Contents

List of Figures

List of Tables

Abbreviations

GM	Grey matter
WM	White matter
CSF	Cerebrospinal fluid
PD	Proton density
FLAIR	Fluid attenuation inversion recovery
WML	White matter lesion
WMH	White matter hyperintensity
DAWM	Dirty appearing WM
MS	Multiple Sclerosis
AD	Alzheimer's Disease
PVE	Partial volume effect
VLR	Voxel-Wise Logistic Regression
MLE	Maximum likelihood estimation
MAP	Maximum a posteriori
SI	Similarity index
ICC	Interclass Correlation Coefficient
LL	Lesion load
CV	Cross validation
LOO-CV	Leave-one-out CV
KF-CV	K-fold CV
LOSO-CV	Leave-one-source-out
SVM	Support vector machine
K-NN	K-nearest neighbours
MRF	Markov random field
FPR	False positive reduction

Notation

Variables		
y	feature	$\in \mathbb{R}$
$\tilde{y}$	standardized feature	$\in \mathbb{R}$
c	true class	$\in \{0,1\}$
$\hat{c}$	estimated class	$\in [0,1]$
β	logistic model feature weight	$\in \mathbb{R}$

Indexing – e.g. arbitrary variable a		
k	feature index	$\in \{1, \ldots, K\}$
n	subject index	$\in \{1, \ldots, N\}$
t	iteration index	$\in \{1, \ldots, T\}$
x	spatial location	$= [\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3]$
$a_n^{k\,(t)}(x)$	k^{th} feature; n^{th} subject; t^{th} iteration; location x	

Images & sets – e.g. arbitrary variable a	
a	one voxel, one feature, one subject
$\boldsymbol{a}$	one voxel, all features, one subject
$\mathrm{A}(x)$	image in native space
$A(x)$	image in standard space (MNI)
$\boldsymbol{A}(x)$	image set: all features, one subject
$\mathcal{A}(x)$	image set: one feature, all subjects
$\boldsymbol{\mathcal{A}}(x)$	image superset: all features, all subjects
$\mathbb{A}$	full dataset: all features, subjects, voxels

Chapter 1

Introduction

Digitization of medical imaging has facilitated innumerable advances in disease understanding and treatment. From multi-modal image fusion to image guided therapy, software tools now underpin research and clinical workflows in almost every domain of medical imaging.

This work concerns an unsolved segmentation problem in 3D brain magnetic resonance imaging (MRI), in which the objective is to automatically predict the class, or label, of every voxel ("volume pixel") in the image. The objects of interest are white matter hyperintensities (WMH), non-cancerous brain lesions which are correlated with several neurodegenerative diseases. This chapter presents the motivation for automated WMH segmentation, gives a problem definition, and explores the previously proposed solutions.

1.1   Background

The brain is composed of three major classes of tissue: grey matter (GM), white matter (GM), and cerebrospinal fluid (CSF). Grey matter constitutes the peripheral surface of the brain – the cortex, approximately 5mm thick – as well as some deeper structures called the basal ganglia. It contains neuronal cell bodies, and performs the bulk of neural processing. The white matter is composed primarily of myelinated axons, and functions to relay information between different GM structures in the brain. The brain is surrounded by CSF, which provides mechanical and immunological defence. It is produced by the choroid plexuses in the ventricles of the brain – a series of 4 connected cavities.

1.1.1 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) provides superior and flexible brain tissue contrast versus computed tomography (CT) imaging, and is the primary modality for imaging brain disease. Whereas CT measures tissue density via attenuation of transmitted X-rays, which does not vary significantly among brain tissues, MRI measures a mutable combination of 3 tissue characteristics: the proton density (PD)[1], and T1 and T2 relaxation constants [1]. The physics of signal generation are described below.

In an MR scanner, a powerful magnetic field induces alignment of proton dipoles with the field. Only a tiny fraction of the total protons align, but they create a small magnetic field M_z which is distinct from the main field [2]. The aligned protons also rotate about the axis of alignment, imperfectly, like a spinning top; this is called precession, and the frequency of rotation is roughly homogeneous and proportional to the main field strength [2]. If a second magnetic field is applied which is $90°$ perpendicular to the first, and rotating at the precession frequency, the aligned protons can be forced into temporary alignment with this transverse rotating field, before decaying back towards their original state, as illustrated in Figure 1.1 [2]. This transient applied magnetic field is induced by a radio frequency (RF) pulse, and the rate at which the original magnetization M_z is regained is described by the tissue-specific T1 relaxation constant,

$$M_z = M_0 \left(1 - e^{-\left(\frac{t}{T1} \right)} \right). \tag{1.1}$$

The T1 constant is dictated by the ability of protons in the tissue to transfer energy to bonded atoms and surrounding molecules, since this energy transfer defines the transition from the high energy transverse state to the low energy original state [2, 3]. Large macromolecules, membranes, and lipids are generally able to facilitate this energy transfer more effectively than small molecules like water, producing a shorter

---

[1] MRI can be used to image any nucleus with a net nuclear dipole, but proton (hydrogen) imaging is most common since hydrogen is biologically abundant and gives a strong signal intensity.
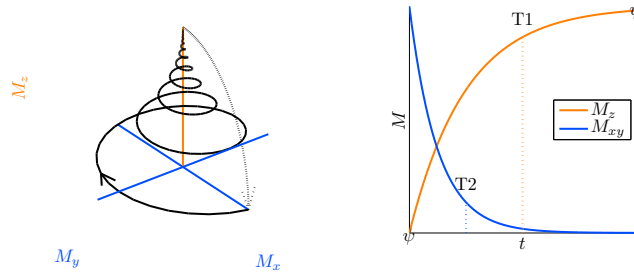
Figure 1.1: Visualization of T1 and T2 relaxation.

Table 1.1: T1 and T2 constants for brain tissues at 1.5 Tesla.

Tissue	T1 (ms)	T2 (ms)	$K[H]$ (a.u.)	Ref
WM	719 ± 33	73 ± 6	0.81 ± 0.03	[6]
GM	1165 ± 88	92 ± 11	0.98 ± 0.07	[6]
CSF	3337 ± 111	2562 ± 123	1.00 ± 0.07	[6]
WML	1124 ± 372	136 ± 79	$-$	[7][a]

[a] Estimated from Fig 1 supratentorial data (numerical results not given); $\pm$ IQR, not SD; cf. § 1.1.2 for definition.

T1 [4]. For this reason, myelinated WM has a shorter T1 than GM, which in turn has a shorter T1 than CSF, which is mostly water [5].

The rate of decay of the transverse moment M_{xy} is actually not equal to the rate of regeneration of M_z. Rather, this is governed by the T2 relaxation constant,

$$M_{xy} = M_0 \left(e^{-\left(\frac{t}{T2} \right)} \right), \tag{1.2}$$

which is always shorter that T1. This is because, in addition to T1 effects, the net rotating moment M_{xy} is eroded by proton dephasing. When precessing protons, having a net dipole, interact with other dipoles or charged particles, their rotational frequency can be increased or decreased, but overall less coherent, reducing the perceptible net magnetization M_{xy} [2]. In highly structured tissues like GM and WM, these interactions are more variable, dephasing is faster, and T2 is shorter [5]. In fluid environments like CSF, proton interactions are more homogeneous, yielding longer T2 [5]. For this reason, T2-weighted images are especially useful in identifying pathologies which degrade tissue structure, since they will have abnormally high T2 [5]. Both relaxation constants depend in a small way on the main magnetic field strength, measured in Tesla (T); T1 and T2 values for various brain tissues at 1.5T are summarized in Table 1.1.

Image acquisition involves sensing the transverse magnetization M_{xy} following proton excitation by an RF pulse. The problem is that this small signal decays very quickly due to proton dephasing, which occurs even faster than $T2$ would predict due to a third factor, inhomogeneity in the main magnetic field [8]. The time constant for this decays is termed $T2^*$, and its effects are usually undesirable [8]. As a result, M_{xy} is easily overpowered by the magnetic moment from the RF pulse, even after it is turned off, due to resonance. An important solution to this, called the spin-echo, was proposed by Erwin Hahn in 1950 [9]. If $T2^*$ for each proton is assumed to be constant, then reversing the direction of rotation at a time t should cause all protons to align again at exactly 2t. Therefore, at 2t the transverse magnetization M_{xy} – the image signal – manifests again for sensing, no longer confounded by RF coil resonance [9].
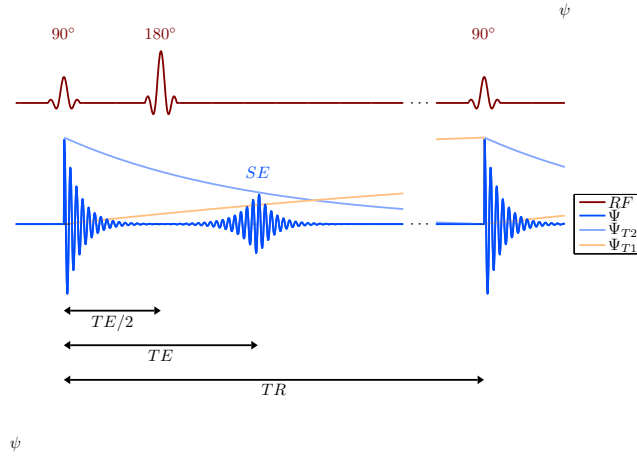
Figure 1.2: RF and MR signal for a basic Spin Echo sequence.

This second signal is called the Spin Echo (SE), and the interval $2t$ is termed the echo time (TE). Reversing the direction of rotation can be achieved by a $180°$ RF pulse at time $TE/2$, in the same way the original excitation is achieved using a $90°$ RF pulse (amount of rotation is proportional to the energy of the pulse). Acquisition of an entire image requires repetitions of this sequence with an interval called the repetition time (TR). An example spin echo sequence, showing TE and TR, as well as $T1$ and $T2$ decay, is illustrated in Figure 1.2. Spatial encoding for creation of 2D and 3D images requires the use of additional electromagnetic gradients; however this topic is omitted here since it is quite involved, and not essential to the current work[2].

Using these principals, the nature of MR image contrast can finally be understood. That is, the signal intensity Ψ for a spin echo sequence at location x can be described by the following 3-term equation,

$$\Psi_{SE}(x) = \left[K[H](x)\right]\left[e^{-\left(\frac{TE}{T2(x)}\right)}\right]\left[1 - e^{-\left(\frac{TR}{T1(x)}\right)}\right], \tag{1.3}$$

where K is scaling factor, and $[H]$ denotes the proton density. If TR is chosen to be relatively long, then the longitudinal magnetization M_z is allowed to recover completely after each repetition, the third term tends towards 1 for all tissues, and differences in tissue specific $T1$ are nullified. Similarly, if TE is relatively short, then M_{xy} has little time to dephase, the second term is maintained close to 1, and differences in $T2$ are nullified. In order to emphasize differences in $T1$, therefore, TR can be chosen shorter; for $T2$-weighted contrast, TE can be chosen longer; and if differences in $[H]$ (proton density, PD) are to be emphasized, TR can be kept long and TE short. An example MRI slice using each of these image sequences is shown in Figure 1.3a, 1.3b, and 1.3c.

---

[2] The interested reader is directed to this comprehensive resource on the topic: http://mri-q.com/

(a) T1   (b) T2   (c) PD   (d) FLAIR
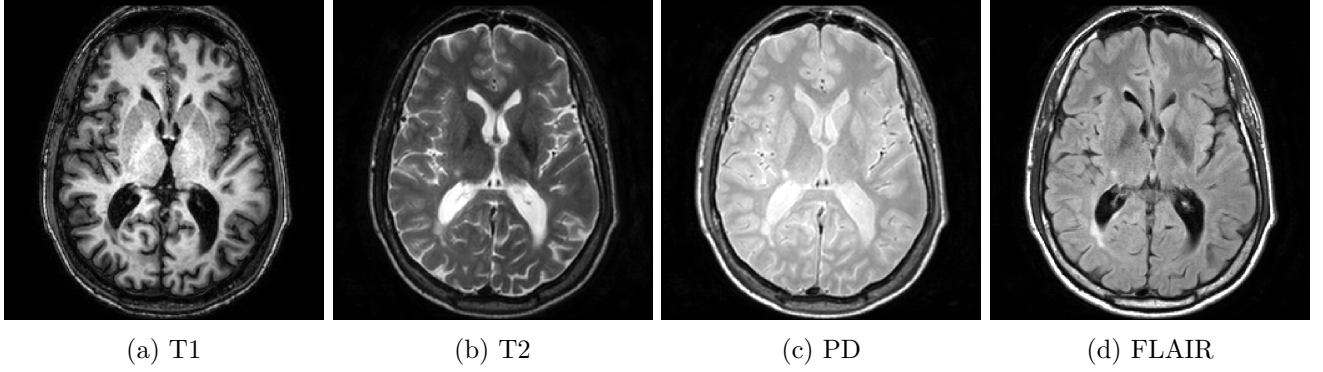
Figure 1.3: Example MRI image set with WMH pathology; from [12].

For identifying WML, T2-weighted images were conventionally used, since the lesions appear bright. However, CSF in the sulci and ventricles also appears bright on T2 images, making delineation of lesions – especially periventricular ones – difficult in T2 images (Figure 1.3b). To solve this problem, an adaptation of the spin echo RF pulse sequence can be used, called an inversion recovery (IR) [10]. In this sequence, an additional 180° inverting RF pulse is added before the 90° pulse, so that the longitudinal magnetization M_z is inverted, then recovers to the original state, passing for a brief moment through zero net magnetization. The rate of recovery is governed by $T1$, so it is tissue specific. Furthermore, if the 90° pulse is applied at the instant of zero net magnetization, no transverse moment will develop, nor the subsequent spin echo. Therefore this time interval, called the inversion time (TI), can be chosen to null the signal from any tissue with a unique $T1$. The equation governing the image signal simply adds an inversion term,

$$\Psi_{IR}(x) = \left[K\left[H(x) \right] \right] \left[e^{-\left(\frac{TE}{T2(x)} \right)} \right] \left[1 + e^{-\left(\frac{TR}{T1(x)} \right)} - 2e^{-\left(\frac{TI}{T1(x)} \right)} \right]. \tag{1.4}$$

This inversion principal is now often used to null the signal from CSF, especially for delineation of WMH, in a sequence called FLuid Attenuation Inversion Recovery (FLAIR) [11]. FLAIR images are usually T2-weighted. Figure 1.3d shows an example FLAIR image, where a WMH can be seen, posterior to the occipital horn of the left lateral ventricle, much more clearly than in the T2 image.

1.1.2 White Matter Disease

"Normal" ageing of the brain is characterized by a variety of physical and cognitive changes. Memory, synaptic plasticity, and brain volume decline, with observable effects on cognitive function [13, 14]. Brain ageing is also expedited in many patients by neurodegenerative diseases targeting the white matter, including Alzheimer's disease (AD), cerebrovascular disease, and in rare cases Multiple Sclerosis (MS). While the etiologies of these diseases are not yet fully understood, there is considerable evidence to suggest

that the they are intertwined [15, 16, 17, 18].

Cerebrovascular disease describes changes to blood vessels in the brain which increase the risk of ischemic injury – a reduction in blood flow due to vessel occlusion or hemorrhage. Ischemic injuries include major events (stroke) [19], transient ischemic attacks [20], and chronic hypoperfusion due to small vessel disease [21]. In all such events, neuronal death occurs from insufficient nutrient supply [19]. Strokes involving major cerebral arteries can be fatal, and post-event quality of life in survivors is highly variable [22]. In the less dramatic courses, clinically quiet disease progression can lead to personality changes, memory loss, and reduced cognitive ability; such changes are termed vascular dementia [23].

Alzheimer's Disease is another subclass of dementia with similar symptoms; in fact it is the most common type, affecting about 6% of the population over age 65 [24]. The cause of Alzheimer's disease is hotly debated. Two 30-year-old theories linking the disease to the build up of amyloid β protein and misfolded protein τ have been widely supported by correlational studies [25, 26, 27], but have lacked clear mechanisms of injury until recently. It is now thought that amyloid β oligomers interfere with neuronal mitochondria and synapse function, leading to cell death [28, 29], while aberrant τ proteins disrupt microtubules necessary for intraneuronal transport [27]. During the search for these mechanisms, competing theories implicating vascular injury [18], immune response [17], and blood brain barrier disruption [30] have emerged, painting the picture of a more complex disease.

