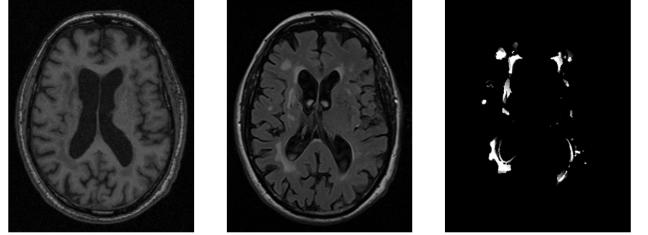


Figure 1: Representative image of a sample MRI slice from T1-weighted (left), FLAIR (middle) and the semi-supervised corresponding annotation of WMH (right).

### 2.2 Post-mortem MRI

To answer the second research question (*How can we apply the WMH segmentation method to post-mortem data to identify the transition zones between NAWM and WMH?*), we make use of a post-mortem dataset from the department of Medical Imaging, Anatomy, Radboud University Medical Center, created to investigate the underlying pathology of CSVD radiological hallmarks, including WMH [25]. This data consists of 22 post-mortem human brains that were included through the body donors’ program at the Radboud university medical center, Nijmegen. For each case, the left hemisphere was horizontally divided into a dorsal and a ventral part using reference landmarks like the corpus callosum to keep tissue processing comparable between individuals. These parts were scanned using a Bruker 7 Tesla Cliniscan MRI system (Bruker Biospin, Ettlingen, Germany) to visualize radiological hallmarks of CSVD.

This study focused on utilizing only a subset of the imaging techniques described in the paper proposed by Sole et al. [25]. Specifically, we used T1-weighted and FLAIR to segment WMH. The T1-weighted sequence has a resolution of 400x400x400  $\mu\text{m}$  and was acquired with a repetition time (TR) of 20 ms and an echo time (TE) of 1 ms, with 1 average. The FLAIR sequence has a resolution of 500x500x500  $\mu\text{m}$  and was acquired with a TR of 8200 ms and an TE of 39 ms, with 2 averages. This data is in 2D. An example case can be seen in figure 2. For complete study details, see [25].

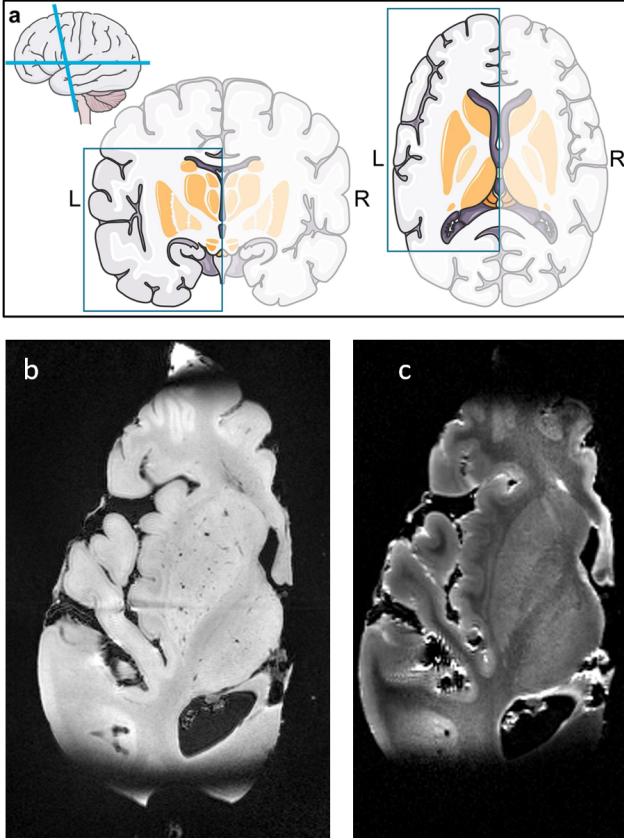


Figure 2: Post-mortem MRI scans. The left hemisphere was divided into dorsal and ventral part for HF (7 Tesla) MRI scanning. **a.** Schematic brain images. The blue squares illustrate the dimensions of the ventral part of the left hemisphere. **b.** Corresponding HF MRI T1 axial slab. **c.** Corresponding HF MRI FLAIR axial slab. Parts of the figure were drawn by using pictures from Servier Medical Art. Servier Medical Art by Servier is licensed under a Creative Commons Attribution 3.0 Unported License (<https://creativecommons.org/licenses/by/3.0/>).

Semi-automated volumetric segmentation was carried out to identify periventricular WMH within the left ventral hemisphere. Segmentation of NAWM, GM and WMH on the smaller slabs was performed manually in this region. The MRI FLAIR scans were independently evaluated by three experienced raters who were blinded to the clinical information and demographic characteristics of the study participants. The final result was a binary weighted average of the WMH, NAWM, and GM labels, based on the agreement among the raters. Regions were only labeled as WMH, GM, or NAWM if at least two out of three raters agreed on the label. An example case can be seen in figure 3.

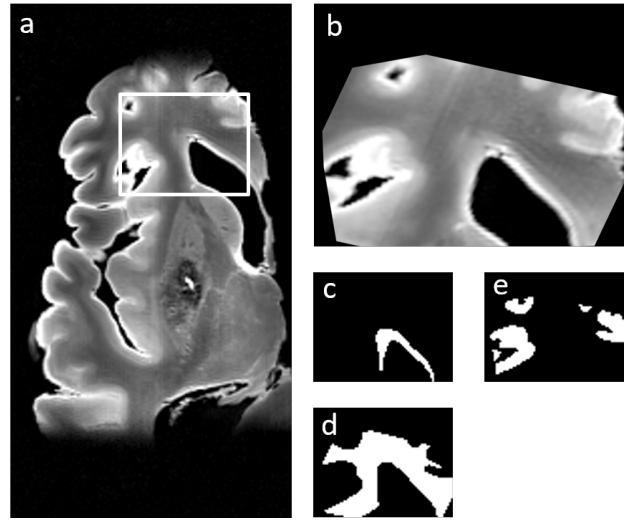


Figure 3: Post-mortem MRI FLAIR scan and corresponding labels. **a.** HF MRI FLAIR axial slab. **b.** Periventricular region, FLAIR. **c, d, e.** Corresponding WMH, NAWM and GM labels, respectively.

### 2.3 (Immuno-)histochemistry stainings

Tissue blocks from the scanned post-mortem parts were embedded in paraffin and sectioned at 4  $\mu\text{m}$  thickness. These sections were stained with luxol fast blue (LFB) to detect myelin using standard histology protocols. In this stain, myelin appears blue and the unstained regions, such as cell nuclei, appear pink or red. This is because the LFB stain binds to myelin, allowing the neuroanatomist to visualize the distribution and extent of myelin in the tissue. By using this stain, researchers can evaluate myelin pallor as an indirect measure of demyelination.

#### 2.3.1 MRI-pathology co-registration

Prior to the registration of the FLAIR MRI and LFB staining images, the MRI data of the 22 individuals was loaded into MANGO [42]. The FLAIR MRI data was compared to the LFB reference section to select the corresponding 2D MRI slide. The selection was based on the comparison of different anatomical landmarks (e.g., corpus callosum, caudate nucleus, cortex) across MRI-pathology performed by experienced neuroanatomists. The FLAIR MRI data was then registered to the LFB reference section based on manual landmark selection using a custom MATLAB script. This was performed in a previous study, [25].

At least 10 landmarks were selected on both 2D MRI slide and LFB reference section, and the MRI 2D images were warped and cropped based on the LFB reference. After MRI-pathology registration, WMH, NAWM and GM were manually segmented, as mentioned before. The ROIs corresponding to WMH and NAWM were defined for further analysis when agreement was met for at least 2 assessors. An example case, where a FLAIR image and LFB staining are registered can be seen in figure 4.

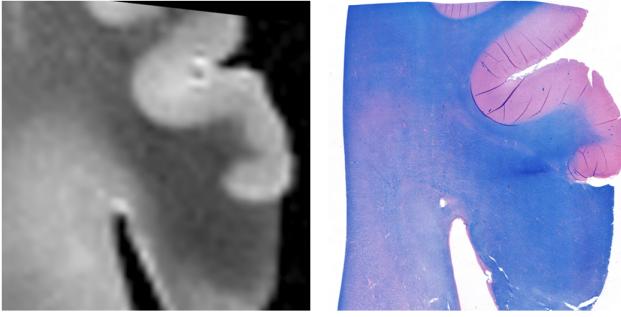


Figure 4: Example case where a FLAIR image (left) and LFB staining (right) are registered.

## 2.4 Preprocessing

### 2.4.1 In vivo MRI

Preprocessing is a crucial step in the analysis of medical imaging data, as it ensures images quality and prevent image variations and other potential sources of bias. FSL, or FMRIB’s Software Library, is a powerful open-source tool for medical image analysis [43]. In our study, we used FSL to preprocess the in vivo MRI FLAIR, T1 and semi-supervised WMH labels. MRI preprocessing helps to correct for any variations that can arise due to differences in the scanning process, such as subject positioning, and head movements.

Firstly, the T1-weighted MRI scan was mapped to FLAIR using FSL FLIRT [44]. This allowed the two scans to be processed consistently. The second step involved the removal of the skull from the mapped T1 scan using FSL BET [45]. This step helped to eliminate any non-brain tissue from the T1 scan, allowing for a clearer and more accurate representation of the brain structure. The third step involved removing the skull from the FLAIR scan, using the mask obtained from the previous step and FSL `fslmaths -mul`. This step also helped to eliminate any non-brain tissue from the FLAIR scan. The fourth step involved mapping the extracted T1 brain to MNI-152 standard space [46] using FSL FLIRT [44] and FNIRT [47]. This step was necessary as it allowed for the comparison of the extracted T1 brain to a standard reference, which provided a consistent and reliable basis for analysis. The fifth and sixth steps involved mapping the extracted FLAIR brain and WMH label to MNI-152 standard space using the same transformations as in step four. This ensured that both scans were in a standardized format, making it easier to compare to other scans and perform further analysis.

Having the T1, FLAIR, and corresponding WMH label in MNI-152 standard space is important as it allows for cross-comparison between studies, and provides a consistent reference for data analysis and interpretation. This standardization enables researchers to perform more accurate and reliable analyses, helping to ensure that the results obtained are meaningful and representative of the underlying biology.

After using FSL, FLAIR, T1, and WMH label im-

ages were normalized so that their intensities lie between 0 and 1. Furthermore, since WMH labels were obtained through a semi-supervised process, their values were not be limited to just 0 or 1. Hence, the label values were thresholded to either 0 or 1 to make the labels binary, which helps in clearly defining the regions of interest in the images.

The result of these preprocessing steps can be seen in figure 5.

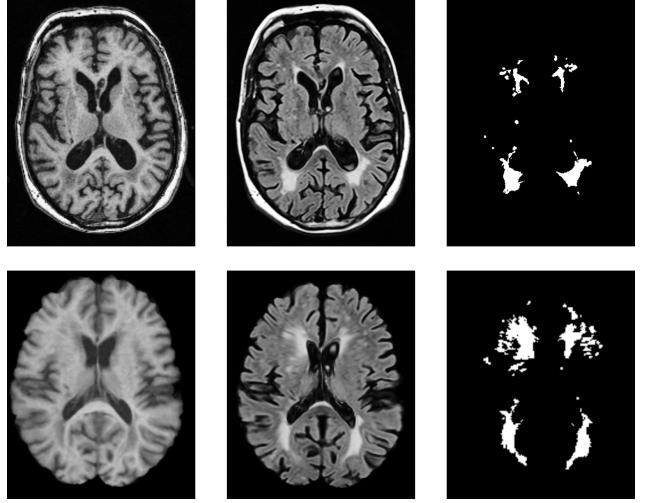


Figure 5: In vivo MRI before and after preprocessing. **From left to right:** T1-weighted image, FLAIR image and semi-supervised WMH label. **Top row:** raw MRI images. **Bottom row:** preprocessed MRI images.

It is worth mentioning that, since the MNI-152 standard space has a fixed size of  $182 \times 218 \times 182$  (w×h×d), there is no need to rescale the images as they are already the same size.

### 2.4.2 Post-mortem MRI

In this study, the preprocessing steps for post-mortem MRI data differed from the in vivo MRI data due to the differences in data format. The post-mortem MRI data is 2D and the FLAIR and label images have already been aligned and labeled with the regions of WMH, NAWM, and GM.

Prior to analysis, T1 image is firstly mapped to FLAIR image using the FSL FLIRT tool [44]. This step is necessary to ensure that the T1 and FLAIR images are in the same coordinate space. Next, the intensities of both FLAIR and T1 images are normalized to lie between 0 and 1. This normalization is important as it allows for a better comparison between the images. Finally, the FLAIR, T1 and label images are resized to  $200 \times 200$ . This resizing is necessary to ensure that the data can be efficiently processed by the FCN U-Net model. In addition to these preprocessing steps, the WMH, NAWM, and GM labels are also merged into a single label image. This is achieved by taking the values of each label image, WMH, NAWM, and GM, and multiplying each of them

by a scalar factor. The resulting values are then summed to produce the merged label image. This operation effectively assigns a unique label to each voxel in the merged label image, representing the WMH, NAWM, and GM regions in the post-mortem MRI data. Table 1 shows the values to the corresponding classes.

| Class      | Value |
|------------|-------|
| Background | 0     |
| GM         | 1     |
| NAWM       | 2     |
| WMH        | 3     |

Table 1: Values assigned to different classes in the multi-class label.

#### 2.4.3 IHC stainings

In order to perform correlation between these stainings and the transition zones of the WMH and NAWM, these stainings needed to have the same resolution of  $200 \times 200$ . However, the resolution of the raw LFB stainings ( $0.25 \mu\text{m}/\text{pixel}$ ) was really high. Therefore, a pipeline was designed that uses Fiji [48] to remove the background noise and downscale the image to the MRI resolution. Then, the image was rescaled to  $200 \times 200$ .

### 2.5 Data augmentation

#### 2.5.1 In vivo MRI

Data augmentation was not performed on in vivo MRI. In this study, the 3D in vivo MRI data of 642 patients were processed as axial slices of 2D images. The vast amount of data obtained through this approach ensured that the model was not prone to overfitting. Additionally, since the images were mapped to the MNI-152 standard space, there was no need for data augmentation. This is because augmentation can lead to increased complexity in the data and the mapping to a standard space already accounts for variations in imaging.

#### 2.5.2 Post-mortem MRI

For the post-mortem MRI dataset, which consists of 44 2D slices of T1, FLAIR, and corresponding labels for WMH, NAWM, and GM, the limited number of images is a challenge for training a deep learning model for segmentation. To address this issue, several data augmentation techniques were applied to increase the size of the dataset. These techniques included vertical flipping, random rotations of 90 degrees, elastic transformations, grid distortions, Gaussian noise, random brightness contrast, and random gamma. Each original image was augmented 50 times, each time with a different random variation of the previously listed methods.

The data augmentation was done using the Python library Albumentations [49]. This data augmentation approach includes a vertical flip with a probability of 0.5

and a random rotation with a probability of 0.5. Additionally, the pipeline includes one of the augmentation methods such as elastic transformation, grid distortion, and Gaussian noise with a probability of 0.8. Finally, the pipeline includes random brightness contrast and random gamma with a probability of 0.8. The overall probability of the pipeline is 0.95, meaning that 95% of the time the augmentation will be applied to the image.

As a result, the dataset size was increased 50-fold, providing a much larger and more diverse set of images for model training. Figure 6 shows an example of the resulting slices and labels after applying the augmentation methods.

### 2.6 Fully Convolutional Network (FCN) U-Nets

In this study, we utilized the fully convolutional network (FCN) architecture in the form of U-Nets for the segmentation of both in vivo (WMH) and post-mortem (WMH, NAWM, GM) MRI scans. FCN is a type of neural network architecture that is specifically designed for image processing tasks, as it processes an entire image rather than a sequence of patches [50]. The use of FCN in medical image analysis has become increasingly popular due to its ability to accurately extract features from images. The U-Net architecture, in particular, has been widely used in medical image segmentation due to its ability to retain spatial information while reducing computational costs [51]. The architecture of the U-Net is based on the winning entry [20] of the WMH Segmentation Challenge [19] and the results of the challenge itself. The challenge is an international competition that evaluates the performance of different algorithms for detecting and quantifying WMH in MRI scans. The code of this winning method was made publicly available.

The resulting architecture incorporated techniques and parameters that have been demonstrated to be effective in the winning method of the challenge, and incorporated the insights gained from the challenge results themselves. In this regard, the same number of features, kernel size in each convolutional layer, number of layers and loss function as the winning method were used. The main difference with the winning method is the introduction of dropout and batch normalization layers after each convolutional layer. Dropout reduces the risk of overfitting by randomly excluding certain neurons during training [52]. Batch normalization, on the other hand, ensures that the inputs to each layer have a consistent distribution, improving the performance of the model [53]. The results of the WMH Segmentation Challenge indicate that the use of dropout in U-Net architecture can significantly improve performance for this task. Together, these techniques help to improve the accuracy and generalization of the U-Net, making it a powerful tool for image segmentation tasks. This is especially useful for the segmentation of the post-mortem MRI images, which are scarce in quantity.