The pathophysiology of multiple sclerosis is similarly unclear, though genetics are a necessary factor, and it is known that symptoms arise from erosion of myelin – a fatty insulating layer surrounding axons which is critical for normal neuron firing [31]. Several theories hypothesize either that this damage is driven by autoimmune attack, followed by neuronal dysfunction and death, or that neurodegenerative changes stimulate recruitment of immune cells as part of the usual response to injury [32, 31]. Recent evidence favours the former mechanism, particularly with inflammatory injury as the initiating event [33, 34].

Connecting all these diseases are white matter lesions (WML, AKA Leukoariosis), which represent the macroscopic changes to brain tissue in regions of white matter damage [15, 35, 36]. WML are very common in elderly populations, and a small volume of lesion does not necessarily implicate one of the above diseases; in one study of 1077 subjects aged 60-90, 95% had at least one WML [37]. WML appear as bright tissue regions in T2-weighted MRI due to some combination of inflammatory injury and degradation of tissue structure [35, 36]; in this imaging context, WML are often called white matter hyperintensities (WMH). Lesions are often focal, as opposed to diffuse, but there is evidence to suggest that surrounding regions of moderate hyperintensity, sometimes called "dirty appearing white matter (DAWM)", are also related to the diseases [38]. As biomarkers of the most common WM diseases – conditions with unsolved

etiologies and inadequate treatments – WML are of special interest to many brain researchers. The next section discusses how they are used.

1.1.3 MRI in White Matter Disease

MR imaging plays important roles in diagnosis and research of white matter diseases. Typical MRI protocols include T1, T2, and FLAIR sequences, though only the latter two sequences depict WML as hyperintense [39, 40]. Depending on the disease and context, WMH can be quantified in several ways, including binary criteria (e.g. is there a lesion in a specific location) [41, 23], rating scales (e.g. a summary of several criteria) [42], or explicit manual segmentation of the lesions by an expert [43].

WMH are arguably most important in MS. Particularly since WMH are more specific to this disease in younger patients, WMH have long been used in the diagnosis of MS, and can even be used to replace some clinical criteria, as in the 2010 McDonald Criteria [41]. MRI can also be used to discriminate between MS subtypes, which stratify disease aggression and course [41, 44, 45]. In numerous clinical trials, WMH have also been used as biomarkers of treatment efficacy [46, 47, 48], since WMH have been shown to be more sensitive to disease progression than clinical features in certain subtypes [49]. In fact, despite the central role of MRI in management and research of MS, there exists a so-called "clinico-radiological" paradox, which is the surprisingly limited correlation between WMH and clinical MS symptoms like physical and cognitive impairment [50]. However, this only strengthens the case for continued WMH research, particularly considering the recommendations by Mollison et al. in [50] to standardize image analysis in order to better understand the paradox.

In dementia (including vascular and AD), WMH are used to discriminate between disease subtypes during diagnosis. For example, the presence of at least one WML was deemed *necessary* for diagnosis of vascular dementia in 1993 [23], and subsequent revisions to these widely used criteria (NINCDS-ADRDA) have added this feature as an *exclusionary* criteria for AD [51]. While diagnosis of additional dementia subtypes may be improved using imaging [52, 53], diagnosis of the most prevalent – AD – continues to be based on clinical features alone [54]. As a result, WMH have not been used as an endpoint to any AD clinical trial. In fact, only recently have specific standards for use of WMH in vascular dementia studies been outlined [40, 36], with some subsequent uptake [55]. And yet, a 2010 meta-analysis found that WMH in brain MRI were independently correlated with stroke risk, dementia (including AD) and death [15], suggesting that much more can be done to make use of WMH as hallmarks of neurodegenerative disease.

1.2 Problem Statement

White matter hyperintensities, as ubiquitous biomarkers of several diseases with unsolved pathophysiology, are of great interest to brain researchers. Segmentation of WMH, compared to visual rating scales, provides a finer resolution for quantification of lesion load, and gives the explicit spatial distribution of pathology. This spatial information can be very useful, since diagnostic criteria often consider lesion location [52] and there are several correlations between lesion location and suspected etiology of WMH [56, 36].

Unfortunately, manual segmentation of WMH is laborious, and subject to large inter- and intra-rater variability, as reported in several works. Table 1.2 summarizes these reports, where similarity index (SI $\in [0, 1]$) is a measure of voxel-wise agreement, and interclass correlation coefficient (ICC $\in [0, 1]$) measures total volume agreement (cf. § 2.7 for definitions). Table 1.2 also gives the results using four semi-automated approaches, since these methods are reported to reduce variability and task time over strictly manual segmentation. Yet, for very large scale research studies, any approach requiring human intervention would be prohibitively time consuming and subjective.

Therefore, a fully automated algorithm to segment WMH in MRI is required. Such an algorithm would have, by construction, perfect repeatability, and consistent bias – which is especially important for perceiving small changes in longitudinal studies [57]. Additionally, while an automated approach may not necessarily be faster than manual or semi-automatic segmentation on a per-case basis, it could be run on several computers in parallel, yielding significant overall speed up.

Furthermore, while T1, T2, and FLAIR sequences are typically recommended for both MS and dementia investigations [39, 40, 45], FLAIR sequences are at least as sensitive as T2 images for the detection of WML[3]. As noted above, FLAIR images also have the advantage of easily distinguishing WMH from confounding CSF hyperintensity, which is important for highly prevalent periventricular lesions, and also for excluding lacunar infarcts [60, 61]. Consequently, it should be feasible to detect WMH using FLAIR MRI alone. This has several advantages, including minimizing the required MR sequences available during retrospective analyses, decreasing cost and scan time in prospective studies, and eliminating the need for intra-subject image registration if sequences are acquired at different resolutions (as is often the case).

---

[3] Early studies exploring the utility of FLAIR sequences may contradict this claim [58, 59], but FLAIR imaging has since improved [36].

Table 1.2: Mean inter-rater agreement measures for manual and semi-automated WMH segmentation reported in previous works.

	Ref	Raters	Data	SI	ICC
Manual	[62]	5	10 images	0.64	—
	[63]	2	6 images	0.75	—
	[64]	2	120 slices	0.83	0.96
	[43]	3	50 images	0.66	0.97
Semi-Automated	[65]	1	16 images	—	0.99
	[66]	1	2 images	0.70	—
	[67]	1	33 slices	0.78	—
	[68]	2	30 images	0.78	—

1.2.1 Objective

The primary objective of this thesis is to develop an algorithm for fully automatic segmentation of WMH, using FLAIR MRI alone. Secondary objectives include:

- analysis of the limitations of prior work in this area;

- exploration and definition of appropriate cross validation techniques for the task;

- validation of the proposed algorithm on a large and heterogeneous database of FLAIR images.

1.2.2 Challenges to Automatic Segmentation

While fully automated segmentation of WMH is attractive, translation of expert knowledge into algorithmic constructs is difficult, and often requires assumptions which induce sensitivity of the model to seemingly extraneous image features. Moreover, human understanding of MR acquisition physics help radiologists to distinguish WML from image artifacts. Thus, there are several challenges to automatic segmentation. These can be summarized as follows:

1. **Overlapping distributions:**

   Using the relaxation constants in Table 1.1 in conjunction with the FLAIR signal equation (1.4), the predicted intensity distributions of GM and WMH can overlap, depending on the choice of TE, TR, and TI (cf. Table 3.2 for typical values and § A.1 for modelling). As a result, the intensity of image voxels alone cannot be used to determine their class [69].

2. **Bias field:**

   The most common image artifact in MRI is due to inhomogeneity in the main magnetic field during acquisition, which is difficult to eliminate in strong electromagnets; this creates a low frequency

variation in signal intensity over the imaged volume [70]. The overall effect is that the same tissues may have different graylevels in different locations, further confounding the uniqueness of WMH graylevels [36].

3. **DAWM:**

   Most of the inter-rater disagreement in manual segmentation of WMH is arguably due to ambiguity of pathological extent at the lesion borders, where the core lesion meets so-called DAWM [38]. If human judgement of this boundary is difficult, then programmatic definitions could be expected to be similarly challenged.

4. **Partial volume effect:**

   With finite image resolution, voxels located on tissue boundaries will inevitably contain tissues of two or more tissues. This is known as partial volume effect (PVE), and the resulting signal intensity can be modelled as a linear mixture of the components [71]. Niessen et al. [72] show that inadequate modelling of PVE can result in significant errors in tissue segmentation, though the widely reported 30% figure from this work is derived from unrealistic conditions.

5. **Artifacts:**

   Due to the complexity of signal acquisition, there are several artifacts which can manifest in MR images. Artifacts which appear hyperintense in T2 images (including FLAIR) are of particular importance to the current work, since these confound bright pathologies, and must therefore be excluded using other features; the most notable artifacts include [36]:

   - CSF flow artifacts – ventricular hyperintensities resulting from movement of magnetically polarized CSF fluid during the inversion interval (cf. § 1.1.1) [73];

   - Perivascular spaces – minuscule spaces adjacent to cerebral vessels whose properties differ from ventricular and sulcal CSF, and are therefore not attenuated in FLAIR images [36];

   - Motion artifacts – artifacts which originate during frequency-domain encoding of spatial image content with subject motion, which is more common in MRI due to long acquisition times (several minutes); these typically manifest as high frequency "ringing" artifacts [74].

6. **Image variability:**

   There are a large number variable characteristics of MR images; some of these can be selected at acquisition time based on time constraints, and physician preferences, while others are immutable. "Image variability" is taken to comprise:

   - differences in image contrasts (and tissue graylevel distributions), due to selection of MRI parameters;

- differences in image resolution (voxel size);

- differences in MRI scanner, including field strength and proprietary image reconstruction;

- inter-subject anatomical variability and lesion heterogeneity.

Modelling this immense gamut of possible image characteristics (e.g. using parametric distributions or task-specific assumptions) represents perhaps the most challenging aspect to automated image analysis. Some specific impacts will be further discussed in § 1.3.3.

An optimal WMH segmentation algorithm will therefore consider and address each of these challenges.

1.3 Prior Work

This endeavour is far from original. Efforts to automate segmentation of WMH date back to 1990 [75] and the task has been the subject of several major reviews in 2012 [76, 69], 2013 [77], and 2015 [78]. The task has also been featured in four international competitions at the MICCAI (Medical Image Computing and Computer Assisted Intervention) Conference – 2008 [79], 2016 [80], and 2017 [12] – and the ISBI (International Symposium on Biomedical Imaging) Conference – 2015 [57] – in which researchers vie to produce the best segmentation algorithms (for more on competitions, cf. § 3.1.2). A summary of many of the proposed approaches is given in Table 1.3[4]. Below, the most important contributions in this area are introduced and summarized.

1.3.1 Segmentation Models & Features

Segmentation models represent a mapping from the content of an observed image to an image of labels or classes – in this case, tissues. The output class image comprises an estimated label for each observed voxel, or, in probabilistic models, the probability of each class for each voxel. As in many classification problems, models can be described as either supervised or unsupervised. Supervised models have relatively large capacity to model arbitrary mappings, but learn a mapping relevant to the current task using feedback from labelled examples (i.e. by a human). Unsupervised models, by contrast, are usually problem-specific, and leverage prior knowledge and the image features to predict the label image; they do not require labelled data for optimization, at least in principle.

Features used for segmentation can be derived from individual voxels (e.g. graylevel), groups of voxels

---

[4] A more detailed and interactive version of this table is available at www.uoguelph.ca/~jknigh04/wmlseg/table.html

Table 1.3: Summary of previous approaches to WMH segmentation with respect to image variability and reported performance (SI).

#	Ref.	Year	Authors	MRI Sequences	I	S	SI
1	[81]	2001	Van Leemput et al.	T1, T2, PD	20	1	0.51
2	[82]	2001	Jack et al.	FLAIR	39	1	
3	[83]	2002	Zijdenbos et al.	T1, T2	10	1	0.6
4	[84]	2004	Anbeek et al.	T1, T2, PD, FLAIR, IR	20	1	0.61
5	[85]	2005	Anbeek et al.	T1, T2, PD, FLAIR, IR	10	1	0.78
6	[86]	2005	Admiraal-Behloul et al.	T2, PD, FLAIR	100	1	0.75
7	[87]	2006	Lao et al.	T1, T2, PD, FLAIR	45	1	
8	[88]	2006	Wu et al.	T1, T2, PD	12	1	
9	[89]	2006	Sajja et al.	T2, PD, FLAIR	23	1	0.78
10	[62]	2006	Harmouche et al.	T1, T2, PD	10	1	0.61
11	[90]	2008	Khayati et al.	FLAIR	20	1	0.75
12	[91]	2008	Wels et al.	T1, T2, FLAIR	6	1	0.57
13	[92]	2008	Herskovits et al.	T1, T2, PD, FLAIR	42	2	0.6
14	[93]	2008	Bricq et al.	T2, FLAIR	25	2	
15	[94]	2008	Dyrby et al.	T1, T2, FLAIR	362	10	0.56
16	[95]	2008	Souplet et al.	T1, T2, FLAIR	25	2	
17	[63]	2009	Boer et al.	T1, PD, FLAIR	20	2	0.72
18	[96]	2009	García-Lorenzo et al.	T1, T2, PD	10	1	0.63
19	[97]	2009	Akselrod-Ballin et al.	T1, T2, PD, FLAIR	41	1	0.53
20	[98]	2009	Schwarz et al.	T1, T2, PD	165	2	
21	[99]	2010	Gibson et al.	T1, T2, FLAIR	18	1	0.81
22	[100]	2010	Shiee et al.	T1, FLAIR	10	1	0.63
23	[101]	2010	Scully et al.	T1, T2, FLAIR	17	1	
24	[102]	2011	García-Lorenzo et al.	T1, T2, FLAIR	10	1	0.65
25	[103]	2011	Geremia et al.	T1, T2, FLAIR	20	2	
26	[104]	2011	Smart et al.	T1, FLAIR	30	1	
27	[105]	2012	Samaille et al.	T1, FLAIR	67	6	0.72
28	[106]	2012	Khademi et al.	FLAIR	24	1	0.83
29	[107]	2012	Schmidt et al.	T1, FLAIR	53	1	0.75
30	[108]	2012	Abdullah et al.	T1, T2, FLAIR	61	3	
31	[109]	2013	Sweeney et al.	T1, T2, PD, FLAIR	111	1	0.61
32	[110]	2013	Datta et al.	T1, T2, FLAIR	90	3	
33	[64]	2013	Steenwijk et al.	T1, FLAIR	40	2	0.8
34	[111]	2014	Khademi et al.	FLAIR	25	1	0.78
35	[112]	2014	Ithapu et al.	T1, FLAIR	38	1	0.67
36	[113]	2014	Yoo et al.	FLAIR	32	2	0.76
37	[114]	2015	Harmouche et al.	T1, T2, PD, FLAIR	100	35	0.56
38	[115]	2015	Guizard et al.	T1, T2, PD, FLAIR	108	32	0.6
39	[116]	2015	Jain et al.	T1, FLAIR	20	1	0.67
40	[117]	2015	Tomas-Fernandez et al.	T1, T2, FLAIR	51	2	
41	[118]	2015	Wang et al.	T1, T2, FLAIR	70	2	0.84
42	[119]	2015	Roy et al.	FLAIR	38	3	0.56
43	[120]	2015	Brosch et al.	T1, T2, FLAIR	20	2	0.36
44	[121]	2015	Fartaria et al.	FLAIR	39	1	0.55
45	[122]	2015	Deshpande et al.	T1, T2, PD, FLAIR	52	1	0.5
46	[123]	2015	Roura et al.	T1, FLAIR	20	2	0.34
47	[124]	2016	Knight et al.	FLAIR	15	3	0.7
48	[125]	2016	Mechrez et al.	T1, T2, FLAIR	20	2	0.31
49	[126]	2016	Strumia et al.	T1, FLAIR	20	3	0.52
50	[127]	2016	Griffanti et al.	T1, FLAIR	130	2	0.76
51	[128]	2017	Valverde et al.	T1, FLAIR	33	2	
52	[129]	2017	Dadar et al.	T1, FLAIR	80	3	0.62
53	[130]	2017	Zhan et al.	T1, T2, FLAIR	50	2	0.76

Abbreviations. I: number of MR image sets used for validation; S: number of MRI scanners used for validation; SI: reported validation similarity index.

12

(e.g. local mean graylevel), the entire image (e.g. a histogram feature), spatial location (e.g. coordinates in a standardized space), or prior knowledge (e.g. class prior probability). It is often useful to imagine the space spanned by all possible values of all features; this is called the feature space. Each observed voxel, having a unique value for each feature, therefore represents a unique location in this space. The task of segmentation is therefore to divide the feature space into subspaces corresponding to each class. In probabilistic models, these subspaces are better described as distributions of each class over the features.

Previous approaches to WMH segmentation have generally employed three types of features:

- **Graylevel:** graylevels of MRI sequences, often following standardization (e.g. T1, T2, PD, FLAIR);

- **Prior:** prior tissue probability, often derived from a coregistered prior image (e.g. ICBM [131]);

- **Spatial:** spatial location, often normalized to a common space (e.g. x_1, x_2, x_3).

Additional features types are rarely used, since the combination of the above features are typically the only features employed by human raters. At least one graylevel feature is always used, since it is the only image-specific information (i.e. the evidence).

1.3.2 Proposed Methods

The specific methods proposed for WMH segmentation are now reviewed.[5]

Thresholding Techniques

Since WMH are brighter than healthy brain tissue in FLAIR images, many unsupervised works have used thresholding of FLAIR intensities as the initial lesion segmentation. For example, in the works by Jack et al. [82], Boer et al. [63], and Smart et al., [104] optimal FLAIR thresholds are empirically estimated relative to histogram statistics, though Boer et al. use only estimated GM voxels in the histogram. Gibson et al. use a conservative FLAIR threshold initially, but then classify the remaining voxels using Fuzzy C Means clustering [99]. Samaille et al. use nonlinear diffusion filtering and watershed segmentation, before classifying candidate regions based on a FLAIR image threshold. Yoo et al. estimate the optimal threshold for FLAIR images using histogram statistics, derived from a regression model primarily considering the total lesion load [113]. In works by Khademi et al., a peak in the conditional probability of edge content on graylevel is used to predict the transition between healthy tissue and lesion [111, 133, 124].

---

[5] This section copied verbatim from a paper in submission [132] – is this allowed?

Mixture Models

Most other unsupervised approaches are probabilistic models, often framed as a mixture model. The work by Van Leemput et al. [81] uses a similar framework as the early work by Ashburner et al. [134], later incorporated into the SPM "segment" tool [135], which jointly estimates Gaussian graylevel distributions for each tissue class, and also bias field, using expectation maximization. In the model by Van Leemput et al., distribution parameters are estimated using outlier-insensitive estimators, and WMH are derived from model outliers using heuristic rules. The predicted classes are also smoothed spatially using a Markov Random Field (MRF).

Similar works by Bricq et al. [93], Schmidt et al. [107], Jain et al. [116], and Roura et al. [123] use parametric mixture models to predict WMH as model outliers, and all but [123] embed the model in a MRF. Khayati et al. [90] and Subbanna et al. [136] also use MRF-constrained mixture models, but model WMHs as a Gaussian-distributed tissue class, rather than as outliers. In the works by Harmouche et al., parametric distributions are also used to model lesions, but such distributions are parameterized independently per brain region, in order to reflect lobe heterogeneity; a MRF is again used for regularization [62, 114]. Schwarz et al. again employ a Bayesian MRF model, but use lognormal distributions for WM and WMH [98]. Souplet et al. use an augmented mixture model which includes partial volume averaging classes and an outlier class to perform initial brain tissue segmentation; WMH are subsequently classified using a FLAIR intensity threshold after contrast enhancement [95]. The work by Herskovits et al. is much the same, but uses statistical information from training data to classify lesions (i.e. it is supervised) [92]. More recently, Graph-Cuts have been used in conjunction with mixture models, as in the works by García-Lorenzo et al. [96], Tomas-Fernandez et al. [117], and Strumia et al. [126].

The Lesion-TOADS method by Shiee et al. [100], a lesion-specific adaptation of the TOADS algorithm [137], presents an entirely new non-Gaussian paradigm for modelling class distributions, and incorporates topological energies in the objective function. Other proposed unsupervised methods have used clustering by Fuzzy C-Means, including the works by Admiraal-Behloul et al. [86], Gibson et al. [99], and Valverde et al. [128].

Classic Supervised Methods

Many early supervised methods used K-Nearest Neighbours (K-NN) for voxel-wise WMH classification. Anbeek et al. used a K-NN model with features derived from spatial coordinates and voxel intensities from several modalities [84, 85]. In the works by Wu et al. [88], Steenwijk et al. [64], and Fartaria et al. [121],

spatial coordinates are substituted for tissue priors as K-NN features. In the recently proposed BIANCA algorithm by Griffanti et al. [127], spatial coordinates are added back, along with some patch-based features.

Other works have also explored Support Vector Machines (SVM) for classification. The works by Lao et al. [87], Abdullah et al. [108], and Scully et al. [101] each use a selection of intensity features, neighbouring intensities, tissue priors, morphological, and texture features with an SVM classifier. Several more recent works have used decision tree-based classifiers, including Random Forest (RF) and AdaBoost. Akselrod-Ballin et al. [97] employ over 30 features for multi-scale image representation and classify voxels using RF. Both Geremia et al. [103] and Roy et al. [119] use a combination of intensity and tissue prior features to train a RF classifier, whereas Wels et al. [91] use a large number of Haar-like features to train an AdaBoost model. Ithapu et al. [112] explore the use of texton features in both SVM and RF models.

Logistic regression models have also gained popularity recently. In the OASIS model by Sweeney et al. [109], image intensities from T1, T2, PD, and FLAIR sequences are used individually, in multiplicative combination, and with Gaussian blurring as predictors for a global set of logistic regression parameters. In the work by Zhan et al. [130], a similar logistic model is fitted using only the raw T1, T2, and FLAIR intensities, while bias correction is performed as preprocessing and spatial smoothness using MRF post processing. In the work by Dadar et al. [129], spatial and intensity features from a flexible selection of MR sequences are used to train a linear regression model, the results of which are thresholded to give the lesion prediction. Still more works have proposed other supervised models, including nonparametric Parzen classifiers [89].

Deep Learning

A number of deep learning approaches have also been proposed, though their permeation in this problem space is surprisingly limited. Both Zijdenbos et al. [83] and Dyrby et al. [94] train fully-connected voxel-wise Neural Networks with a selection of intensity, spatial, and tissue prior features to predict the lesion class. In contrast, Brosch et al. [120] construct a more modern deep convolutional model, which is capable of capturing both local and global dependencies.

External Toolboxes

Many of the proposed methods use registration, brain extraction, bias field correction, and segmentation tools available in freely available toolkits; these include the SPM[6] toolkit [89, 94, 97, 104, 107, 113, 112, 119, 128] and the FSL[7] toolkit [92, 99, 110, 64, 109, 119, 118, 127, 130], as well as bias correction by the N3/4 [138] algorithm [83, 62, 121, 115, 114, 125, 128, 129, 130].

1.3.3 Limitations

Despite over 50 proposed algorithms and several competitions, no WMH segmentation algorithm has clearly emerged the superior method, nor has any been taken up for use in the wider research community. This is contrasted with other neuroimaging tasks, where several robust tools noted above are now regularly used in analysis pipelines – e.g. N3/4 [138] for bias field correction, BET for brain extraction [139], SPM Segment [135] / FSL FAST [140] for healthy brain segmentation, SPM Coregister [135] / FSL FLIRT [141] for registration. We hypothesize two reasons for this gap in WMH segmentation tools.

First, very few of the proposed methods have been released as either open-source code or compiled applications. Researchers may not want to release source code for reasons related to intellectual property, or the additional work of ensuring robustness and writing documentation. Yet the field of deep learning illustrates how these practices can accelerate progress in the field enormously. Similarly, compiling applications for cross-platform compatibility is no small feat, though there are many examples for SPM extensions[8], as well as events by NA-MIC (National Alliance for Medical Image Computing) for development of 3D Slicer modules[9].

Second, very few of the WMH segmentation methods have been validated on large, multi-centre databases: of the 53 works reviewed (Table 1.3), less than half use more than one scanner for validation, and only 4 use more than three. As noted in § 1.2.2, there are several sources of image variability in MRI, and both supervised and unsupervised methods can be sensitive to these factors, as noted by several authors [76, 109, 129]. Therefore, while many of the proposed methods may be of use for in-house work (i.e. with images from a consistent source), there can be little confidence that they will perform as reported on data from new sources (generalization performance).

---

[6] http://www.fil.ion.ucl.ac.uk/spm/
[7] https://fsl.fmrib.ox.ac.uk/fsl/
[8] http://www.fil.ion.ucl.ac.uk/spm/ext/
[9] https://na-mic.org/wiki/Events

In supervised models, graylevel features must be standardized, since the MRI intensity scale is not consistent across scanners or scan parameters, due to the complexity of signal acquisition [142]. However, this is not an easy task. For example, Steenwijk et al. validate a supervised WMH segmentation algorithm using same-scanner training and testing for two different scanners independently (mean SI = 0.75, 0.84), after variance scaling of intensity features [64]. Yet, a follow-up experiment which saw the method trained on one scanner and tested on the other showed a precipitous drop in performance to mean SI = 0.50.

Unsupervised models also have parameters which can be inadvertently over-tuned to data from one or two sources. For example, mixture models which classify lesions as outliers often employ an outlier definition which depends on mixture model parameters (e.g. tissue graylevel mean and variance), which in turn are subject to MR slice thickness, noise level, and contrast [81, 95, 102, 123]. Graylevel thresholding techniques [82, 104, 105, 107, 111] are similarly affected by changes in image properties.

It is worth noting three works[10] which run counter to this trend, demonstrating robust validation of their proposed methods. These works are summarized in Table 1.4. Perhaps not surprisingly, these papers report lower performance (mean SI ≤ 0.60) than other works; as a result, these algorithms would not likely be used for clinical research. Yet, even these works do not optimally estimate the model generalization performance for data from new scanners, as will be discussed in § 3.1. Moreover, the proposed algorithms in these papers require at least three MRI sequences, which fails the objectives of the current work.

Table 1.4: Works demonstrating excellent validation of a WMH segmentation algorithm.

Ref.	Year	Authors	I	S	SI
[94]	2008	Dyrby et al.	362	10	0.56
[115]	2015	Guizard et al.	108	32	0.60
[114]	2015	Harmouche et al.	100	35	0.56

Abbreviations. I: number of MR image sets used for validation; S: number of MRI scanner-parameter combinations used for validation; SI: reported validation similarity index.

1.4 Contributions

This thesis therefore aims to produce a WMH segmentation algorithm which can be used on FLAIR MRI from any source, and to characterize the expected performance on unseen data. The major contributions are as follows:

---

[10] The work by Samaille et al. (2012) [105] is also a good candidate, having used 6 scanners for validation; however, 43 of the 67 images (64%) come only from one scanner, reducing the robustness of generalization results.

1. A review and critique of the previously proposed WMH segmentation algorithms, especially with respect to expected performance on unseen data;

2. Voxel-Wise Logistic Regression (VLR): a new FLAIR-only WMH segmentation algorithm;

3. Leave-One-Source-Out Cross Validation (LOSO-CV): a validation framework which accurately characterizes the generalization performance of medical image analysis methods;

4. Extensive validation of the proposed method and its components.

The remainder of this thesis is organized as follows: Chapter 2 motivates and develops the voxel-wise logistic regression model, including expected challenges and solutions with this approach; Chapter 3 explores optimization of model components through experiment, and then presents segmentation performance results under various cross validation schemes; Chapter 4 discusses the performance results, summarizes the contributions, and highlights avenues of future work.

Chapter 2

Methodology

In this section we motivate and introduce the proposed model, and explore its parametrization. Wherever possible, concepts related to segmentation and classification will be described in their most general sense, since many of the these may also apply to other image analysis and machine learning tasks.

2.1 Model Design

As suggested in § 1.3.2, three general paradigms have emerged for segmentation of brain MRI. The first and most popular is the pipeline, in which sub-tasks are completed in sequence, such as: pre-processing, classification, post-processing. This approach permits a flexible model definition which can incorporate existing algorithms for individual steps. Most thresholding and classic supervised models fall under this category. The main drawback of this approach is that some steps could be improved by the results of downstream steps. For example, tissue classifying modules typically assume that the bias field is already corrected, but bias field estimation can be more accurate if the tissue segmentation is known.

The second paradigm, a unified generative model, aims to solve this chicken-egg problem. The segmentation is parameterized in one integrated probabilistic model, which often combines the input images with tissue prior probability images, a bias field model, and smoothness terms. Parameters of each sub-model are estimated using several expectation maximization iterations before the final segmentation is inferred. The challenges to this approach include balancing model complexity with estimability, robust convergence issues, and reduced ability to include external tools.

The final paradigm, deep learning, uses error back-propagation to update thousands of parameters in a large, relatively unstructured model mapping the input MRI to the output segmentation image. Using

so-called "end-to-end" training, there is no guarantee that the usual sub-components (e.g. bias field, tissue graylevel distributions, etc.) will be estimated, though such elements could be expected to develop in the internal relationships if they are relevant to the task at hand. Most deep learning models still require standardization in space and graylevel, and more importantly, large training datasets – which are rare in medical imaging.

For several reasons, the supervised pipeline approach was selected for this work. First, preliminary work drawing on existing algorithms [111, 143] showed the feasibility of several relatively simple thresholding methods, which confer robustness through simplicity. Second, the parametric assumptions required by unified probabilistic models may be challenged by data from multiple sources [81]. Third, only a small number of training cases were available during initial development, which limited the feasibility of deep learning approaches. Finally, time constraints favoured the incorporation of existing tools to address challenges like bias correction and image registration, which would have been otherwise difficult to develop in a unified model.

2.1.1   Features

Many FLAIR-only WMH segmentation algorithms use a thresholding technique: map a single graylevel feature directly to a class or class probability (e.g. "healthy" or "lesion"). Such models often have complex methods of deriving this mapping (e.g. using mixture model parameters [123], histogram features [113], or conditional edge probability [144]), but the final rule is applied equally to the entire graylevel image, usually the FLAIR sequence. The most common challenge for these methods is a high number of false positives (c.f. § 2.7 for definitions), since several artifacts and GM can overlap the WMH intensity distribution. As a result, nearly all previously proposed thresholding algorithms employ heuristic "false positive reduction" (FPR) rules as post-processing to remove these errors.

Preliminary investigations sought to characterize the spatial distribution of false positives, in order to explore possible FPR rules. Using a database of 96 FLAIR images in MNI space (c.f. § 3.1.3 for details), the optimal threshold for each image was calculated[1] and used to generate binary class images (0 = healthy, 1 = lesion). Comparing these results with the manual segmentations, the average distribution of true positives (TP), false positives (FP), and false negatives (FN) was computed, as shown in Figure 2.1. The median optimal similarity index (SI) was 0.36.