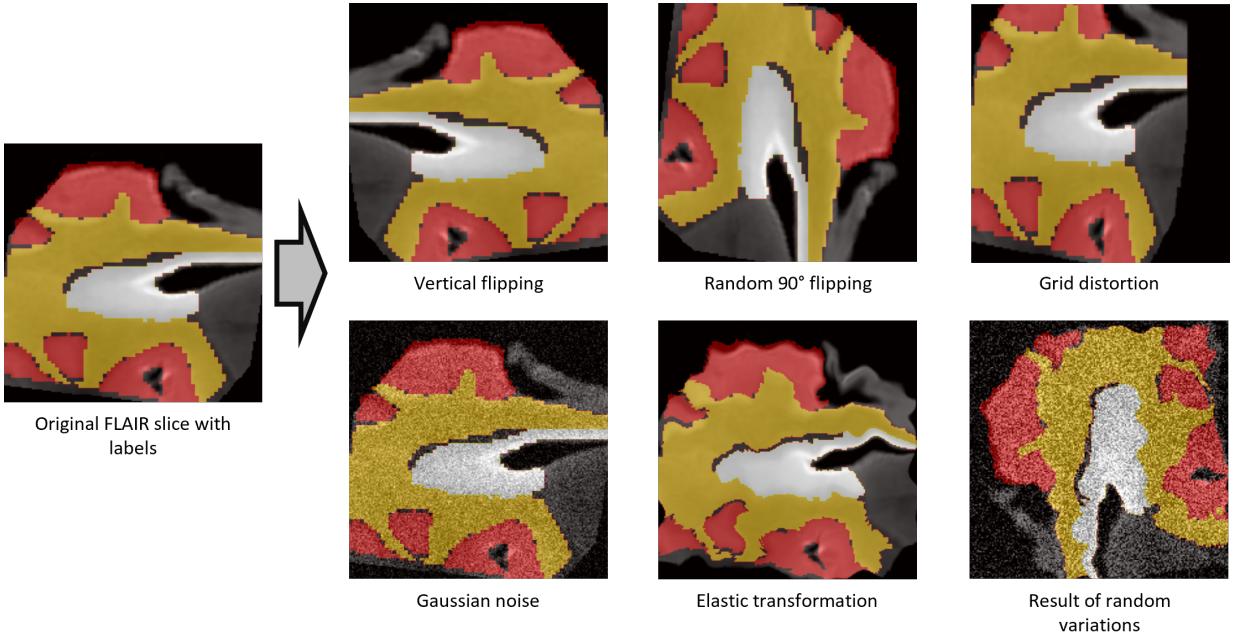


Figure 6: An example of data augmentation results. These images contain a FLAIR image, WMH label (white), NAWM label (yellow) and GM label (red).

### 2.6.1 In vivo MRI

For the in vivo MRI scans, we used a binary U-Net for WMH segmentation. The architecture of the U-Net consists of a down-convolutional part that shrinks the spatial dimensions of the image and an up-convolutional part that expands the score maps. The skip connections between these two parts allow for the transfer of information between them, allowing the network to have a more complete understanding of the image. The final layer of the network is a  $1 \times 1$  convolutional layer with a sigmoid activation function that maps each feature vector to two classes, corresponding to the presence or absence of WMH.

### 2.6.2 Post-mortem MRI

For the post-mortem MRI scans, we employed a multi-class U-Net for WMH, NAWM, and GM segmentation. The architecture of this U-Net is similar to the binary U-Net, but it includes a final layer that maps the feature vectors to four classes, representing WMH, NAWM, GM, and background pixels. This final layer is implemented using a  $1 \times 1$  convolutional layer with four filters and a softmax activation function. The use of the softmax activation function is important in this context because it provides a probability-based prediction for each class, which enables a more nuanced understanding of the image. The multiclass U-Net approach allows for a more detailed characterization of the lesions in the post-mortem scans, as compared to a binary segmentation approach. The difference in the final layer of the two U-Nets is due to the different nature of the segmentation tasks: a binary segmentation task for the in vivo scans and a multiclass

segmentation task for the post-mortem scans. The resulting U-Net can be seen in figure 7.

### 2.6.3 (Categorical) Dice loss

The Dice loss function is a widely used loss function for image segmentation tasks, particularly when using a U-Net architecture. It measures the overlap in percentage between two segmentations, prediction and ground truth. One of the reasons for its popularity is its ability to handle strong imbalances between positive and negative samples, which is common in biomedical images such as the segmentation of WMH versus healthy tissue [54]. The Dice loss is defined as follows:

$$Dice_{Loss} = -\frac{2 \sum_i^N y_i \cdot \hat{y}_i + \epsilon}{\sum_i^N y_i + \sum_i^N \hat{y}_i + \epsilon}$$

where  $y_i$  is the ground truth label,  $\hat{y}_i$  is the predicted label,  $N$  is the total number of pixels, and  $\epsilon$  is a (small) constant to prevent division by zero.  $\epsilon$  was set to 1. The Dice loss measures the similarity between the ground truth and the predicted labels, with a value close to 1 indicating a good segmentation and 0 being the worst possible segmentation with no overlap between label and prediction. The use of the Dice loss can help to optimize the U-Net for accurate segmentation, by penalizing predictions that are significantly different from the ground truth.

The categorical Dice loss is a variant of the Dice loss, designed for multi-class image segmentation problems, where each pixel belongs to one of several categories or classes. In this scenario, the standard Dice loss is extended to account for multiple classes by calculating the Dice loss for each class separately. The categorical Dice

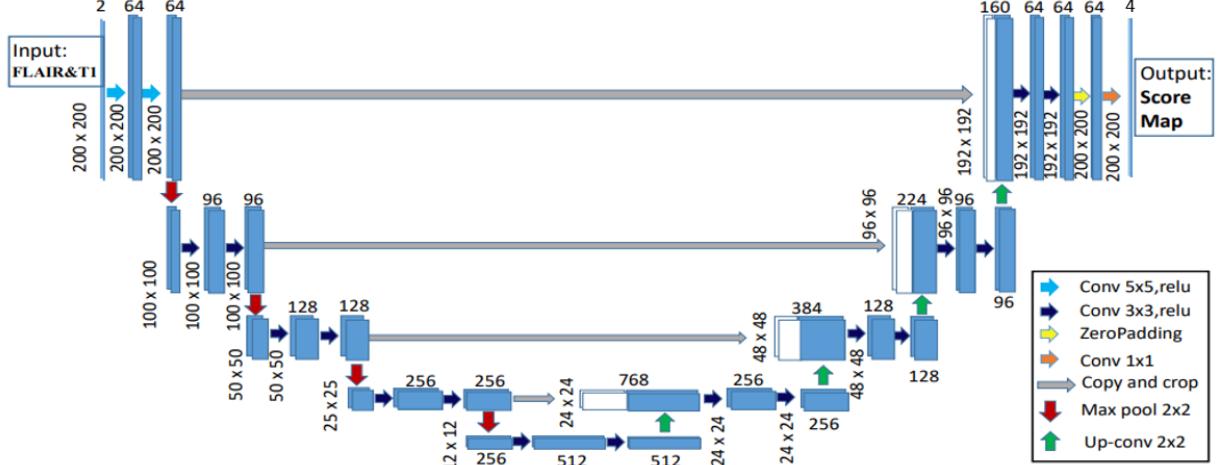


Figure 7: The 2D U-Net architecture illustrated is designed for post-mortem MRI and consists of two parts: a shrinking part on the left side, and an expansive part on the right side. The model uses both FLAIR and T1 channels as inputs and is trained to detect and locate WMHs. The 2D U-Net for in vivo MRI is slightly different, with differences in the input resolution, output resolution, and other layer resolutions. Additionally, the final layer in the in vivo model has one filter, as opposed to the post-mortem 2D U-Net. This image is taken from the work of Li et al. [20].

loss is defined as follows:

$$CatDiceLoss = - \sum_{c=1}^C \frac{2 \sum_i^N y_{i,c} \cdot \hat{y}_{i,c} + \epsilon}{\sum_i^N y_{i,c} + \sum_i^N \hat{y}_{i,c} + \epsilon}$$

where  $C$  is the total number of classes,  $N$  is the total number of pixels,  $y_{i,c}$  is the ground truth label for the  $i^{th}$  pixel in class  $c$ ,  $\hat{y}_{i,c}$  is the predicted label for the same pixel and class.  $\epsilon$  is a small positive constant added to the numerator and denominator to prevent division by zero, again set to 1.

The categorical Dice loss can become unreliable in situations where some of the classes have much fewer instances than others, a phenomenon known as class imbalance. To account for class imbalance, class weights were added to the calculation of the categorical Dice loss. The class weights indicate the relative importance of each class, with larger weights assigned to classes that are underrepresented in the data. The final step in the calculation of the categorical Dice loss was to divide it by the sum of the class weights. This normalization step ensures that the loss value is between 0 and 1, making it easier to interpret and compare across different datasets with different class distributions. The weighted categorical Dice loss is defined as follows:

$$WCatDiceLoss = - \frac{1}{\sum_{c=1}^C w_c} \sum_{c=1}^C w_c \cdot \frac{2 \sum_i^N y_{i,c} \cdot \hat{y}_{i,c} + \epsilon}{\sum_i^N y_{i,c} + \sum_i^N \hat{y}_{i,c} + \epsilon}$$

where the only difference with the categorical dice loss is the division by the sum of the class weights.

The class weights were computed using scikit-learn [55] and can be seen in table 2.

| Class      | Weight |
|------------|--------|
| Background | 0.61   |
| GM         | 1.60   |
| NAWM       | 0.67   |
| WMH        | 3.46   |

Table 2: Class weights.

The weighted categorical Dice loss calculation gives a higher penalty for misclassification of samples from a class with fewer samples, which helps to prevent the model from overfitting to the class with more samples by ensuring a more balanced representation of the different classes in the training data.

## 2.7 Ensemble networks

The use of ensemble networks has become a popular technique in computer vision and medical image analysis due to its ability to improve predictive performance. Ensemble networks combine multiple learning models to obtain better results than any of the constituent learning algorithms alone. This approach is useful in reducing over-fitting problems of complex models on the training data [56].

The WMH Segmentation Challenge [19] demonstrated that ensemble methods are effective for this task. The winning entry of the challenge used ensemble networks to improve the automated WMH segmentation. The authors of the winning method [20] noted that different models could learn different attributes of the training data, resulting in improved segmentation results. Additionally, they discussed the use of ensemble networks in addressing the bias-variance trade-off. The ensemble of models

can lower the variance by averaging the model outputs, reducing the overall error of the network.

In this study, we also incorporated the use of ensemble networks in our fully convolutional network (FCN) architecture, in the form of U-Nets, for the segmentation of both in vivo and post-mortem MRI scans. Our network architecture is inspired by the winning method [20] and incorporates the use of three multiple U-Net models, each trained with random parameter initialization and shuffled data in the batch learning process. For each U-Net model, a probability segmentation map is generated when given a test image. The resulting maps are then averaged, and an empirically-picked threshold is used to transform the scores map into a binary segmentation map. The use of ensemble networks in our architecture serves to improve the accuracy and generalization of the model, making it a powerful tool for image segmentation tasks. The resulting pipeline can be seen in figure 8.

## 2.8 Post Processing

### 2.8.1 In vivo MRI

In our study, we used a 2D U-Net to segment in vivo MRI data. However, since the original MRI images consist of 182 slices, we treated each slice as a separate 2D input for the U-Net. This required us to run the network 182 times for each patient, where each run produced a prediction for a single slice. The only post-processing step required was to combine these 182 individual slice predictions into the 3D format of the MNI-152 standard space, resulting in a complete prediction for the entire MRI volume of the patient.

### 2.8.2 Post-mortem MRI

In the post-processing of the post-mortem dataset, the T1-weighted and FLAIR images were registered using FSL to align the two images. However, sometimes the images visualized slightly different regions, making it necessary to use both T1 and FLAIR as an inclusion mask.

The first step in the script involved creating a binary mask from the T1-weighted image. This function created a mask with zeros where the T1 image had zero intensity values, and ones where it had non-zero values. Similarly, a binary mask was created from the FLAIR image. This post-processing step was crucial in ensuring that both T1 and FLAIR pixels were included in the segmentation of WMH, NAWM, and GM, as both images were necessary to accurately segment the three regions. The result of this step was a post-processed prediction mask that accurately reflected the T1 and FLAIR pixels.

## 2.9 Transition zones

In this study, we aimed to investigate the progression of WMH and NAWM in post-mortem data using FLAIR and T1 weighted images. It is important to note that prior to creating the transition zones, an accurate segmentation

of WMH, NAWM, and GM was necessary. This is because without segmenting the regions, it is not possible to distinguish between WMH and GM, as both show high intensity values in FLAIR.

After successful segmentation of the three regions, we used the intensity values from FLAIR image and KMeans clustering algorithm from scikit-learn [55] to create three transition zones within the WMH and NAWM segmentations. The intensity values were used to partition the pixels of the WMH/NAWM region into three clusters. The mean intensity for each cluster was then calculated and the clusters were sorted based on the mean intensity.

## 2.10 MRI-pathology correlation

To study the underlying pathology of WMH, a correlation analysis was performed between the MRI data and the LFB staining results. The correlation between the MRI clusters and the LFB staining intensity was analyzed using the Pearson correlation coefficient ( $r$ ). In this study, LFB stainings were transformed to greyscale images, where high pixel intensity corresponds to white or bright regions, indicating a lack of myelin in the image. To obtain the LFB intensity values for each cluster, the voxel intensity values from each cluster were extracted and correlated with the corresponding LFB intensity values. This was repeated for all the images, and the results were combined to obtain an overall correlation between the MRI clusters and the LFB staining.

## 3 Results

The results section of this research paper presents the findings from the in-vivo MRI WMH segmentation, post-mortem MRI periventricular WMH, NAWM and GM segmentation, clustering of the three regions within the WMH and NAWM regions, and the correlations between these clusters and the Luxol Fast Blue staining of these regions. This section provides valuable insights into the progression of WMH and NAWM and their relationship to the pathology of the white matter.

### 3.1 In vivo segmentation

To evaluate the convergence of this model, we present the training and validation curve of one of the three ensemble models. Figure 9 shows the training and validation Dice coefficient over epochs. The figure indicates that the model converges and generalizes well, with the training curve reaching a peak value of 0.76, and the validation curve reaching a peak value of 0.77, compared to 0.805 for an independent human observer [57]. These results indicate that the model has learned a robust representation of the data and has good generalization performance. The training and validation curves of the other two ensemble models are also very similar, displaying a comparable trend in accuracy during training.

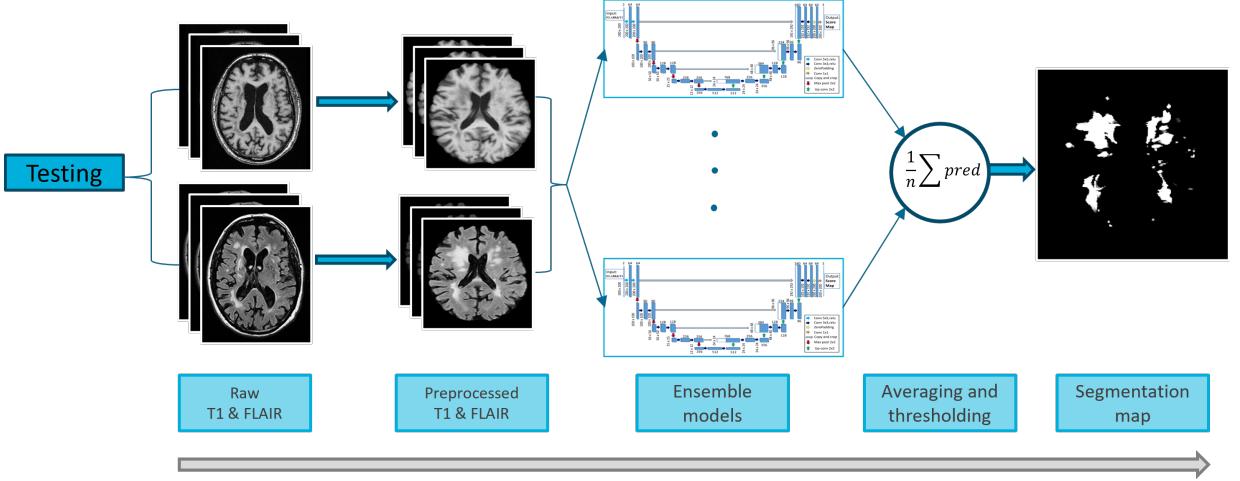


Figure 8: Overall framework of the testing stage.

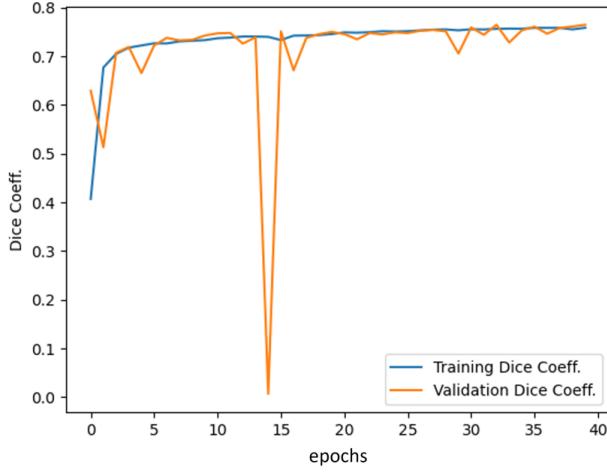


Figure 9: Training and validation curve of one out of three ensemble models. Training and validation Dice coefficient over epochs. This is equal to the inverse loss.

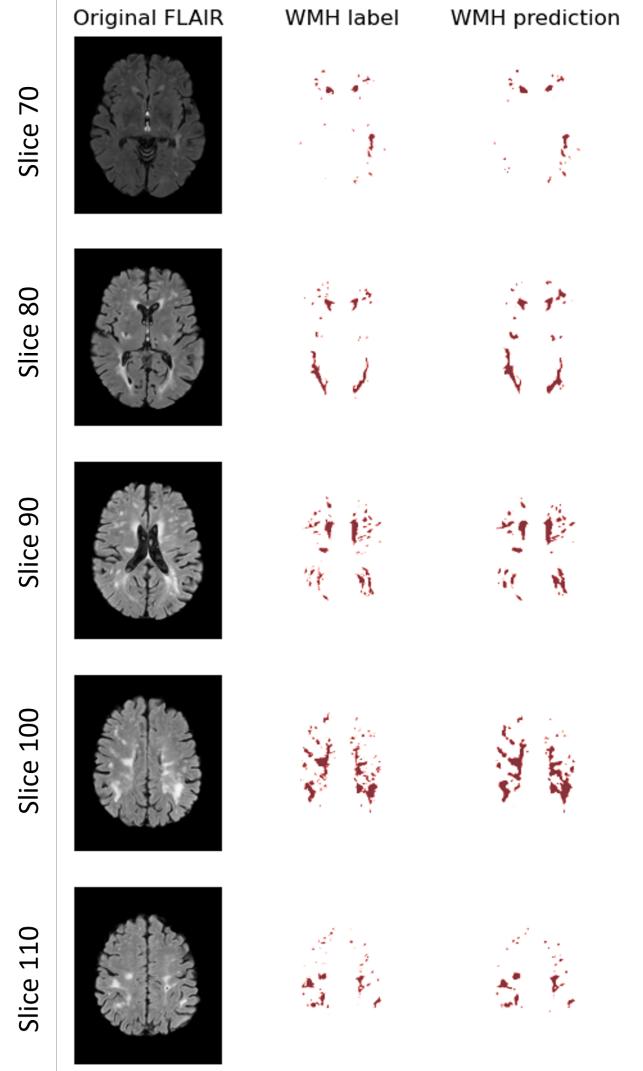


Figure 10: Example of in vivo MRI WMH segmentation results. The first column shows the original FLAIR image, the second column displays the ground truth WMH label, and the third column shows the corresponding WMH prediction generated by the model.