Evidently, even if an optimal threshold could be estimated for these data, a high proportion of FP and FN

---

[1] The `fminsearch` function in Matlab maximized the similarity index with respect to the manual segmentation.
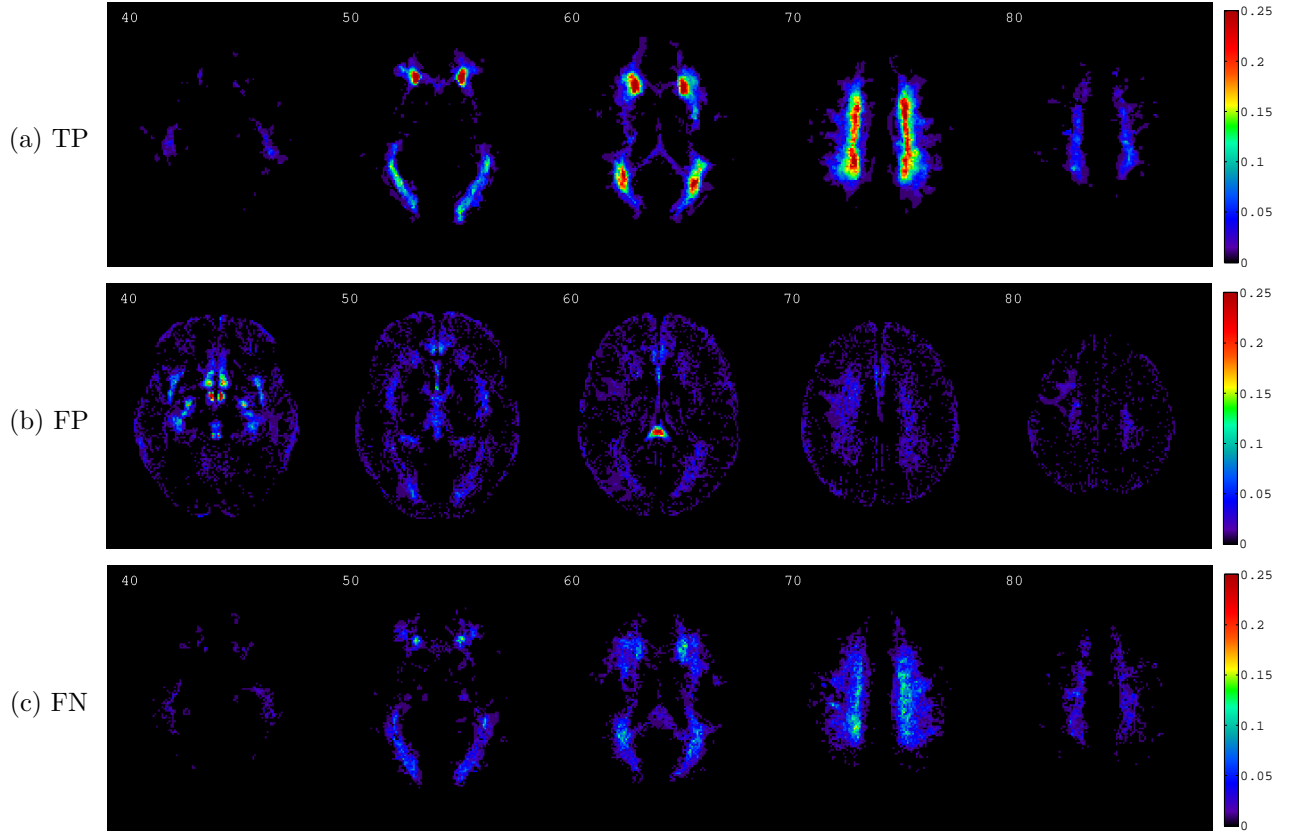
Figure 2.1: Distributions of TP, FP, and FN in 96 FLAIR MRI, following supervised optimal thresholding in MNI space.

would be incurred. Yet in Figure 2.1 there are regular and distinct spatial distributions of these errors. Moreover, is it has been suggested that there is regional heterogeneity in relaxation rates of brain tissues [145], and that WML intensity depends in part on location [7, 114]. This implies that spatial coordinates would be useful features for WMH segmentation, especially when incorporated into the main classification model (i.e. not as post-processing). Indeed such features have previously been used [84, 85, 94, 127, 129].

Spatial features could be used with FLAIR intensities in supervised classification models like K-NN, SVM, RF, etc. However, such models treat all features equally, which can lead to artifacts in the decision boundary that contradict prior knowledge. For example, in spatial locations which do not observe any lesions during training, the optimized model may learn to never predict the lesion class, regardless of graylevel features. Similarly, there is no guarantee that FLAIR graylevels will map monotonically to the lesion class, as would be expected. Therefore, spatial features should be treated differently.
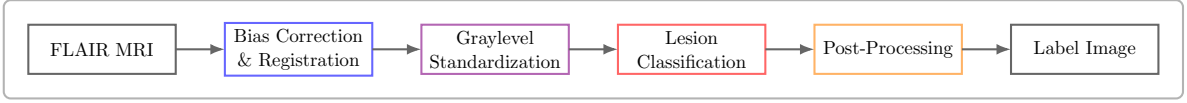
Figure 2.2: Overview of the necessary processing steps.

2.1.2 Overview

The proposed pipeline can be summarized as follows. Pre-processing steps will aim to correct any bias field effect and standardize spatial coordinates. The classification model will employ FLAIR intensities and spatial features to give the initial segmentation. Since most classification models assume features are drawn from a consistent distribution, a graylevel standardization step is appended to the pre-processing. Post-processing steps will generally aim to improve segmentation performance. This pipeline is illustrated in Figure 2.2. Next, the proposed classification model is developed.

2.2 Voxel-Wise Logistic Regression

This section presents the proposed classification model: Voxel-wise Logistic Regression (VLR). The VLR model solves the problem of differential treatment for spatial and graylevel features, and produces uniquely interpretable parameters following training. The model is inspired by the LPA algorithm in the LST toolbox for SPM by Schmidt [143, 146][2], however, novel methods of model parametrization and estimation are presented here, in addition to differences in pre- and post-processing from the LPA algorithm.

Logistic regression models have recently gained popularity for WMH segmentation [147, 109, 146, 130], and have several advantages. First, model parameters are generally more interpretable than those in other models, permitting better design of regularizations. Second, the simplicity of the model reduces its capacity for over-fitting. Third, model foundations in statistical theory allow probabilistic interpretations of the outputs, which may be helpful for quantifying marginally pathological tissues like DAWM. The major drawback of logistic models for WMH segmentation is that standardization of graylevels is required.

In the two class formulation, the lesion class is denoted $c = 1$, and the healthy class $c = 0$. The probability of the lesion class, given a set of features $\boldsymbol{y} = [1, y^1, \ldots, y^K]^T$ is modelled by a logistic function,

---

[2] This algorithm was originally understood through the Matlab code, after downloading the toolbox. Section 6.1 of [146], published in 2017, provides more details.

parameterized by a vector of feature weights $\boldsymbol{\beta} = [\beta^0, \beta^1, \ldots, \beta^{\mathrm{K}}]^T$,

$$P(c = 1 \mid \boldsymbol{y}, \boldsymbol{\beta}) = \frac{1}{1 + e^{-\eta}}, \qquad \eta = \boldsymbol{\beta}^T \boldsymbol{y}. \tag{2.1}$$

This probability – the estimated lesion class label – is denoted $\hat{c} = P(c = 1 \mid \boldsymbol{y}, \boldsymbol{\beta}) \in [0, 1]$. Considering the spatial location $x = [\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3]$, the label image is defined as $\hat{C}(x) = P(C(x) = 1 \mid \boldsymbol{Y}(x), \boldsymbol{\beta})$.

Previous works have not included spatial dimensions in the logistic feature set $\boldsymbol{y}$, since $\hat{c}$ is not expected to be monotonically related to any spatial dimension. The circumvention proposed by Schmidt is to parameterize the feature weights spatially – that is $\boldsymbol{\beta} \to \boldsymbol{\beta}(x)$,

$$\hat{C}(x) = \frac{1}{1 + e^{-\eta(x)}}, \qquad \eta = \boldsymbol{\beta}(x)^T \boldsymbol{Y}(x). \tag{2.2}$$

In this way, the characteristics of the logistic regression model are maintained with respect to graylevel features, but spatial features can have more complex relationships with the output. A second pre-processing step, however, is now required: standardization of spatial dimensions – i.e. registration of the input image to a consistent target – so that the same voxel in every input image corresponds to the same brain region. Training the VLR model then yields a unique vector $\boldsymbol{\beta}$ for each voxel, or equivalently, one complete image for each parameter. Such "parameter images", as they will be called, are then readily interpretable and can be regularized intuitively.

2.2.1   Model Fitting

Fitting the VLR model involves estimating $\boldsymbol{\beta}$ for each voxel x. This requires some training data: feature vectors from a population of N observations $\boldsymbol{\mathcal{Y}} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{\mathrm{N}}\}$, and the corresponding labels $\mathcal{C} = \{c_1, \ldots, c_{\mathrm{N}}\}$. Typically, parameter estimation involves maximizing the likelihood of the model, given this data – i.e. maximum likelihood estimation (MLE).

Three major challenges emerge during model fitting. These challenges involve contradictions between prior knowledge and the fitted model using the available training data. That is, these challenges could all be overcome by a more complete training set, but this is usually not available. The three challenges are:

1. **Separable classes:** When data from two classes are perfectly separable, the MLE-fitted logistic model can approach a step-function – i.e. $\beta^k \to +\infty$. This implies that on either side of a specific graylevel threshold, the model is either 100% confident in predicting the healthy class, or 100% confident in predicting the lesion class. In fact, no threshold is ever so perfect, and instead a level
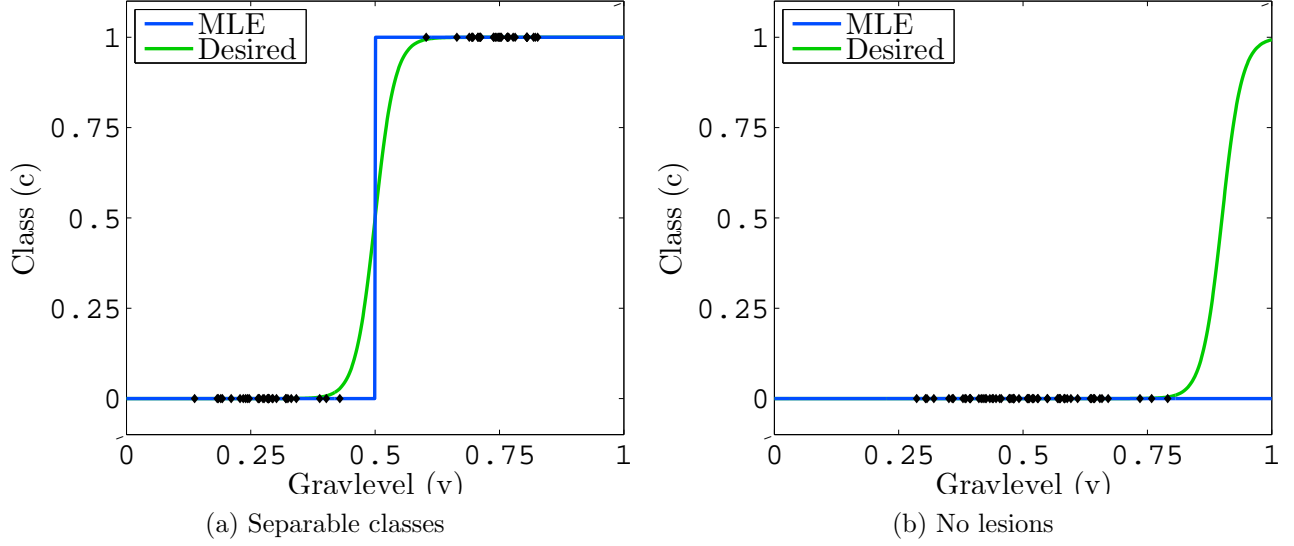
Figure 2.3: Challenges encountered during estimation of a logistic model.

of uncertainty should be maintained around the decision boundary. These two cases are illustrated in Figure 2.3a.

2. **Sparsely observed lesion class:** Since WML are often distributed in consistent locations, many brain regions contain no lesions across the entire training dataset. These voxels will be termed "healthy training" voxels. In some locations, this is expected (e.g. the GM, since by definition WMH manifest in the WM), while in others, our prior knowledge predicts lesions will eventually be observed (e.g. the rest of the WM). As illustrated in Figure 2.3b, the MLE-fitted model may not maintain the ability to predict $\hat{c} = 1$ in such locations, regardless of the features. However, the ability to predict lesion should be maintained in many of these locations.

3. **Smooth parameter images:** It is assumed that similar locations will contain similar training data, yielding smooth parameter images. If this assumption is sometimes invalid, parameter images could contain noise or discontinuities, creating artifacts in estimated lesion class images.

Joint Estimation

In the LPA algorithm, the logistic feature weights for each voxel are not considered independent. The FLAIR graylevel effect is tied across all voxels – i.e. $\beta^1(x) \rightarrow \beta^1$, so $\eta(x) = \beta^0(x) + \beta^1 Y(x)$ – and a Gaussian MRF model is used to estimate the spatially varying intercept $\beta^0(x)$. This design addresses the aforementioned challenges by greatly increasing the number of observations used to estimate each

parameter. However, it requires joint estimation of all parameters simultaneously, and dramatically increases the complexity of estimation [146].

Efficient fitting of such models was the subject of major works by Schmidt et al. [148, 146], but several drawbacks remain. First, Markov Chain Monte Carlo estimation of the model appears to create discontinuity artifacts in the spatial effect image $\beta^0(x)$, as shown in Figure 3.2. Second, MRF modelling of the parameter images assumes that the missing data (i.e. WMH training examples in the more superficial brain regions) can be interpolated spatially, but this may not be justified. Third, the proposed joint estimation procedures are computationally expensive (versus the methods proposed here), requiring approximately two hours to estimate $\boldsymbol{\beta}(x)$ [146]. Finally, it is not clear whether any tied β are necessary or advantageous in this context. These deficiencies then motivate investigations into alternative solutions to the above challenges.

In addition to these potential modelling weaknesses, there were also several limitations to the validation methodology for the LPA algorithm, worth noting here. First, the "ground truth" segmentations were generated using an automated algorithm – the Lesion Growth Algorithm (LGA) [107] of the same toolbox – rather than a human expert. Second, the graylevel standardization procedure employed does not consider the variance of image graylevels (only the mean is subtracted); this strongly assumes that user images will have graylevels spanning a similar range. Third, all 53 training cases were obtained on the same MRI scanner, which may limit generalization performance. Finally, no segmentation performance results are given in either of the associated publications [148, 146]. Therefore, while the open-source release of the LPA algorithm is greatly appreciated, significant improvements can be made to this algorithm.

Independent Estimation

An alternative approach is to consider each parameter vector $\boldsymbol{\beta}(x)$ independent. Doing so greatly simplifies model fitting, but does not address the challenges described above. These must instead be addressed using regularization strategies, as discussed below in § 2.3. In this section, ML estimation of independent parameter vectors $\boldsymbol{\beta}$ is developed. For clarity, only a single voxel is considered – i.e. y from $Y(x)$, etc.

As note above, the optimal $\boldsymbol{\beta}$ for each independent voxel can be resolved using MLE. If the training data are also assumed to be independently observed, then the likelihood (conditioned on the data) is defined

from binomial theory as

$$L(\boldsymbol{\beta} \mid \mathcal{C}, \boldsymbol{\mathcal{Y}}) = \prod_{n=1}^{N} P(c = 1 \mid \boldsymbol{y}_n, \boldsymbol{\beta})^{c_n} \left(1 - P(c = 1 \mid \boldsymbol{y}_n, \boldsymbol{\beta})^{1-c_n}\right)$$

$$= \prod_{n=1}^{N} \left[\hat{c}_n{}^{c_n} \left(1 - \hat{c}_n{}^{1-c_n}\right)\right]. \tag{2.3}$$

For computational reasons, it is simpler and asymptotically equivalent to maximize the log-likelihood,

$$\mathcal{L}(\boldsymbol{\beta}) = \log \prod_{n=1}^{N} \left[\hat{c}_n{}^{c_n} \left(1 - \hat{c}_n{}^{1-c_n}\right)\right]$$

$$= \sum_{n=1}^{N} \left[c_n \log \hat{c}_n + (1 - c_n) \log(1 - \hat{c}_n)\right]$$

$$= \sum_{n=1}^{N} \left[c_n \boldsymbol{\beta}^T \boldsymbol{y}_n - \log(1 + e^{\boldsymbol{\beta}^T \boldsymbol{y}_n})\right]. \tag{2.4}$$

The optimal $\boldsymbol{\beta}$ is therefore resolved by maximizing the log-likelihood,

$$\boldsymbol{\beta}^* = \underset{\boldsymbol{\beta}}{\arg\max} \ \mathcal{L}(\boldsymbol{\beta})$$

$$= \underset{\boldsymbol{\beta}}{\arg\max} \ \sum_{n=1}^{N} \left[c_n \boldsymbol{\beta}^T \boldsymbol{y}_n - \log(1 + e^{\boldsymbol{\beta}^T \boldsymbol{y}_n})\right] \tag{2.5}$$

Iterative Updates

Estimation of $\boldsymbol{\beta}^*$ can be performed using iterative optimization, using an initial estimate $\boldsymbol{\beta}^{(0)}$ and an update term $\Delta\boldsymbol{\beta}^{(t)}$,

$$\boldsymbol{\beta}^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t)} + \alpha \, \Delta\boldsymbol{\beta}^{(t)}, \tag{2.6}$$

where α is a small valued learning rate parameter. There are many possible definitions of $\Delta\boldsymbol{\beta}$, including simply the gradient of $\mathcal{L}(\boldsymbol{\beta})$, denoted $\nabla_{\boldsymbol{\beta}}\mathcal{L}$. However, it can be shown that $\mathcal{L}(\boldsymbol{\beta})$ is convex, so higher order update equations can be used. The work by Minka [149] compares several options, including Newton's method (and variants), conjugate gradient, iterative scaling (and variants), and dual optimization[3]. For small feature dimensionality (K), performance differences among the options were small. Classic Newton updates gave a good balance between memory requirements and computational order, so they are used.

---

[3] Matlab code available at https://github.com/tminka/logreg/

If the gradient $\nabla_{\boldsymbol{\beta}} \mathcal{L}$ and Hessian matrix $\nabla_{\boldsymbol{\beta}}^2 \mathcal{L}$ are defined as

$$\nabla_{\boldsymbol{\beta}} \mathcal{L} = \begin{bmatrix} \frac{\partial L}{\partial \beta^1} \\ \vdots \\ \frac{\partial L}{\partial \beta^{\mathrm{K}}} \end{bmatrix}, \tag{2.7}$$

$$\nabla_{\boldsymbol{\beta}}^2 \mathcal{L} = \begin{bmatrix} \frac{\partial^2 L}{\partial \beta^1 \partial \beta^1} & \cdots & \frac{\partial^2 L}{\partial \beta^1 \partial \beta^{\mathrm{K}}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 L}{\partial \beta^{\mathrm{K}} \partial \beta^1} & \cdots & \frac{\partial^2 L}{\partial \beta^{\mathrm{K}} \partial \beta^{\mathrm{K}}} \end{bmatrix}, \tag{2.8}$$

then the Newton update is given by

$$\Delta\boldsymbol{\beta} = -\nabla_{\boldsymbol{\beta}}^2 \mathcal{L}^{-1} \nabla_{\boldsymbol{\beta}} \mathcal{L}. \tag{2.9}$$

In the current model, the gradient is given by

$$\nabla_{\boldsymbol{\beta}} \mathcal{L} = \sum_{n=1}^{N} \boldsymbol{y}_n \left(c_n - \hat{c}_n \right), \tag{2.10}$$

and the Hessian by

$$\nabla_{\boldsymbol{\beta}}^2 \mathcal{L} = \sum_{n=1}^{N} \boldsymbol{y}_n {\boldsymbol{y}_n}^T \left(c_n - \hat{c}_n \right). \tag{2.11}$$

Substituting (2.10) and (2.11) into (2.9), the explicit update $\Delta\boldsymbol{\beta}$ for (2.6) is obtained. At each iteration, $\Delta\boldsymbol{\beta}^{(t)}$ is re-computed, and the process continues until some convergence criterion is satisfied. This procedure is repeated for all voxels of interest. Such voxels can be selected using a "brain mask", denoted $M(x)$ – a binary image denoting the expected location of the brain in standardized space (c.f. § B.3.2 for details about the brain mask used in this work). Since the parameters of each voxel are estimated independently, this can also be computed in parallel. The details of this implementation are presented in § B.2, after incorporation of certain regularization terms described in the next section.

2.3 Regularization

Regularizations are methods of injecting prior knowledge about the expected model into the optimization. Assuming voxel-wise independence of model parameters requires the use of regularization strategies to solve the challenges outlined in § 2.2.1. Model fitting which includes regularization is termed maximum a posteriori (MAP) estimation. Several regularization methods are explored below.

2.3.1 Data Augmentation

Noting the central role of training data in each of the challenges, methods of artificially increasing the training dataset size may be particularly useful in solving them. Data augmentation has long been used in machine learning tasks with limited training data, and there are several methods of generating synthetic data. In low dimensional input/output spaces, random sampling of fitted class-conditional posterior distributions can produce reasonable samples with known labels [150]. In higher dimensional problem spaces, however, imputation is more difficult [151]. For example, the space of potential $100 \times 100 \times 100$-sized images has 100^3 dimensions (one per voxel), yet only a small inner subspace represents plausible images. Generating synthetic examples in this space is therefore challenging, especially for segmentation tasks, where the outputs have dimensionality roughly equal to the input.

Alternatively, simple image manipulations can still afford model improvements [152]. In segmentation tasks, both the input image(s) and the corresponding label images can be translated, reflected, rotated, and perhaps resized, thereby avoiding the generation of genuinely synthetic examples. In the current work, reflections and small (one-voxel) translations can be applied to the label and FLAIR images following registration to the MNI brainspace.

2.3.2 Classic Regularization

The separable classes challenge is well-known in regression problems, and a good solution is to penalize the magnitude of model parameters using the L_p-norm: $\lambda \, ||\boldsymbol{\beta}||_p$ [153]. It can be shown that L_1 regularization corresponds to a Laplacian prior on elements of $\boldsymbol{\beta}$, with scale parameter inversely proportional to λ (equivalently, this assumes that the model error follows this distribution); similarly, L_2 regularization implies a Gaussian prior, with standard deviation inversely proportional to λ [153]. This penalty can then be appended to the objective function (2.5), as in

$$
\begin{aligned}
\boldsymbol{\beta}^* &= \arg\max_{\boldsymbol{\beta}} \; \mathcal{J}(\boldsymbol{\beta}) \\
&= \arg\max_{\boldsymbol{\beta}} \; \mathcal{L}(\boldsymbol{\beta}) - \lambda \, ||\boldsymbol{\beta}||_p \\
&= \arg\max_{\boldsymbol{\beta}} \; \sum_{n=1}^{N} \left[c_n \boldsymbol{\beta}^T \boldsymbol{y}_n - \log(1 + e^{\boldsymbol{\beta}^T \boldsymbol{y}_n}) \right] - \lambda \, ||\boldsymbol{\beta}||_p
\end{aligned}
\tag{2.12}
$$

Due to its relatively large gradient near zero, L_1 regularization is typically used to encourage sparsity in the feature weights (i.e. $\beta^k \to 0$) [154]. This is not desirable in the current model, since the feature

(FLAIR graylevel) is known to be discriminative. Moreover, the expansion of the $||\boldsymbol{\beta}||_1$ term in the gradient of the objective function is not straightforward, since it is non-differentiable at zero [154, 155]. Conversely, L_2 regularization is more effective at limiting parameter magnitude – which is the current aim – and the first and second order gradients of (2.12) derive easily [149]. For these reasons, only L_2 regularization is considered, yielding the following change to the Newton update expression (2.9),

$$
\begin{aligned}
\Delta\boldsymbol{\beta} &= -\nabla_{\boldsymbol{\beta}}^2 \mathcal{J}^{-1} \nabla_{\boldsymbol{\beta}} \mathcal{J} \\
&= -\left(\nabla_{\boldsymbol{\beta}}^2 \mathcal{L} - \lambda I\right)^{-1} \left(\nabla_{\boldsymbol{\beta}} \mathcal{L} - \lambda \boldsymbol{\beta}\right)
\end{aligned}
\tag{2.13}
$$

What remains is to select an appropriate value of λ. This is explored experimentally in § 3.3.1 using a toy model.

2.3.3   Pseudo-Lesions

The Sparsely observed lesion class challenge is less common, since discriminative models are rarely fit in the absence of one class altogether. This occurs here because all voxels are modelled independently. It is therefore tempting to simply sample features (FLAIR graylevels) from the lesion class at other spatial locations in order to fit the logistic model in the healthy training voxels. However, as noted in § 1.2.2 and § 2.1, there may be spatially-correlated heterogeneity in the appearance of lesions [145, 7], and some locations will likely never contain any lesion. Considering these facts, generating additional synthetic lesion-class samples, or "pseudo-lesions", for these locations using static, spatially parameterized definitions would have two advantages. First, this approach would not be subject to variations in the training data. Second, it would be feasible to consider the prior probability of healthy tissues (GM / WM / CSF) in the definition.

...

2.3.4   Parameter Image Smoothing

Finally,

...

While an alternative approach might involve modelling the parameter images as a spatial function (e.g. band-limited discrete cosine / Fourier transform), there are two challenges with this approach. First, deriving the update gradients for such a model would be challenging, and their computation might increase

training time. Second, such an encoding may introduce artifacts in the resulting parameter images. More-over, it is well known that frequency domain band-limiting can be equivalently achieved by convolution (i.e. filtering) in the spatial domain [156].

...

Harmouche et al. [114] also note that in smoothly varying models, small registration errors can be expected to have only a minor impact.

2.4   Bias Correction & Registration

Two objectives of preprocessing

...

2.5   Graylevel Standardization

...

2.6 Post-Processing

2.7 Performance Metrics

Segmentation performance of the model is characterized in two respects: voxel-wise agreement and total lesion load (LL) volume agreement. When comparing the estimated segmentation ($\hat{c}$) to the ground truth segmentation (c), each individual voxel can occupy one of four states (colours shown for future reference):

- ■ True Positive (TP): $c = 1$ and $\hat{c} = 1$, correctly predicted "lesion".

- ■ False Positive (FP): $c = 0$ and $\hat{c} = 1$, incorrectly predicted "lesion".

- ■ False Negative (FN): $c = 1$ and $\hat{c} = 0$, incorrectly predicted "healthy".

- ■ True Negative (TN): $c = 0$ and $\hat{c} = 0$, correctly predicted "healthy".

Summing the number of voxels in each state, voxel-wise agreement can then be quantified using the following measures:

- **Similarity Index (SI)** (AKA Dice Similarity Coefficient, F1-Score)
  Measures overall segmentation performance.

$$SI = \frac{2TP}{2TP + FP + FN} \tag{2.14}$$

- **Precision (Pr)** (AKA Overlap Fraction, Positive Predictive Value)
  Fraction of predicted predicted positives which are true positives.

$$Pr = \frac{TP}{TP + FP} \tag{2.15}$$

- **Recall (Re)** (AKA Sensitivity, True Positive Rate)
  Fraction of true positives which are predicted positive.

$$Re = \frac{TP}{TP + FN} \tag{2.16}$$

Each measure is $\in [0, 1]$, where higher is better. Note that typical performance metrics like accuracy and sensitivity are avoided, since they include the TN count in the numerator, which is typically much larger than $TP + FP + FN$ combined (i.e. $c = 1$ is a rare event).

Overall volume agreement between segmentations is characterized using the 2-way mixed-effects single-
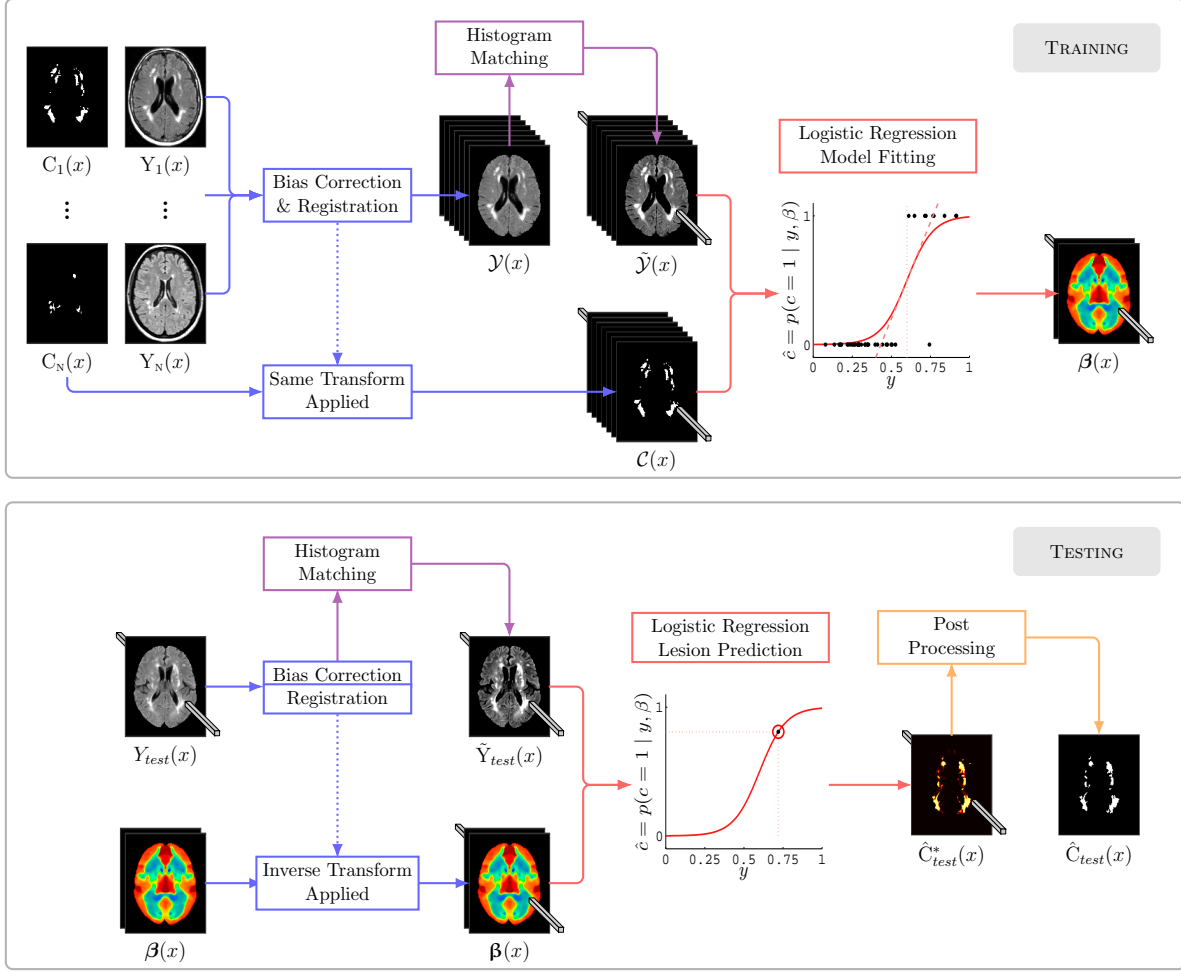
Figure 2.4: Overview of the proposed algorithm. Typefaces – upright Roman: images in native space; italic Roman: images in standard (MNI) space; calligraphic: a set of images from several patients; bold: a set of images corresponding to different features; Variables – $C(x)$: manual segmentation; $Y(x)$: FLAIR image, $\beta(x)$: parameter image; $\hat{C}(x)$: estimated lesion segmentation.

rater absolute intraclass correlation coefficient (ICC)[4] [157], while trends in over/undersegmentation with lesion load are illustrated using Blant-Altman plots [158].

2.8 Model Summary

A summary of the model is shown in Figure 2.4.

---

[4] Option 'A-1' in the Matlab function `ICC` from https://www.mathworks.com/matlabcentral/fileexchange/22099

Table 2.1: Model hyperparameters.

Stage	Parameter	Notation	Type	Default
Pre-Processing	Reflect Augmentation	A_R	$\mathbb{B}$	0
	Shift Augmentation	A_S	$\mathcal{N}_p$	$\mathcal{N}_0$
	Graylevel Transform	T_y	$f : y \mapsto \tilde{y}$	f_{he}
	Transform Mask	$M_T(x)$	$\mathbb{B}(x)$	M_{brain}
VLR Fitting	Iterations	N_t^{LR}	$\mathbb{Z}$	30
	Initial $\boldsymbol{\beta}$	$\boldsymbol{\beta}^{(0)}$	$\mathbb{R}^2$	$[0,0]$
	Learning Rate	α	$\mathbb{R}$	1
	Regularization	λ	$\mathbb{R}$	0
	Pseudo-Lesions	$\{\boldsymbol{\mathcal{Y}}_\rho, \boldsymbol{\mathcal{C}}_\rho\}$	$\{\{\cdot \in \mathbb{R}\}, \{\cdot \in [0,1]\}\}$	$\{\{\}, \{\}\}$
	$\boldsymbol{\beta}$ Filter	$F_{\boldsymbol{\beta}}$	$f : \boldsymbol{\beta}(\mathcal{N}_p(x)) \mapsto \tilde{\boldsymbol{\beta}}(x)$	$\tilde{\boldsymbol{\beta}}(x) = \boldsymbol{\beta}(x)$
Post-Processing	Min Lesion Size	$\mathcal{X}_{\text{min}}^c$	$\mathbb{R}$ (mm^3)	0
	MRF Iterations	N_t^{MRF}	$\mathbb{Z}$	0
	MRF Weights	W_{MRF}	$\{w_\Delta, w_F, w_y\}$	$\{1,1,1\}$

Notation. $\mathbb{B}$: boolean value; $\mathbb{Z}$: integer value; $\mathbb{R}^n$: real value ($n = 1$) or vector ($n > 1$); $\mathcal{N}_p$: nearest p voxel neighbourhood.