In addition to the performance metrics, we also provide a qualitative evaluation of the model’s performance by presenting a few examples of the segmentation results. Figure 10 displays a comparison of the original FLAIR image, ground truth WMH label, and the corresponding WMH prediction generated by the model. The results indicate that the model is able to accurately capture the presence and location of WMH in the FLAIR images, with high correspondence between the prediction and ground truth label. This suggests that the model is able to effectively identify WMH and can be used for practical applications.

### 3.2 Post-mortem segmentation

To evaluate the convergence and generalization ability of the models, we plotted the training and validation curves of one of the three ensemble models. Figure 11 displays the training and validation Dice coefficient over epochs, showing the performance of the model as it trains on the dataset. The three ensemble models displayed similar performance, with close training Dice coefficients around 0.942 and validation Dice coefficients around 0.92. The Dice score on the test set was 0.94, indicating a good overall performance of the models in terms of their ability to correctly segment the target structures in the MRI data.

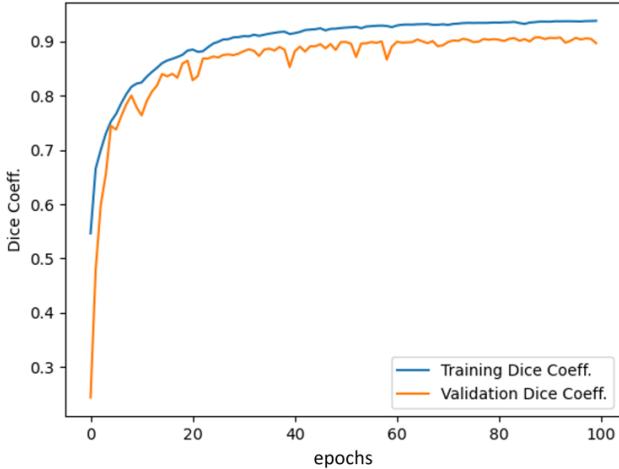


Figure 11: Training and validation Dice coefficient over epochs. This is equal to the inverse loss.

In addition to the training and validation curves, we also provide a qualitative evaluation of the model’s performance by presenting a few examples of the segmentation results. Figure 12 displays three examples of the multi-class segmentation of WMH, NAWM, and GM. The results indicate that the model is able to accurately differentiate between the three regions in the post-mortem MRI data and can be used for practical applications in the medical imaging field.

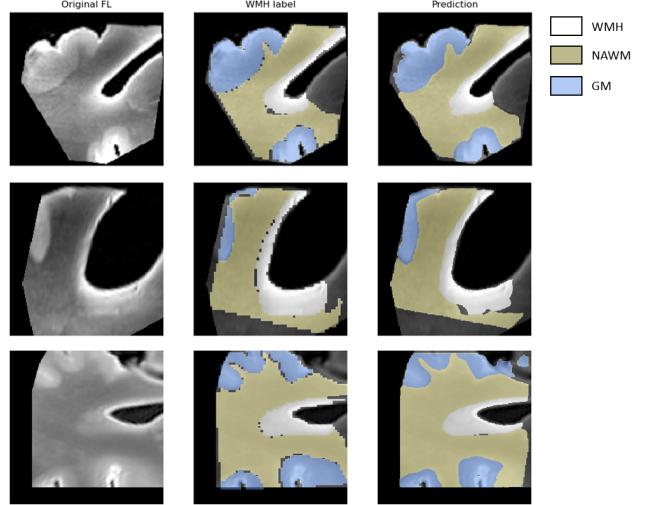


Figure 12: Example results of the multi-class segmentation of WMH, NAWM, and GM in post-mortem MRI data on a test set. The original MRI data, the ground truth labels, and the predicted segmentations are shown for three different cases.

### 3.3 Transition zones

Figure 13 shows the transition zones between WMH and NAWM for three example images. The image displays the original FLAIR image, alongside the corresponding labels and segmentation results. The darker red regions indicate a more severe case of WMH, while the darker blue regions indicate a more severe case of NAWM. The figure demonstrates the ability of our method to accurately capture the transition zones between the two regions, thus providing a more detailed understanding of the underlying structures in the MRI data.

It is worth noting that, in addition to WMH and NAWM, the model also segments GM in the images. However, for this analysis, we have focused on the transition zones between WMH and NAWM and have excluded the GM segmentation from the visual representation.

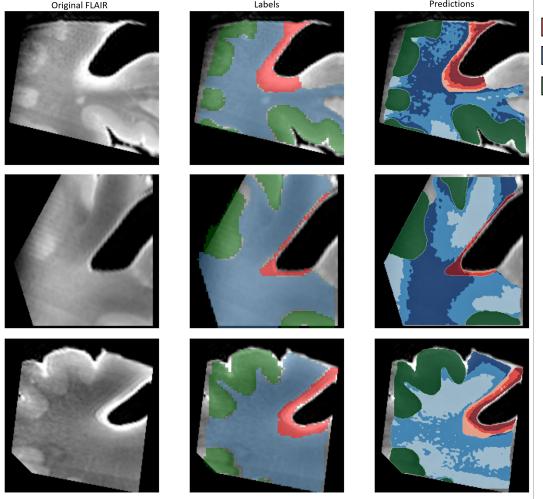


Figure 13: Clustered representation of WMH and NAWM transition zones in three example MRI slices. From left to right: raw FLAIR, labels and transition zones. The clusters are formed based on the FLAIR intensity, with darker red indicating a more severe case of WMH and darker blue indicating a more severe case of NAWM.

### 3.4 MRI-pathology correlation

In this section, we present the results of the correlations between the clustering of WMH and NAWM regions and the LFB staining intensity of these regions. Luxol Fast Blue is a histological stain used to assess the presence of myelin in the brain tissue. We used the LFB staining pixel intensity as a measure of the degree of myelin loss in the WMH and NAWM regions. The positive correlation between the clustering of these regions and the LFB staining intensity, as shown in the box plots presented in figures 14 and 15, confirms the relationship between myelin loss and the severity of WMH and NAWM.

We found a significant positive correlation between the clustering of WMH and NAWM regions and the LFB staining pixel intensity. The Pearson correlation coefficients are 0.26 and 0.15, respectively, with p-values  $< 0.0001$ . The results are shown in 3 box plots each for the WMH and NAWM regions, illustrating the correlation between the clustering of each region and the corresponding LFB pixel intensity.

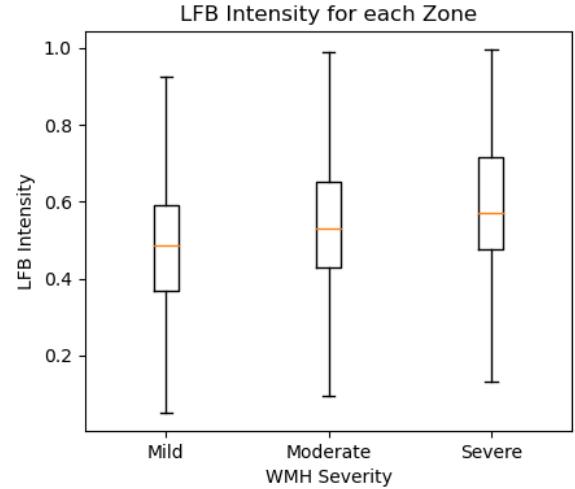


Figure 14: Box plot illustration of the correlation between clustering of WMH regions and LFB staining intensity.

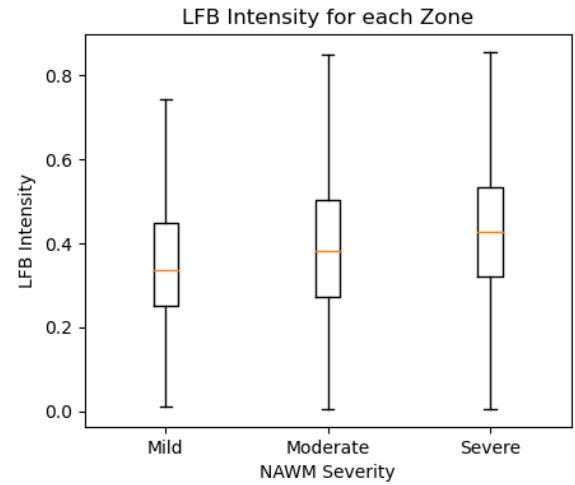


Figure 15: Box plot illustration of the correlation between clustering of NAWM regions and LFB staining intensity.

The findings suggest that there is a positive relationship between the clustering of WMH and NAWM regions and the degree of myelin loss as measured by LFB staining pixel intensity.

## 4 Discussion

This study aimed to develop an automatic segmentation pipeline for identifying and segmenting WMH in MRI scans and to investigate the relationship between WMH and myelin loss in post-mortem tissue. We found a significant positive correlation between the clustering of WMH and NAWM regions and the degree of myelin loss, indicating that myelin loss is a contributor to the development of WMH.