2.8.1 Tunable Parameters

In order to achieve the best possible model performance, it is prudent to track tunable model parameters (AKA hyperparameters) which are distinct from those fitted during each cross validation fold. Considering both the main VLR model and the pre- and post-processing aspects, the parameters of the proposed algorithm are summarized in Table 2.1 and also programmatically in `hypdef.m`. The optimization of these model components will be the subject of the next chapter.

Chapter 3

Experiment & Results

This section explores model validation. Performance of model components is characterized with respect to intermediate objectives, including graylevel standardization and regularization in toy scenarios. The segmentation performance of the full model is then presented under several cross validation frameworks, and compared to a similar algorithm.

3.1 Cross Validation Frameworks

Supervised segmentation models require the capacity to model complex relationships between the input image(s) and output label images. When models with large capacity are trained on a dataset which does not represent the full gamut of potential input data, they risk *overfitting*: acquiring a bias towards the training data [159]. The main problem associated with overfitting is decreased performance on new data (AKA generalization performance) [159]. Popular techniques for characterizing this expected decrease include cross validation (CV) procedures. These involve splitting the N available examples into training (t) and testing (e) subsets, where the training data are used to fit the model parameters, and the test data are used to approximate the expected generalization performance; the data splits are usually repeated, randomly or exhaustively, to ensure robust results [160]. The most popular CV frameworks include:

- **LOO – Leave-One-Out:** Withhold one example from the training set, and use it as the test case ($N_t = N - 1$; $N_e = 1$); repeat N times.
  Benefit: Close approximation of the expected generalization performance
  Drawback: Expensive to compute – $\mathcal{O}(N)$

- **KFCV – K-Fold Cross Validation:** Withhold a random batch of $B = N/K$ examples from the training set, and use it as the test set ($N_t = N - B$; $N_e = B$); repeat K times (without replacement).

  Benefit: Less expensive to compute – $\mathcal{O}(N/K)$

  Drawback: Worse approximation of the expected generalization performance

3.1.1 Leave-One-Source-Out CV

The choice of cross validation framework can have significant impacts on the reported model performance (see [160] for an in-depth review), and there is at least one assumption of the above methods which is not always valid: that examples are independent and identically distributed (iid). This is not true for data originating from multiple sources with different underlying distributions (e.g. MRI with different scan-parameter combinations) [161]. In fact, Geras et al. show that in multi-source problems where the expected use case involves data from entirely new sources, random KFCV (and therefore also LOO, as a special case of KFCV with $B = 1$) significantly overestimates the generalization performance. This is because random training fold selection allows the model to perceive source-specific characteristics of the test examples, which cannot be repeated for truly new examples. In such scenarios, the authors propose the following:

- **LOSO – Leave-One-Source-Out[1]:** Withhold all examples from source $s \in 1 \ldots S$ from the training set, and use these as the test set ($N_t = N - N_s$; $N_e = N_s$); repeat S times.

  Benefit: Best approximation of the expected generalization performance in multi-source problems

  Drawback: Still only an approximation

As noted in the introduction (cf. § 1.3.3), there has been surprisingly limited use of data from multiple sources for validation of WMH segmentation algorithms. Moreover, CV frameworks vary widely among papers, and to the best of this author's knowledge, no WMH algorithm has yet been validated using LOSO CV. This represents a significant caveat to reported performances, since MRI have many sources of variability (cf. § 1.2.2), including scanner manufacturer, field strength, sequence parameters, resolution, anatomical and disease variability. As the aim of this work is to develop a segmentation algorithm which will perform well on any given FLAIR MRI, the LOSO framework was initially developed without knowledge of the work by Geras et al. However, this paper happily corroborates the importance of LOSO CV to the current work. In this case, one data source is defined as a unique scanner-parameter combination.

---

[1] Authors' original name was "Multi-Source Cross Validation"

3.1.2 Competition Frameworks

One notable exception to this trend in WMH algorithms have been segmentation competitions [79, 57, 80, 12]. These generally provide both multi-source datasets and a robust validation framework. Table 3.1

Table 3.1: Summary of competition image databases.

Database	Ref.	Training		Testing	
		I	S	I	S
WMH 2017	[12]	60	3	110	5
MS 2016	[80]	15	3	38	4
MS 2015 ISBI	[57]	21	1	61	1
MS 2008	[79]	20	2	32	2

3.1.3 Data

For the reasons outlined above, it was important to collect a large and diverse database of FLAIR images for model validation. We use 129 FLAIR images from 10 different scanners. The number of images and scan parameters are summarized in Table 3.2. With the exception of the MS 2008[2] and In-House datasets, all of the data are freely available as part of the segmentation competitions. Since direct comparison of results on equal datasets is important for establishing state-of-the-art, we present results using only these freely available data ("Dataset A") in addition to all .

3.2 Graylevel Standardization

3.3 Regularization

3.3.1 Toy Model

In order to investigate the effect of different regularization strategies, a toy model is used. The model represents a single voxel during training, with synthetic observations. Regularizations are then chosen by hand to maintain desired characteristics in the fitted functions.

It is also possible to plot the MAP objective function $\mathcal{J}(\boldsymbol{\beta})$ in the 2D plane composed of $\boldsymbol{\beta} = [\beta^0, \beta^1]$.

---

[2] The manuals used with these data

Table 3.2: Summary of experimental image database.

Img (#)	Database	Ref.	Scanner	TE (ms)	TR (ms)	TI (ms)	Voxel Size (mm)	Manuals (#)
20	WMH 2017 (1) ■	[12]	3T Philips Achieva	125	11000	2800	$0.96 \times 0.96 \times 3.00$	1 [a]
20	WMH 2017 (2) ■	[12]	3T Siemens TrioTim	82	9000	2500	$1.00 \times 1.00 \times 3.00$	1 [a]
20	WMH 2017 (3) ■	[12]	3T GE Signa HDxt	126	8000	2340	$0.98 \times 1.20 \times 3.00$	1 [a]
5	MS 2016 (1) ■	[80]	3T Philips Ingenia	360	5400	1800	$0.50 \times 1.10 \times 0.50$	7 [b]
5	MS 2016 (2) ■	[80]	1.5T Siemens Aera	336	5400	1800	$1.04 \times 1.25 \times 1.04$	7 [b]
5	MS 2016 (3) ■	[80]	3T Siemens Verio	399	5000	1800	$0.74 \times 0.70 \times 0.74$	7 [b]
21	MS 2015 ISBI ■	[57]	3T Philips	68	11000	2800	$0.43 \times 0.43 \times 3.00$	2 [c]
13	In-House ■	—	3T Philips Achieva	125	9000	2800	$1.00 \times 1.00 \times 1.00$	1 [d]
10	MS 2008 CHB ■	[79]	—	—	—	—	$0.50 \times 0.50 \times 0.50$	1 [e]
10	MS 2008 UNC ■	[79]	3T Siemens Allegra	125	9000	2800	$0.50 \times 0.50 \times 0.50$	1 [f]

[a] Manuals were generated following the standards outlined in [78], and were subsequently reviewed by a second rater, only WMH labels were included; [b] Manuals were fused using the LOP-STAPLE method [162]; [c] Manuals were fused using logical 'and'; [d] Manuals were generated in-house; [e] REVIEW; [f] REVIEW.

3.4 Full Model

3.4.1 Convergence

The rate of convergence in each voxel will be unique. During parallel fitting, it is prudent to stop training after a reasonable number of iterations, rather than wait for all voxels to achieve a certain stopping criterion, in case a few aberrant voxels do not converge. In order to determine this number, the model was fitted using [X] selection(s) of the training data, and the value of the MAP objective

3.4.2 Parameter Images

For comparison with the LPA algorithm, the

3.4.3 Quantitative Performance

3.4.4 Comparison with Another Method

3.5 Discussion

(a) $\mathcal{J}(\boldsymbol{\beta})$ surface
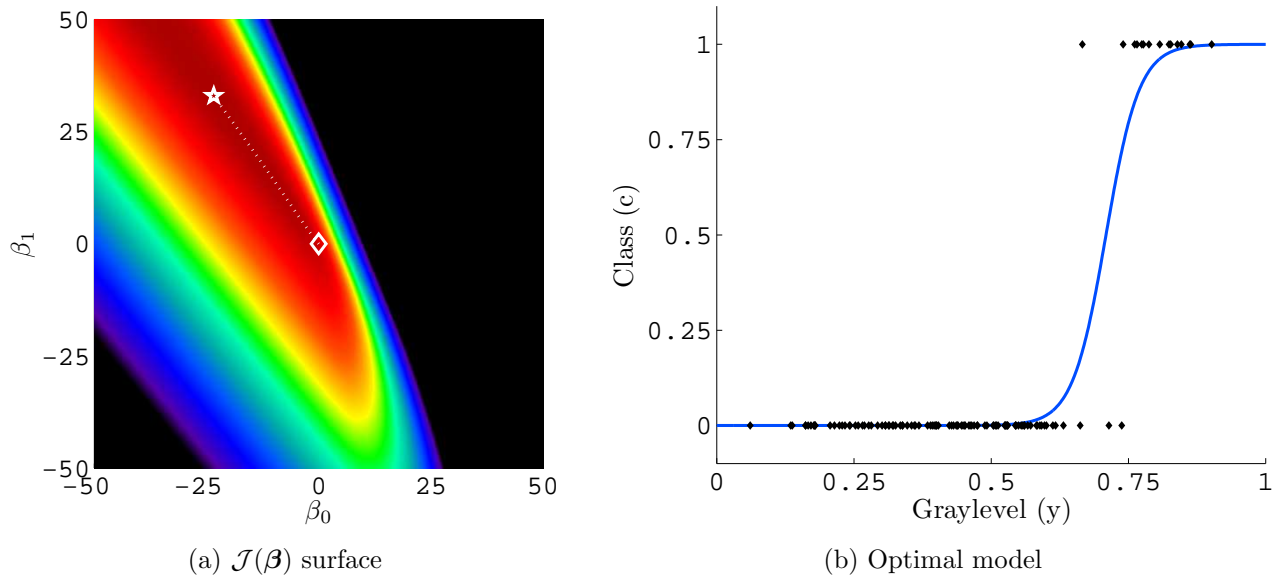
(b) Optimal model

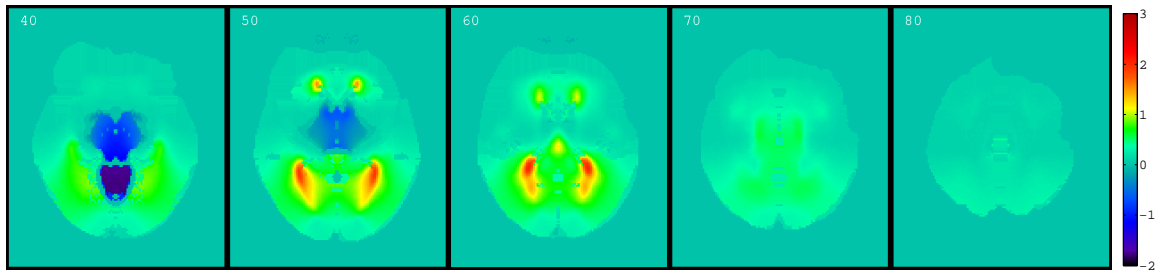Figure 3.1: MAP objective function in 2D; initialization (diamond), fitting (dotted), optimum (star).

Figure 3.2: Spatial effect parameter image from the LPA toolbox after registration to MNI space.

Chapter 4

Conclusion

4.1 Summary

4.2 Future Work

Use SPM segment with the FLAIR + T1 to obtain nice GM/WM segmentation. This allows: - mixture-model-based standardization of graylevels, since now more accurate? - estimated GM to clean up false positives

deep learning advantages: - context-aware - no graylevel standardization necessary

4.2.1 Other Features

4.2.2 Open Sourcing

Bibliography

[1] Robert A. Pooley. "Fundamental Physics of MR Imaging". In: *RadioGraphics* 25.4 (July 2005), pp. 1087–1099. DOI: 10.1148/rg.254055027.

[2] F Bloch, W W Hansen, and Martin Packard. "Nuclear Induction". In: *Phys. Rev.* 69.3-4 (Feb. 1946), p. 127. DOI: 10.1103/PhysRev.69.127.

[3] Robert G Bryant and Jean Pierre Korb. "Nuclear magnetic resonance and spin relaxation in biological systems". In: *Magnetic Resonance Imaging*. Vol. 23. 2 SPEC. ISS. 2005, pp. 167–173. DOI: 10.1016/j.mri.2004.11.026.

[4] S. H. Koenig et al. "Relaxometry of brain: Why white matter appears bright in MRI". In: *Magnetic Resonance in Medicine* 14.3 (June 1990), pp. 482–495. DOI: 10.1002/mrm.1910140306.

[5] T. P L Roberts and David Mikulis. "Neuro MR: Principles". In: *Journal of Magnetic Resonance Imaging* 26.4 (Oct. 2007), pp. 823–837. DOI: 10.1002/jmri.21029.

[6] Peter Schmitt et al. "Inversion Recovery TrueFISP: Quantification of T1, T 2, and Spin Density". In: *Magnetic Resonance in Medicine* 51.4 (Apr. 2004), pp. 661–667. DOI: 10.1002/mrm.20058.

[7] V.L L. Stevenson et al. "Variations in T1 and T2 relaxation times of normal appearing white matter and lesions in multiple sclerosis". In: *Journal of the neurological sciences* 178.2 (Sept. 2000), pp. 81–87. DOI: 10.1016/S0022-510X(00)00339-7.

[8] Govind B. Chavhan et al. "Principles, Techniques, and Applications of T2*-based MR Imaging and Its Special Applications". In: *RadioGraphics* 29.5 (Sept. 2009), pp. 1433–1449. DOI: 10.1148/rg.295095034.

[9] E. L. Hahn. "Spin echoes". In: *Physical Review* 80.4 (1950), pp. 580–594. DOI: 10.1103/PhysRev.80.580.

[10] G M Bydder and I R Young. "MR imaging: clinical use of the inversion recovery sequence." In: *Journal of computer assisted tomography* 9.4 (1985), pp. 659–675.

[11] J V Hajnal et al. "Use of fluid attenuated inversion recovery (FLAIR) pulse sequences in MRI of the brain." In: *Journal of computer assisted tomography* 16.6 (1992), pp. 841–844. DOI: 10.1097/00004728-199211000-00001.

[12] Hugo J. Kuijf et al. *WMH Segmentation Challenge*. 2017.

[13] Margaret M. Esiri. "Ageing and the brain". In: *Journal of Pathology* 211.2 (Feb. 2007), pp. 181–187. DOI: 10.1002/path.2089.

[14] C.D. Good et al. "A voxel-based morphometric study of ageing in 465 normal adult human brains". In: *5th IEEE EMBS International Summer School on Biomedical Imaging, 2002*. IEEE, 2002, II_5_1–II_5_16. DOI: 10.1109/SSBI.2002.1233974.

[15] S. Debette and H. S. Markus. "The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis". In: *BMJ* 341 (July 2010). DOI: 10.1136/bmj.c3666.

[16] John Conklin et al. "Are acute infarcts the cause of leukoaraiosis? Brain mapping for 16 consecutive weeks." eng. In: *Annals of neurology* 76.6 (Dec. 2014), pp. 899–904. DOI: 10.1002/ana.24285.

[17] Frank L Heppner, Richard M Ransohoff, and Burkhard Becher. "Immune attack: the role of inflammation in Alzheimer disease". In: *Nat Rev Neurosci* 16.6 (June 2015), pp. 358–372.

[18] Heather M. Snyder et al. "Vascular contributions to cognitive impairment and dementia including Alzheimer's disease". In: *Alzheimer's and Dementia* 11.6 (2015), pp. 710–717. DOI: 10.1016/j.jalz.2014.10.008.

[19] H. Bart van der Worp and Jan van Gijn. "Acute Ischemic Stroke". In: *New England Journal of Medicine* 357.6 (Aug. 2007), pp. 572–579. DOI: 10.1056/NEJMcp072057.