The potential impact of this research is significant, as it leads to more accurate diagnosis and treatment of

CSVD, the identification of new targets for treatment and prevention, and the development of new therapies and interventions to slow or halt the progression of WMH and CSVD. Our study provides a foundation for future research to perform comprehensive analyses of WMH and NAWM clusters in conjunction with IHC stainings and other markers. Additionally, the automatic segmentation pipelines for in vivo and post-mortem MRI developed in this study could be utilized to perform large-scale studies on populations with different demographics, genetic risk factors, and disease states.

#### 4.1 Limitations of Dice score in in vivo MRI segmentation model and the need for visual inspection

The in vivo MRI segmentation model has shown promising results in the automatic segmentation of WMH. However, one limitation of the model is that the WMH labels are semi-automatic, which sometimes result in incorrect labels of certain regions. Specifically, during visual inspection, we came to the conclusion that part of the ventricles were sometimes labeled as WMH, while other regions that should have been annotated were not. Therefore, the Dice score, which is commonly used to evaluate segmentation models, cannot be relied on completely in this context.

Despite this limitation, visual inspection revealed that the model’s predictions often outperformed the labels, suggesting that the model is still useful in automatic segmentation, see figure 16. However, the lack of a reliable Dice score also highlights the need for alternative evaluation methods. In this study, we addressed this issue by using a subset of 100 cases with poorly annotated data for testing the model’s performance on unseen data.

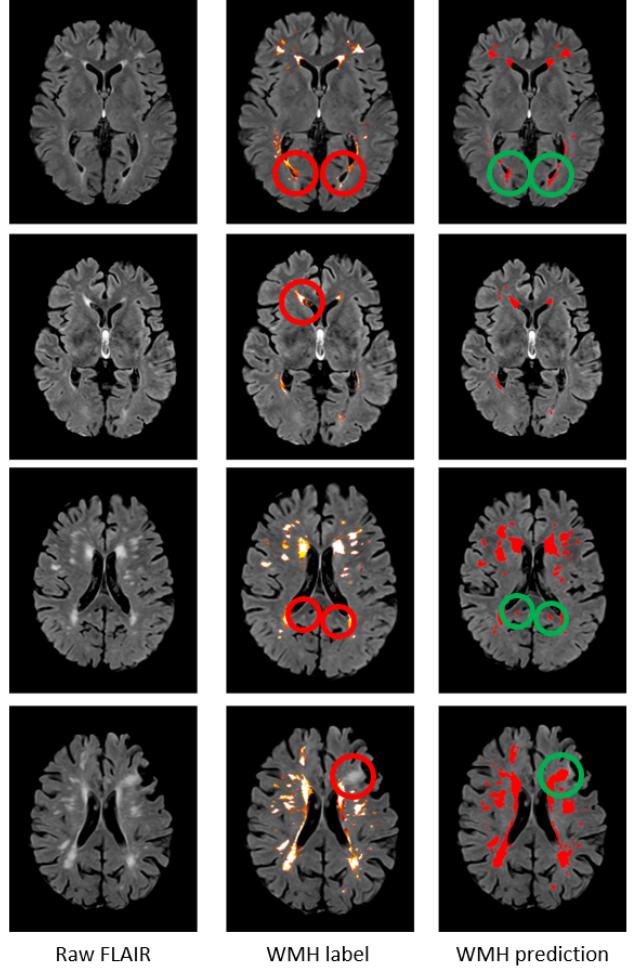


Figure 16: Improvements in WMH Segmentation: Comparison of Raw FLAIR Image, WMH Label, and Model Prediction. This image shows the raw FLAIR image on the left, the WMH label in the middle, and the model’s prediction on the right. Red circles indicate areas with bad annotation in the label, while green circles indicate improvements made by the model’s prediction.

In conclusion, the in vivo MRI segmentation model shows promise for automatic segmentation of WMH, but the limitations of the Dice score emphasize the need for visual inspection. The use of a test set with poorly annotated data allows for a more comprehensive evaluation of the model’s performance on unseen data.

The fact that the in vivo MRI segmentation model filters out errors during training means that the resulting model is robust. By excluding poorly annotated cases from the training data, we ensured that the model is not biased towards such errors. Figure 17 shows example cases for the test set. This leads to a more accurate segmentation, even when the data contains some inaccuracies. Additionally, the fact that the model outperforms the labels in some cases when visually inspected suggests that it has learned to identify patterns in the data that even human annotators may have missed. This can be a valuable feature when dealing with large amounts of data, as it can reduce the need for manual inspection and cor-

rection of the labels. Despite the limitations of the Dice score, it remains a commonly used metric for evaluating the performance of segmentation models. In this study, we used it in combination with visual inspection to ensure the quality and reliability of our results.

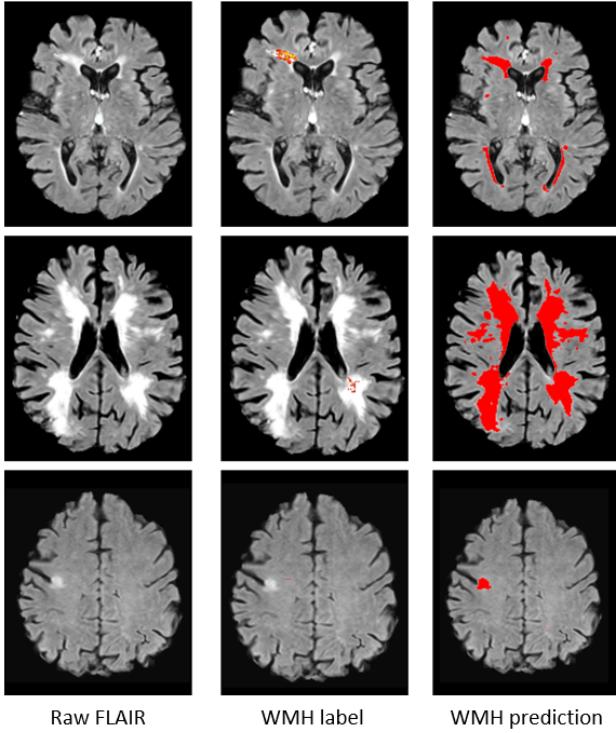


Figure 17: Test set of WMH segmentation: Comparison of Raw FLAIR Image, Bad Label, and Model Prediction. This image shows the raw FLAIR image on the left, a bad label in the middle, and the model’s prediction on the right. The bad label contains errors, making it unsuitable for training the model. This image was used for testing the model’s ability to perform well on unseen data.

## 4.2 Applying in vivo method to post-mortem data

The development of an automatic segmentation pipeline for identifying and segmenting WMH in MRI scans was the first research question of this project. However, when we applied this method to post-mortem data to identify the transition zones between NAWM and WMH, we found that the results were not satisfactory. The reason for this was that the post-mortem data looked significantly different from the in-vivo data that was used to develop the original model. Initially, we tried to augment the in-vivo data to make it look like the post-mortem data, since we had a lot of in-vivo data available. We hoped that this new model, with augmented in-vivo data, would work well on post-mortem data. However, this was not the case, and we had to create a new model from scratch for the post-mortem data.

Creating a new model from scratch for the post-mortem data was a challenging task because we only had

44 2D images for training the model. This is a relatively small amount of data for training a deep learning model, and there was a high risk of overfitting. To address this issue, we used data augmentation techniques and methods to prevent overfitting, such as dropout and ensemble networks. With these techniques, we were able to develop a new model that gave satisfactory results for post-mortem data.

The development of a new model for post-mortem data has several implications. Firstly, it shows that the model developed for in-vivo data cannot be directly applied to post-mortem data without modification. This is an important consideration for researchers who may want to apply the same method to both types of data. Additionally, our success in developing an accurate segmentation model for post-mortem data with a small amount of data suggests that it is possible to develop accurate segmentation models for other types of data with limited resources.

## 4.3 Different approaches to multi-class segmentation

In this study, two approaches were explored for the segmentation of WMH, NAWM and GM in post-mortem MRI scans. The first approach involved creating a multi-class segmentation model, while the second approach involved creating three separate models for each of the three tissue types and combining the segmentations.

The results of the study showed that the multi-class segmentation model outperformed the three separate models in terms of segmentation accuracy. While the three separate models had a tendency to leave space between the segmented classes (resulting in more background), the multi-class segmentation model was able to accurately segment each tissue type without any spaces between them, see figure 18.

The better performance of the multi-class segmentation model can be attributed to the fact that it takes into account the relationships between the different tissue types and uses this information to improve the segmentation accuracy. In contrast, the three separate models do not consider the relationships between the different tissue types, which can lead to errors in the segmentation results.

It’s important to note that the transition zone between WMH and NAWM is a critical area of interest in this study, and accurately identifying the borders of these regions is essential for characterizing the pathology of white matter diseases. Any space or overlap between the segmented classes in this area can lead to misclassification of tissue and potentially impact the results of the study. Therefore, the fact that the multi-class segmentation model performs better in this region is a significant advantage and should be emphasized in the discussion.