[20] Gregory W. Albers et al. "Transient Ischemic Attack - Proposal for a New Definition". In: *New England Journal of Medicine* 347.21 (Nov. 2002), pp. 1713–1716. DOI: 10.1056/NEJMsb020987.

[21] Leonardo Pantoni. "Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges". In: *The Lancet Neurology* 9.7 (July 2010), pp. 689–701. DOI: 10.1016/S1474-4422(10)70104-6.

[22] Shyam Prabhakaran, Ilana Ruff, and Richard A. Bernstein. "Acute Stroke Intervention A Systematic Review Clinical Review & Education Review". In: *JAMA* 313.14 (Apr. 2015), pp. 1451–1462. DOI: 10.1001/jama.2015.3058.

[23] G C Román et al. "Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop." In: *Neurology* 43.2 (Feb. 1993), pp. 250–260. DOI: 10.1212/WNL.43.2.250.

[24] Alistair Burns and Steve Iliffe. "Alzheimer's disease". In: *BMJ* 338 (2009). DOI: 10.1136/bmj.b158.

[25] C L Masters et al. "Amyloid plaque core protein in Alzheimer disease and Down syndrome." In: *Proceedings of the National Academy of Sciences* 82.12 (June 1985), pp. 4245–4249. DOI: 10.1073/pnas.82.12.4245.

[26] J. Hardy. "The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics". In: *Science* 297.5580 (July 2002), pp. 353–356. DOI: 10.1126/science.1072994.

[27] Virginia M Y Lee et al. "Developing therapeutic approaches to tau, selected kinases, and related neuronal protein targets." In: *Cold Spring Harbor perspectives in medicine* 1.1 (Sept. 2011), a006437–a006437. DOI: 10.1101/cshperspect.a006437.

[28] T. Kim et al. "Human LilrB2 Is a -Amyloid Receptor and Its Murine Homolog PirB Regulates Synaptic Plasticity in an Alzheimer's Model". In: *Science* 341.6152 (Sept. 2013), pp. 1399–1404. DOI: 10.1126/science.1242077.

[29] Shichun Tu et al. "Oligomeric Aβ-induced synaptic dysfunction in Alzheimer's disease". In: *Mol Neurodegener* 9.1 (Nov. 2014), pp. 1–12. DOI: 10.1186/1750-1326-9-48.

[30] Robert D Bell and Berislav V Zlokovic. "Neurovascular mechanisms and blood-brain barrier disorder in Alzheimer's disease." In: *Acta neuropathologica* 118.1 (July 2009), pp. 103–13. DOI: 10.1007/s00401-009-0522-3.

[31] Bruce D. Trapp and Klaus-Armin Nave. "Multiple sclerosis: an immune or neurodegenerative disorder?" In: *Annual review of neuroscience* 31.1 (July 2008), pp. 247–69. DOI: 10.1146/annurev.neuro.30.051606.094313.

[32] Claudia Lucchinetti et al. "Heterogeneity of multiple sclerosis lesions: Implications for the pathogenesis of demyelination". In: *Annals of Neurology* 47.6 (June 2000), pp. 707–717. DOI: 10.1002/1531-8249(200006)47:6<707::AID-ANA3>3.0.CO;2-Q.

[33] Olga Ciccarelli et al. "Pathogenesis of multiple sclerosis: insights from molecular and metabolic imaging." In: *The Lancet. Neurology* 13.8 (Aug. 2014), pp. 807–22. DOI: 10.1016/S1474-4422(14)70101-2.

[34] Don H Mahad, Bruce D Trapp, and Hans Lassmann. "Pathological mechanisms in progressive multiple sclerosis". In: *The Lancet Neurology* 14.2 (2015), pp. 183–193. DOI: 10.1016/S1474-4422(14)70256-X.

[35] Rohit Bakshi et al. "Imaging of multiple sclerosis: role in neurotherapeutics." In: *NeuroRx : the journal of the American Society for Experimental NeuroTherapeutics* 2.2 (Apr. 2005), pp. 277–303. DOI: 10.1602/neurorx.2.2.277.

[36] Joanna M Wardlaw, Maria C Valdés Hernández, and Susana Muñoz-Maniega. "What are white matter hyperintensities made of? Relevance to vascular cognitive impairment." In: *Journal of the American Heart Association* 4.6 (June 2015), p. 001140. DOI: 10.1161/JAHA.114.001140.

[37] F-E de Leeuw et al. "Prevalence of cerebral white matter lesions in elderly people: a population based magnetic resonance imaging study. The Rotterdam Scan Study". In: *Journal of Neurology, Neurosurgery & Psychiatry* 70.1 (Jan. 2001), 9 LP –14.

[38] Yulin Ge et al. "Dirty-Appearing White Matter in Multiple Sclerosis: Volumetric MR Imaging and Magnetization Transfer Ratio Histogram Analysis". In: *American Journal of Neuroradiology* 24.10 (2003), pp. 1935–1940.

[39] J.H. Simon et al. "Standardized MR Imaging Protocol for Multiple Sclerosis: Consortium of MS Centers Consensus Guidelines". In: *AJNR Am J Neuroradiol* 27.2 (Feb. 2006), pp. 455–461.

[40] Joanna M Wardlaw et al. *Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration*. Aug. 2013. DOI: 10.1016/S1474-4422(13)70124-8.

[41] Chris H Polman et al. "Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria." In: *Annals of neurology* 69.2 (Feb. 2011), pp. 292–302. DOI: 10.1002/ana.22366.

[42] F Fazekas, J B Chawluk, and A Alavi. "MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging". In: *American Journal of Neuroradiology* 8.3 (Aug. 1987), pp. 421–426. DOI: 10.2214/ajr.149.2.351.

[43] Christine Egger et al. "MRI FLAIR lesion segmentation in Multiple Sclerosis: Does automated segmentation hold up with manual annotation?" In: *NeuroImage: Clinical* 13 (2016), pp. 264–270. DOI: 10.1016/j.nicl.2016.11.020.

[44] F D Lublin et al. "Defining the clinical course of multiple sclerosis: the 2013 revisions". In: *Neurology* 83 (2014). DOI: 10.1212/WNL.0000000000000560.

[45] Anthony Traboulsee et al. "Canadian Expert Panel Recommendations for MRI Use in MS Diagnosis and Monitoring." In: *The Canadian journal of neurological sciences. Le journal canadien des sciences neurologiques* 42.3 (May 2015), pp. 159–67. DOI: 10.1017/cjn.2015.24.

[46] Maria Pia Sormani and Paolo Bruzzi. "MRI lesions as a surrogate for relapses in multiple sclerosis: A meta-analysis of randomised trials". In: *The Lancet Neurology* 12.7 (2013), pp. 669–676. DOI: 10.1016/S1474-4422(13)70103-0.

[47] Kyle Fahrbach et al. "Relating Relapse and T2 Lesion Changes to Disability Progression in MS: A Systematic Literature Review and Regression Analysis (P05.090)". In: *Neurology* 78.Meeting Abstracts 1 (Nov. 2012), P05.090–P05.090. DOI: 10.1212/WNL.78.1_MeetingAbstracts.P05.090.

[48] Tjalf Ziemssen et al. *Optimizing therapy early in multiple sclerosis: An evidence-based view.* Sept. 2015. DOI: 10.1016/j.msard.2015.07.007.

[49] J I O'Riordan et al. "The prognostic value of brain MRI in clinically isolated syndromes of the CNS. A 10-year follow-up". In: *Brain* 121.3 (Mar. 1998), pp. 495–503. DOI: 10.1093/brain/121.3.495.

[50] Daisy Mollison et al. "The clinico-radiological paradox of cognitive function and MRI burden of white matter lesions in people with multiple sclerosis: A systematic review and meta-analysis". In: *PLoS ONE* 12.5 (May 2017). Ed. by Orhan Aktas, e0177727. DOI: 10.1371/journal.pone.0177727.

[51] Bruno Dubois et al. "Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria". In: *Lancet Neurology* 6.8 (Aug. 2007), pp. 734–746. DOI: 10.1016/S1474-4422(07)70178-3.

[52] S.a Sorbi et al. "EFNS-ENS Guidelines on the diagnosis and management of disorders associated with dementia". In: *European Journal of Neurology* 19.9 (2012), pp. 1159–1179. DOI: 10.1111/j.1468-1331.2012.03784.x.

[53] Martijn V. Verhagen et al. "The impact of MRI combined with visual rating scales on the clinical diagnosis of dementia: a prospective study". In: *European Radiology* 26.6 (June 2016), pp. 1716–1722. DOI: 10.1007/s00330-015-3957-z.

[54] Guy M. McKhann et al. "The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease". In: *Alzheimer's and Dementia* 7.3 (May 2011), pp. 263–269. DOI: 10.1016/j.jalz.2011.03.005.

[55] Danielle van Westen et al. "Cerebral white matter lesions - associations with $A\beta$ isoforms and amyloid PET." In: *Scientific reports* 6 (2016), p. 20709. DOI: 10.1038/srep20709.

[56] Ki Woong Kim, James R. MacFall, and Martha E. Payne. "Classification of White Matter Lesions on Magnetic Resonance Imaging in Elderly Persons". In: *Biological Psychiatry* 64.4 (Aug. 2008), pp. 273–280. DOI: 10.1016/j.biopsych.2008.03.024.

[57] Aaron Carass et al. "Longitudinal multiple sclerosis lesion segmentation data resource". In: *Data in Brief* 12 (June 2017), pp. 346–350. DOI: 10.1016/j.dib.2017.04.004.

[58] Tomoko Okuda et al. "Brain Lesions: When Should Fluid-attenuated Inversion-Recovery Sequences Be Used in MR Evaluation?" In: *Radiology* 212.3 (Sept. 1999), pp. 793–798. DOI: 10.1148/radiology.212.3.r99se07793.

[59] M. Rovaris et al. "The contribution of fast-FLAIR MRI for lesion detection in the brain of patients with systemic autoimmune diseases." In: *Journal of neurology* 247.1 (Jan. 2000), pp. 29–33. DOI: 10.1007/s004150050006.

[60] Rohit Bakshi et al. "Fluid-attenuated inversion recovery magnetic resonance imaging detects cortical and juxtacortical multiple sclerosis lesions". In: *Archives of neurology* 58.5 (May 2001), pp. 742–748. DOI: 10.1001/archneur.58.5.742.

[61] Frederik Barkhof and Philip Scheltens. "Imaging of white matter lesions." In: *Cerebrovascular diseases (Basel, Switzerland)* 13 Suppl 2.suppl 2 (2002), pp. 21–30. DOI: 49146.

[62] Rola Harmouche et al. "Bayesian MS lesion classification modeling regional and local spatial information". In: *Proceedings - International Conference on Pattern Recognition*. Vol. 3. IEEE, 2006, pp. 984–987. DOI: 10.1109/ICPR.2006.318.

[63] Renske de Boer et al. "White matter lesion extension to automatic brain tissue segmentation on MRI." In: *NeuroImage* 45.4 (May 2009), pp. 1151–61. DOI: 10.1016/j.neuroimage.2009.01.011.

[64] Martijn D Steenwijk et al. "Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs)." In: *NeuroImage. Clinical* 3 (Jan. 2013), pp. 462–9. DOI: 10.1016/j.nicl.2013.10.003.

[65] Martha E Payne et al. "Development of a semi-automated method for quantification of MRI gray and white matter lesions in geriatric subjects". In: *Psychiatry Research - Neuroimaging* 115.1-2 (Aug. 2002), pp. 63–77. DOI: 10.1016/S0925-4927(02)00009-4.

[66] M Ghazel, A Traboulsee, and R K Ward. "Semi-Automated Segmentation of Multiple Sclerosis Lesions in Brain MRI using Texture Analysis". In: *2006 IEEE International Symposium on Signal Processing and Information Technology*. Aug. 2006, pp. 6–10. DOI: 10.1109/ISSPIT.2006.270760.

[67] Yasuo Kawata et al. "Computer-aided evaluation method of white matter hyperintensities related to subcortical vascular dementia based on magnetic resonance imaging". In: *Computerized Medical Imaging and Graphics* 34.5 (July 2010), pp. 370–376. DOI: 10.1016/j.compmedimag.2009.12.014.

[68] Mariangela Iorio et al. "White matter hyperintensities segmentation: a new semi-automated method." In: *Frontiers in aging neuroscience* 5 (Jan. 2013), p. 76. DOI: 10.3389/fnagi.2013.00076.

[69] D Mortazavi, A Z Kouzani, and H Soltanian-Zadeh. "Segmentation of multiple sclerosis lesions in MR images: a review". In: *Neuroradiology* 54 (2012). DOI: 10.1007/s00234-011-0886-7.

[70] Jaber Juntu et al. "Bias field correction for mri images". In: *Computer Recognition Systems* 30 (2005), pp. 543–551. DOI: 10.1007/3-540-32390-2_64.

[71] P. Santago. "Statistical models of partial volume effect". In: *Image Processing, IEEE Transactions on* 4.11 (1995), pp. 1531–1540.

[72] W.J. Niessen et al. "Multiscale Segmentation of Three-Dimensional MR Brain Images". In: *International Journal of Computer Vision* 31.2/3 (1999), pp. 185–202. DOI: 10.1023/A:1008070000018.

[73] Rohit Bakshi et al. "Intraventricular CSF pulsation artifact on fast fluid-attenuated inversion-recovery MR images: Analysis of 100 consecutive normal studies". In: *American Journal of Neuroradiology* 21.3 (2000), pp. 503–508.

[74] Maxim Zaitsev, Julian Maclaren, and Michael Herbst. *Motion artifacts in MRI: A complex problem with many partial solutions*. Oct. 2015. DOI: 10.1002/jmri.24850.

[75] Ioannis Kapouleas. "Automatic detection of white matter lesions in magnetic resonance brain images". In: *Computer Methods and Programs in Biomedicine* 32.1 (May 1990), pp. 17–35. DOI: 10.1016/0169-2607(90)90082-K.

[76] Xavier Lladó et al. "Segmentation of multiple sclerosis lesions in brain MRI: A review of automated approaches". In: *Information Sciences* 186.1 (2012), pp. 164–185. DOI: http://dx.doi.org/10.1016/j.ins.2011.10.011.

[77] Daniel Garcia-Lorenzo et al. "Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging." eng. In: *Medical image analysis* 17.1 (Jan. 2013), pp. 1–18. DOI: 10.1016/j.media.2012.09.004.

[78] Maria Eugenia Caligiuri et al. "Automatic Detection of White Matter Hyperintensities in Healthy Aging and Pathology Using Magnetic Resonance Imaging: A Review". In: *Neuroinformatics* 13.3 (July 2015), pp. 261–276. DOI: 10.1007/s12021-015-9260-y.

[79] *MS Lesion Segmentation Challenge 2008*. 2008.

[80] *MS Lesion Segmentation Challenge 2016*. 2016.

[81] K. Van Leemput et al. "Automated segmentation of multiple sclerosis lesions by model outlier detection". In: *IEEE Transactions on Medical Imaging* 20.8 (2001), pp. 677–688. DOI: 10.1109/42.938237.

[82] Clifford R Jack et al. "FLAIR histogram segmentation for measurement of leukoaraiosis volume". In: *Journal of Magnetic Resonance Imaging* 14.6 (Dec. 2001), pp. 668–676. DOI: 10.1002/jmri.10011.

[83] Alex P Zijdenbos, Reza Forghani, and Alan C Evans. "Automatic pipeline analysis of 3-D MRI data for clinical trials: application to multiple sclerosis." In: *IEEE transactions on medical imaging* 21.10 (Oct. 2002), pp. 1280–91. DOI: 10.1109/TMI.2002.806283.

[84] Petronella Anbeek et al. "Probabilistic segmentation of white matter lesions in MR imaging." In: *NeuroImage* 21.3 (Mar. 2004), pp. 1037–44. DOI: 10.1016/j.neuroimage.2003.10.012.

[85] Petronella Anbeek et al. "Probabilistic segmentation of brain tissue in MR imaging." In: *NeuroImage* 27.4 (Oct. 2005), pp. 795–804. DOI: 10.1016/j.neuroimage.2005.05.046.

[86] F. Admiraal-Behloul et al. "Fully automatic segmentation of white matter hyperintensities in MR images of the elderly". In: *NeuroImage* 28.3 (2005), pp. 607–617. DOI: 10.1016/j.neuroimage.2005.06.061.