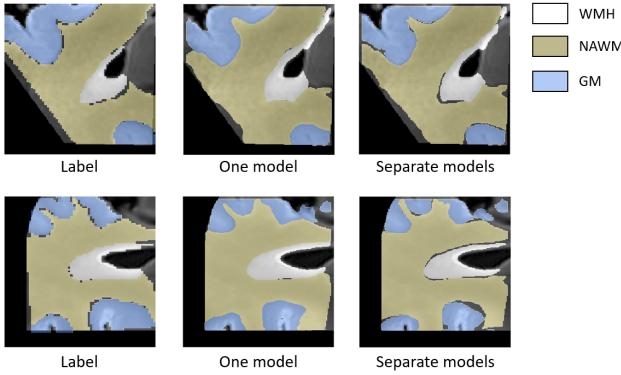


Figure 18: Comparison of segmentation results using different methods for two cases. The multi-class segmentation approach produces more accurate segmentations with minimal or no gaps between classes, while the three separate models approach leaves spaces between the segmented classes. This is particularly important in the transition zone between WMH and NAWM as this is a critical region of interest in the study.

#### 4.4 Transition zones

The creation of three transition zones within the WMH and NAWM regions provides a clearer understanding of the progression of WMHs and NAWM. The clusters with higher mean intensity can be considered as a more severe case of WMH and can be used to better understand the progression and severity of the condition. The segmentation of WMH, NAWM, and GM followed by the creation of three transition zones within the WMH and NAWM using the intensity values from the FLAIR image and KMeans clustering, provided valuable information about the progression and severity of WMHs and NAWM in post-mortem data. These results have important implications for the understanding of the pathology of WMH and NAWM and their relationship with myelin loss.

## 5 Conclusions

This research paper presents findings from various experiments conducted to understand the progression of WMHs and NAWM in human brains. The paper reports results from experiments related to in vivo and post-mortem MRI segmentation, clustering of different regions within the WMH and NAWM, and correlations between these clusters and LFB staining of these regions.

The in vivo MRI WMH segmentation experiment evaluated the performance of the proposed method using the Dice coefficient as the performance metric. The results showed a peak Dice coefficient of 0.76 for the training curve, 0.77 for the validation curve, and 0.805 for an independent human observer. Visual inspection revealed that the model’s predictions often outperformed the labels. The post-mortem MRI WMH segmentation experiment aimed to accurately segment three classes of

tissue in MRI data, namely WMH, NAWM, and GM. The performance of the model was evaluated using the weighted categorical Dice coefficient, which takes into account the imbalanced distribution of the different classes in the data. The training and validation curves of one of the three ensemble models were presented to evaluate the model’s convergence and generalization ability, with close training Dice coefficients around 0.942, validation Dice coefficients around 0.92 and test Dice coefficient of 0.94.

The clustering of transition zones between WMH and NAWM in post-mortem MRI data was conducted using k-means clustering to separate the different levels of FLAIR intensity. The results showed that the proposed method can effectively distinguish between different severity levels of WMH and NAWM using FLAIR intensity.

Finally, our study provides new insights into the underlying mechanisms of WMH and their positive relationship with myelin loss. The findings presented here have important implications for the development of new therapies and interventions to slow or halt the progression of WMH and related disorders. Future research directions may include investigating the relationship between WMH and other markers of white matter damage, such as axonal injury and inflammation, and performing larger-scale studies to further explore the pathogenesis of WMH.

## References

- [1] J. M. Wardlaw, E. E. Smith, G. J. Biessels, C. Cor弁onnier, F. Fazekas, R. Frayne, R. I. Lindley, J. T O’Brien, F. Barkhof, O. R. Benavente, *et al.*, “Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration,” *The Lancet Neurology*, vol. 12, no. 8, pp. 822–838, 2013.
- [2] S. J. van Veluw, A. Y. Shih, E. E. Smith, C. Chen, J. A. Schneider, J. M. Wardlaw, S. M. Greenberg, and G. J. Biessels, “Detection, risk factors, and functional consequences of cerebral microinfarcts,” *The Lancet Neurology*, vol. 16, no. 9, pp. 730–740, 2017.
- [3] F. De Leeuw, J. C. de Groot, E. Achten, M. Oudkerk, L. Ramos, R. Heijboer, A. Hofman, J. Jolles, J. Van Gijn, and M. Breteler, “Prevalence of cerebral white matter lesions in elderly people: a population based magnetic resonance imaging study. the rotterdam scan study,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 70, no. 1, pp. 9–14, 2001.
- [4] D. Bos, F. J. Wolters, S. K. Darweesh, M. W. Verhooij, F. de Wolf, M. A. Ikram, and A. Hofman, “Cerebral small vessel disease and the risk of dementia: a systematic review and meta-analysis of population-based evidence,” *Alzheimer’s & Dementia*, vol. 14, no. 11, pp. 1482–1492, 2018.
- [5] A. Kapasi, C. DeCarli, and J. A. Schneider, “Impact of multiple pathologies on the threshold for clinically

- overt dementia,” *Acta neuropathologica*, vol. 134, no. 2, pp. 171–186, 2017.
- [6] E. M. Van Leijsen, M. I. Bergkamp, I. W. Van Uden, M. Ghafoorian, H. M. Van Der Holst, D. G. Norris, B. Platel, A. M. Tuladhar, and F.-E. de Leeuw, “Progression of white matter hyperintensities preceded by heterogeneous decline of microstructural integrity,” *Stroke*, vol. 49, no. 6, pp. 1386–1393, 2018.
- [7] P. Sachdev, W. Wen, X. Chen, and H. Brodaty, “Progression of white matter hyperintensities in elderly individuals over 3 years,” *Neurology*, vol. 68, no. 3, pp. 214–222, 2007.
- [8] E. M. van Leijsen, I. W. Van Uden, M. Ghafoorian, M. I. Bergkamp, V. Lohner, E. C. Kooijmans, H. M. van der Holst, A. M. Tuladhar, D. G. Norris, E. J. van Dijk, et al., “Nonlinear temporal dynamics of cerebral small vessel disease: the run dmc study,” *Neurology*, vol. 89, no. 15, pp. 1569–1577, 2017.
- [9] K. W. Kim, J. R. MacFall, and M. E. Payne, “Classification of white matter lesions on magnetic resonance imaging in elderly persons,” *Biological psychiatry*, vol. 64, no. 4, pp. 273–280, 2008.
- [10] S. Debette and H. Markus, “The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis,” *Bmj*, vol. 341, 2010.
- [11] C. DeCarli and S. Lockhart, “Cerebrovascular disease,” in *Encyclopedia of the Neurological Sciences*, pp. 722–726, Elsevier Inc., 2014.
- [12] A. A. Gouw, A. Seewann, W. M. Van Der Flier, F. Barkhof, A. M. Rozemuller, P. Scheltens, and J. J. Geurts, “Heterogeneity of small vessel disease: a systematic review of mri and histopathology correlations,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 82, no. 2, pp. 126–135, 2011.
- [13] H. M. A. Abraham, L. Wolfson, N. Moscufo, C. R. Guttmann, R. F. Kaplan, and W. B. White, “Cardiovascular risk factors and small vessel disease of the brain: blood pressure, white matter lesions, and functional decline in older persons,” *Journal of Cerebral Blood Flow & Metabolism*, vol. 36, no. 1, pp. 132–142, 2016.
- [14] S. Hilal, V. Mok, Y. C. Youn, A. Wong, M. K. Ikram, and C. L.-H. Chen, “Prevalence, risk factors and consequences of cerebral small vessel diseases: data from three asian countries,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 88, no. 8, pp. 669–674, 2017.
- [15] J. M. Wardlaw, C. Smith, and M. Dichgans, “Mechanisms of sporadic cerebral small vessel disease: insights from neuroimaging,” *The Lancet Neurology*, vol. 12, no. 5, pp. 483–497, 2013.
- [16] J. H. Veldink, P. Scheltens, C. Jonker, and L. J. Launer, “Progression of cerebral white matter hyperintensities on mri is related to diastolic blood pressure,” *Neurology*, vol. 51, no. 1, pp. 319–320, 1998.
- [17] W. Van der Flier, P. Scheltens, and J. de Groot, “White matter hyperintensities on magnetic resonance imaging and cognitive function in healthy elderly people: a systematic review,” *Journal of Neurology*, vol. 256, no. 5, pp. 717–725, 2009.
- [18] B. Dijkstra, H.-H. Kan, M. Smits, J. Van der Grond, and M. Viergever, “Automatic segmentation of white matter hyperintensities in multi-channel mri,” *IEEE transactions on medical imaging*, vol. 35, no. 9, pp. 2199–2210, 2016.
- [19] H. J. Kuijf, J. M. Biesbroek, J. De Bresser, R. Heinen, S. Andermatt, M. Bento, M. Berseth, M. Belyaev, M. J. Cardoso, A. Casamitjana, et al., “Standardized assessment of automatic segmentation of white matter hyperintensities and results of the wmh segmentation challenge,” *IEEE transactions on medical imaging*, vol. 38, no. 11, pp. 2556–2568, 2019.
- [20] H. Li, G. Jiang, J. Zhang, R. Wang, Z. Wang, W.-S. Zheng, and B. Menze, “Fully convolutional network ensembles for white matter hyperintensities segmentation in mr images,” *NeuroImage*, vol. 183, pp. 650–665, 2018.
- [21] C. A. Humphreys, C. Smith, and J. M. Wardlaw, “Correlations in post-mortem imaging-histopathology studies of sporadic human cerebral small vessel disease: a systematic review,” *Neuropathology and applied neurobiology*, vol. 47, no. 7, pp. 910–930, 2021.
- [22] F. De Guio, J.-F. Mangin, M. Duering, S. Ropele, H. Chabriat, and E. Jouvent, “White matter edema at the early stage of cerebral autosomal-dominant arteriopathy with subcortical infarcts and leukoencephalopathy,” *Stroke*, vol. 46, no. 1, pp. 258–261, 2015.
- [23] N. D. Prins and P. Scheltens, “White matter hyperintensities, cognitive impairment and dementia: an update,” *Nature Reviews Neurology*, vol. 11, no. 3, pp. 157–165, 2015.
- [24] Q. Vanderbecq, E. Xu, S. Ströer, B. Couvy-Duchesne, M. D. Melo, D. Dormont, O. Colliot, A. D. N. Initiative, et al., “Comparison and validation of seven white matter hyperintensities segmentation software in elderly patients,” *NeuroImage: Clinical*, vol. 27, p. 102357, 2020.
- [25] G. Solé-Guardia, E. Custers, A. de Lange, E. Clijncke, B. Geenen, J. Gutierrez, B. Küsters, J. A. Claassen, F.-E. de Leeuw, M. Wiesmann, et al., “Association between hypertension and neurovascular

- inflammation in both normal-appearing white matter and white matter hyperintensities,” *Acta neuropathologica communications*, vol. 11, no. 1, pp. 1–12, 2023.
- [26] C. D. Smith, D. Snowdon, and W. R. Markesberry, “Periventricular white matter hyperintensities on mri: correlation with neuropathologic findings,” *Journal of Neuroimaging*, vol. 10, no. 1, pp. 13–16, 2000.
- [27] S. J. Van Veluw, J. J. Zwanenburg, J. Engelen-Lee, W. G. Spliet, J. Hendrikse, P. R. Luijten, and G. J. Biessels, “In vivo detection of cerebral cortical microinfarcts with high-resolution 7t mri,” *Journal of Cerebral Blood Flow & Metabolism*, vol. 33, no. 3, pp. 322–329, 2013.
- [28] S. J. Van Veluw, J. J. Zwanenburg, A. J. Rozemuller, P. R. Luijten, W. G. Spliet, and G. J. Biessels, “The spectrum of mr detectable cortical microinfarcts: a classification study with 7-tesla postmortem mri and histopathology,” *Journal of Cerebral Blood Flow & Metabolism*, vol. 35, no. 4, pp. 676–683, 2015.
- [29] L. Bronge, N. Bogdanovic, and L.-O. Wahlund, “Postmortem mri and histopathology of white matter changes in alzheimer brains,” *Dementia and geriatric cognitive disorders*, vol. 13, no. 4, pp. 205–212, 2002.
- [30] S. T. Grafton, S. Sumi, G. K. Stimac, E. C. Alvord, C.-M. Shaw, and D. Nochlin, “Comparison of postmortem magnetic resonance imaging and neuropathologic findings in the cerebral white matter,” *Archives of neurology*, vol. 48, no. 3, pp. 293–298, 1991.
- [31] T. Revesz, C. Hawkins, E. Du Boulay, R. Barnard, and W. McDonald, “Pathological findings correlated with magnetic resonance imaging in subcortical arteriosclerotic encephalopathy (binswanger’s disease).,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 52, no. 12, pp. 1337–1344, 1989.
- [32] M. Scarpelli, U. Salvolini, L. Diamanti, R. Montironi, L. Chiaromoni, and M. Maricotti, “Mri and pathological examination of post-mortem brains: the problem of white matter high signal areas,” *Neuroradiology*, vol. 36, no. 5, pp. 393–398, 1994.
- [33] V. Marshall, W. Bradley Jr, C. Marshall, T. Bhoopat, and R. Rhodes, “Deep white matter infarction: correlation of mr imaging and histopathologic findings.,” *Radiology*, vol. 167, no. 2, pp. 517–522, 1988.
- [34] K. McAleese, M. Firbank, D. Hunter, L. Sun, R. Hall, J. Neal, D. Mann, M. Esiri, K. Jellinger, J. O’Brien, *et al.*, “Magnetic resonance imaging of fixed post mortem brains reliably reflects subcortical vascular pathology of frontal, parietal and occipital white matter,” *Neuropathology and Applied Neurobiology*, vol. 39, no. 5, pp. 485–497, 2013.
- [35] M. Fernando, J. O’Brien, R. Perry, P. English, G. Forster, W. McMeekin, J. Slade, A. Golkhar, F. Matthews, R. Barber, *et al.*, “Comparison of the pathology of cerebral white matter with postmortem magnetic resonance imaging (mri) in the elderly brain,” *Neuropathology and applied neurobiology*, vol. 30, no. 4, pp. 385–395, 2004.
- [36] M. S. Fernando, J. E. Simpson, F. Matthews, C. Brayne, C. E. Lewis, R. Barber, R. N. Kalaria, G. Forster, F. Esteves, S. B. Wharton, *et al.*, “White matter lesions in an unselected cohort of the elderly: molecular pathology suggests origin from chronic hypoperfusion injury,” *Stroke*, vol. 37, no. 6, pp. 1391–1398, 2006.
- [37] S. J. Van Veluw, Y. D. Reijmer, A. J. Van Der Kouwe, A. Charidimou, G. A. Riley, A. Leemans, B. J. Bacska, M. P. Frosch, A. Viswanathan, and S. M. Greenberg, “Histopathology of diffusion imaging abnormalities in cerebral amyloid angiopathy,” *Neurology*, vol. 92, no. 9, pp. e933–e943, 2019.
- [38] P. Maillard, E. Fletcher, D. Harvey, O. Carmichael, B. Reed, D. Mungas, and C. DeCarli, “White matter hyperintensity penumbra,” *Stroke*, vol. 42, no. 7, pp. 1917–1922, 2011.
- [39] S. M. Maniega, M. C. V. Hernández, J. D. Clayden, N. A. Royle, C. Murray, Z. Morris, B. S. Aribisala, A. J. Gow, J. M. Starr, M. E. Bastin, *et al.*, “White matter hyperintensities and normal-appearing white matter integrity in the aging brain,” *Neurobiology of aging*, vol. 36, no. 2, pp. 909–918, 2015.
- [40] A. G. van Norden, K. F. de Laat, R. A. Gons, I. W. van Uden, E. J. van Dijk, L. J. van Oudheusden, R. A. Esselink, B. R. Bloem, B. G. van Engelen, M. J. Zwarts, *et al.*, “Causes and consequences of cerebral small vessel disease. the run dmc study: a prospective cohort study. study rationale and protocol,” *BMC neurology*, vol. 11, no. 1, pp. 1–8, 2011.
- [41] M. Ghafoorian, N. Karssemeijer, I. W. van Uden, F.-E. de Leeuw, T. Heskes, E. Marchiori, and B. Platel, “Automated detection of white matter hyperintensities of all sizes in cerebral small vessel disease,” *Medical physics*, vol. 43, no. 12, pp. 6246–6258, 2016.
- [42] Research Imaging Institute, University of Texas Health Science Center, “MANGO (medical image analysis with grid and optimization).” <http://www.ric.uthscsa.edu/mango>.
- [43] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, “Fsl,” *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.

- [44] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, “Improved optimization for the robust and accurate linear registration and motion correction of brain images,” *Neuroimage*, vol. 17, no. 2, pp. 825–841, 2002.
- [45] S. M. Smith, “Fast robust automated brain extraction,” *Human brain mapping*, vol. 17, no. 3, pp. 143–155, 2002.
- [46] V. Fonov, A. C. Evans, K. Botteron, C. R. Almlí, R. C. McKinstry, D. L. Collins, B. D. C. Group, *et al.*, “Unbiased average age-appropriate atlases for pediatric studies,” *Neuroimage*, vol. 54, no. 1, pp. 313–327, 2011.
- [47] J. L. Andersson, M. Jenkinson, S. Smith, *et al.*, “Non-linear registration, aka spatial normalisation fmrib technical report tr07ja2,” *FMRIB Analysis Group of the University of Oxford*, vol. 2, no. 1, p. e21, 2007.
- [48] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, *et al.*, “Fiji: an open-source platform for biological-image analysis,” *Nature methods*, vol. 9, no. 7, pp. 676–682, 2012.
- [49] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, “Albumentations: Fast and flexible image augmentations,” *Information*, vol. 11, no. 2, 2020.
- [50] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015.
- [51] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18, pp. 234–241, Springer, 2015.
- [52] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [53] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *International conference on machine learning*, pp. 448–456, pmlr, 2015.
- [54] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in *2016 fourth international conference on 3D vision (3DV)*, pp. 565–571, Ieee, 2016.
- [55] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [56] D. Opitz and R. Maclin, “Popular ensemble methods: An empirical study,” *Journal of artificial intelligence research*, vol. 11, pp. 169–198, 1999.
- [57] M. Ghafoorian, N. Karssemeijer, T. Heskes, I. W. van Uden, C. I. Sanchez, G. Litjens, F.-E. de Leeuw, B. van Ginneken, E. Marchiori, and B. Platel, “Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities,” *Scientific Reports*, vol. 7, no. 1, pp. 1–12, 2017.