[87] Z. Lao et al. "Automated Segmentation of White Matter Lesions in 3D Brain MR Images, using Multivariate Pattern Classification". In: *3rd IEEE International Symposium on Biomedical Imaging: Macro to Nano, 2006.* IEEE, 2006, pp. 307–310. DOI: 10.1109/ISBI.2006.1624914.

[88] Ying Wu et al. "Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI". In: *NeuroImage* 32.3 (Sept. 2006), pp. 1205–1215. DOI: 10.1016/j.neuroimage.2006.04.211.

[89] Balasrinivasa Rao Sajja et al. "Unified approach for multiple sclerosis lesion segmentation on brain MRI." In: *Annals of biomedical engineering* 34.1 (Jan. 2006), pp. 142–51. DOI: 10.1007/s10439-005-9009-0.

[90] Rasoul Khayati et al. "Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and Markov random field model." eng. In: *Computers in biology and medicine* 38.3 (Mar. 2008), pp. 379–390. DOI: 10.1016/j.compbiomed.2007.12.005.

[91]   M. Wels, M. Huber, and J. Hornegger. "Fully automated segmentation of multiple sclerosis lesions in multispectral MRI". In: *Pattern Recognition and Image Analysis* 18.2 (June 2008), pp. 347–350. DOI: 10.1134/S1054661808020235.

[92]   E H Herskovits, R N Bryan, and F Yang. "Automated Bayesian segmentation of microvascular white-matter lesions in the ACCORD-MIND study." eng. In: *Advances in medical sciences* 53.2 (2008), pp. 182–190. DOI: 10.2478/v10039-008-0039-3.

[93]   S Bricq, Christophe Collet, and Jean-Paul Armspach. "MS Lesion Segmentation based on Hidden Markov Chains". In: *The MIDAS Journal - MS Lesion Segmentation (MICCAI 2008 Workshop)*. 2008.

[94]   Tim B Dyrby et al. "Segmentation of age-related white matter changes in a clinical multi-center study." eng. In: *NeuroImage* 41.2 (June 2008), pp. 335–345. DOI: 10.1016/j.neuroimage.2008.02.024.

[95]   Jc Souplet et al. "An Automatic Segmentation of T2-FLAIR Multiple Sclerosis Lesions". In: *MICCAI Grand Challenge Workshop: Multiple Sclerosis Lesion Segmentation Challenge* (2008), pp. 1–11.

[96]   Daniel García-Lorenzo et al. "Multiple sclerosis lesion segmentation using an automatic multimodal graph cuts". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 5762 LNCS. PART 2. 2009, pp. 584–591. DOI: 10.1007/978-3-642-04271-3_71.

[97]   Ayelet Akselrod-Ballin et al. "Automatic Segmentation and Classification of Multiple Sclerosis in Multichannel MRI". In: *IEEE Transactions on Biomedical Engineering* 56.10 (2009), pp. 2461–2469.

[98]   Christopher Schwarz et al. "Fully-automated white matter hyperintensity detection with anatomical prior knowledge and without FLAIR". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 5636 LNCS. NIH Public Access, 2009, pp. 239–251. DOI: 10.1007/978-3-642-02498-6_20.

[99]   Erin Gibson et al. "Automatic segmentation of white matter hyperintensities in the elderly using FLAIR images at 3T". In: *Journal of Magnetic Resonance Imaging* 31.6 (June 2010), pp. 1311–1322. DOI: 10.1002/jmri.22004.

[100] Navid Shiee et al. "A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions." In: *NeuroImage* 49.2 (Jan. 2010), pp. 1524–1535. DOI: 10.1016/j.neuroimage.2009.09.005.

[101] Mark Scully et al. "An Automated Method for Segmenting White Matter Lesions through Multi-Level Morphometric Feature Classification with Application to Lupus." In: *Frontiers in human neuroscience* 4.April (2010), p. 27. DOI: 10.3389/fnhum.2010.00027.

[102] Daniel García-Lorenzo et al. "Trimmed-likelihood estimation for focal lesions and tissue segmentation in multisequence MRI for multiple sclerosis." In: *IEEE Transactions on Medical Imaging* 30.8 (Aug. 2011), pp. 1455–67. DOI: 10.1109/TMI.2011.2114671.

[103] Ezequiel Geremia et al. "Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images." eng. In: *NeuroImage* 57.2 (July 2011), pp. 378–390. DOI: 10.1016/j.neuroimage.2011.03.080.

[104] Sean D Smart, Michael J Firbank, and John T O'Brien. "Validation of automated white matter hyperintensity segmentation." In: *Journal of aging research* 2011 (Jan. 2011), p. 391783. DOI: 10.4061/2011/391783.

[105] Thomas Samaille et al. "Contrast-based fully automatic segmentation of white matter hyperintensities: method and validation." eng. In: *PloS one* 7.11 (2012), e48953. DOI: 10.1371/journal.pone.0048953.

[106] April Khademi, Anastasios Venetsanopoulos, and Alan R Moody. "Robust white matter lesion segmentation in FLAIR MRI." eng. In: *IEEE transactions on bio-medical engineering* 59.3 (Mar. 2012), pp. 860–871. DOI: 10.1109/TBME.2011.2181167.

[107] Paul Schmidt et al. "An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis". eng. In: *Neuroimage* 59.4 (Feb. 2012), pp. 3774–3783. DOI: 10.1016/j.neuroimage.2011.11.032.

[108] Bassem a Abdullah, Akmal a Younis, and Nigel M John. "Multi-Sectional Views Textural Based SVM for MS Lesion Segmentation in Multi-Channels MRIs." eng. In: *The open biomedical engineering journal* 6 (2012), pp. 56–72. DOI: 10.2174/1874230001206010056.

[109] Elizabeth M. Sweeney et al. "OASIS is Automated Statistical Inference for Segmentation, with applications to multiple sclerosis lesion segmentation in MRI". In: *NeuroImage: Clinical* 2.1 (2013), pp. 402–413. DOI: 10.1016/j.nicl.2013.03.002.

[110] Sushmita Datta and Ponnada A. Narayana. "A comprehensive approach to the segmentation of multichannel three-dimensional MR brain images in multiple sclerosis". In: *NeuroImage: Clinical* 2 (2013), pp. 184–196. DOI: 10.1016/j.nicl.2012.12.007.

[111] April Khademi, Anastasios Venetsanopoulos, and Alan R Moody. "Generalized method for partial volume estimation and tissue segmentation in cerebral magnetic resonance images". In: *Journal of Medical Imaging* 1.1 (Apr. 2014), p. 14002. DOI: 10.1117/1.JMI.1.1.014002.

[112] Vamsi Ithapu et al. "Extracting and summarizing white matter hyperintensities using supervised segmentation methods in Alzheimer's disease risk and aging studies". In: *Human Brain Mapping* 35.8 (2014), pp. 4219–4235. DOI: `10.1002/hbm.22472`.

[113] Byung Il Yoo et al. "Application of variable threshold intensity to segmentation for white matter hyperintensities in fluid attenuated inversion recovery magnetic resonance images". In: *Neuroradiology* 56.4 (2014), pp. 265–281. DOI: `10.1007/s00234-014-1322-6`.

[114] Rola Harmouche et al. "Probabilistic multiple sclerosis lesion classification based on modeling regional intensity variability and local neighborhood Information". In: *IEEE Transactions on Biomedical Engineering* 62.5 (2015), pp. 1281–1292. DOI: `10.1109/TBME.2014.2385635`.

[115] Nicolas Guizard et al. "Rotation-invariant multi-contrast non-local means for MS lesion segmentation". In: *NeuroImage: Clinical* 8 (2015), pp. 376–389. DOI: `10.1016/j.nicl.2015.05.001`.

[116] Saurabh Jain et al. "Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images." In: *NeuroImage: Clinical* 8 (2015), pp. 367–75. DOI: `10.1016/j.nicl.2015.05.003`.

[117] Xavier Tomas-Fernandez and Simon K Warfield. "A model of population and subject (MOPS) intensities with application to multiple sclerosis lesion segmentation". In: *IEEE Transactions on Medical Imaging* 34.6 (June 2015), pp. 1349–1361. DOI: `10.1109/TMI.2015.2393853`.

[118] Rui Wang et al. "Automatic segmentation and volumetric quantification of white matter hyperintensities on fluid-attenuated inversion recovery images using the extreme value distribution." eng. In: *Neuroradiology* 57.3 (Mar. 2015), pp. 307–320. DOI: `10.1007/s00234-014-1466-4`.

[119] Pallab Kanti Roy et al. "Automatic white matter lesion segmentation using contrast enhanced FLAIR intensity and Markov Random Field." eng. In: *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society* 45 (Oct. 2015), pp. 102–111. DOI: `10.1016/j.compmedimag.2015.08.005`.

[120] Tom Brosch et al. *Deep Convolutional Encoder Networks for Multiple Sclerosis Lesion Segmentation*. Vol. 9556. Springer International Publishing, 2015, pp. 144–155. DOI: `10.1007/978-3-319-30858-6`.

[121] Mário João Fartaria et al. "Automated detection of white matter and cortical lesions in early stages of multiple sclerosis." In: *Journal of magnetic resonance imaging : JMRI* (Nov. 2015). DOI: `10.1002/jmri.25095`.

[122] Hrishikesh Deshpande, Pierre Maurel, and Christian Barillot. "Classification of multiple sclerosis lesions using adaptive dictionary learning". In: *Computerized Medical Imaging and Graphics* 46 (2015), pp. 2–10. DOI: `10.1016/j.compmedimag.2015.05.003`.

[123] Eloy Roura et al. "A toolbox for multiple sclerosis lesion segmentation". In: *Neuroradiology* 57.10 (Oct. 2015), pp. 1031–1043. DOI: `10.1007/s00234-015-1552-2`.

[124] Jesse Knight and April Khademi. "MS Lesion Segmentation Using FLAIR MRI Only". In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI*. Athens, Greece, 2016, TBD.

[125] Roey Mechrez, Jacob Goldberger, and Hayit Greenspan. "Patch-Based Segmentation with Spatial Consistency: Application to MS Lesions in Brain MRI". In: *International Journal of Biomedical Imaging* 2016 (2016), pp. 1–13. DOI: `10.1155/2016/7952541`.

[126] Maddalena Strumia et al. "White Matter MS-Lesion Segmentation Using a Geometric Brain Model." In: *IEEE transactions on medical imaging* PP.99 (2016), p. 1. DOI: `10.1109/TMI.2016.2522178`.

[127] Ludovica Griffanti et al. "BIANCA (Brain Intensity AbNormality Alassification Algorithm): A new tool for automated segmentation of white matter hyperintensities". In: *NeuroImage* 141 (2016), pp. 191–205. DOI: `10.1016/j.neuroimage.2016.07.018`.

[128] Sergi Valverde et al. "Automated tissue segmentation of MR brain images in the presence of white matter lesions". In: *Medical Image Analysis* 35 (Aug. 2017), pp. 446–457. DOI: `10.1016/j.media.2016.08.014`.

[129] Mahsa Dadar et al. "Validation of a Regression Technique for Segmentation of White Matter Hyperintensities in Alzheimer's Disease". In: *IEEE Transactions on Medical Imaging* 99.PP (2017), pp. 1–1. DOI: `10.1109/TMI.2017.2693978`.

[130] Tianming Zhan et al. "Multimodal spatial-based segmentation framework for white matter lesions in multi-sequence magnetic resonance images". In: *Biomedical Signal Processing and Control* 31 (2017), pp. 52–62. DOI: `10.1016/j.bspc.2016.06.016`.

[131] John Mazziotta et al. "A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM)." In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 356.1412 (2001), pp. 1293–322. DOI: `10.1098/rstb.2001.0915`.

[132] Jesse Knight, April Khademi, and Graham Taylor. "Voxel-Wise Logistic Regression and Leave-One-Scanner-Out Cross Validation for White Matter Hyperintensity Segmentation". In: *NeuroImage* ().

[133] April Khademi and Alan R. Moody. "Multiscale partial volume estimation for segmentation of white matter lesions using flair MRI". In: *Proceedings - International Symposium on Biomedical Imaging*. Vol. 2015-July. IEEE, Apr. 2015, pp. 568–571. DOI: `10.1109/ISBI.2015.7163937`.

[134] John Ashburner and Karl J Friston. "Multimodal Image Coregistration and Partitioning - a Unified Framework". In: *Neuroimage* 6.3 (Oct. 1997), pp. 209–217. DOI: `10.1006/nimg.1997.0290`.

[135] John Ashburner and Karl J Friston. "Unified segmentation." eng. In: *NeuroImage* 26.3 (July 2005), pp. 839–851. DOI: `10.1016/j.neuroimage.2005.02.018`.

[136] N K Subbanna et al. "MS Lesion Segmentation using Markov Random Fields". In: *MICCAI Workshop on Medical Image Analysis on Multiple Sclerosis*. 2009, pp. 37–48.

[137] Pierre Louis Bazin and Dzung L. Pham. "Homeomorphic brain image segmentation with topological and statistical atlases". In: *Medical Image Analysis* 12.5 (Oct. 2008), pp. 616–625. DOI: `10.1016/j.media.2008.06.008`.

[138] Nicholas J Tustison et al. "N4ITK: improved N3 bias correction." In: *IEEE Transactions on Medical Imaging* 29.6 (June 2010), pp. 1310–20. DOI: `10.1109/TMI.2010.2046908`.

[139] Stephen M Smith. "Fast robust automated brain extraction." In: *Human brain mapping* 17.3 (Nov. 2002), pp. 143–55. DOI: `10.1002/hbm.10062`.

[140] Y. Zhang, M. Brady, and S. Smith. "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm". In: *IEEE Transactions on Medical Imaging* 20.1 (2001), pp. 45–57. DOI: `10.1109/42.906424`.

[141] M Jenkinson and S Smith. "A global optimisation method for robust affine registration of brain images". In: *Med Image Anal* 5 (2001). DOI: `10.1016/S1361-8415(01)00036-6`.

[142] L G Nyul and J K Udupa. "On standardizing the MR image intensity scale." eng. In: *Magnetic resonance in medicine* 42.6 (Dec. 1999), pp. 1072–1081.

[143] Paul Schmidt. *LST: A lesion segmentation tool for SPM*. 2015.

[144] Jesse Knight, Alan R Moody, and April Khademi. "Noise in parallel MRI: how to determine whether single-coil assumptions still hold (they don't) (Poster)". In: *Imaging Network Ontario Symposium*. Toronto, 2016. DOI: `10.13140/RG.2.2.11028.91527`.

[145] John G. Sled et al. "Regional Variations in Normal Brain Shown by Quantitative Magnetization Transfer Imaging". In: *Magnetic Resonance in Medicine* 51.2 (Feb. 2004), pp. 299–303. DOI: `10.1002/mrm.10701`.

[146] Paul Schmidt. "Bayesian inference for structured additive regression models for large-scale problems with applications to medical imaging". PhD thesis. Ludwig-Maximilians-Universität München, 2017.

[147] E M Sweeney et al. "Automatic lesion incidence estimation and detection in multiple sclerosis using multisequence longitudinal MRI". In: *American Journal of Neuroradiology* 34.1 (Jan. 2013), pp. 68–73. DOI: `10.3174/ajnr.A3172`.

[148] Paul Schmidt, Mark Mühlau, and Volker Schmid. "Fitting large-scale structured additive regression models using Krylov subspace methods". In: *Computational Statistics and Data Analysis* 105 (2017), pp. 59–75. DOI: `10.1016/j.csda.2016.07.006`.

[149] Thomas P Minka. "A comparison of numerical optimizers for logistic regression". 2003.

[150] Martin A. Tanner et al. "The calculation of posterior distributions by data augmentation". In: *Journal of the American Statistical Association* 82.398 (June 1987), pp. 528–540. DOI: `10.2307/2289457`.

[151] Ian Goodfellow et al. "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems 27*. Ed. by Z Ghahramani et al. Curran Associates, Inc., 2014, pp. 2672–2680.

[152] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems 25*. Ed. by F Pereira et al. Curran Associates, Inc., 2012, pp. 1097–1105.

[153] Hui Zou and Trevor Hastie. "Regularization and variable selection via the elastic net". In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 67.2 (Apr. 2005), pp. 301–320. DOI: `10.1111/j.1467-9868.2005.00503.x`.

[154] Robert Tibshirani. "Regression selection and shrinkage via the lasso". In: *Journal of the Royal Statistical Society B* 58.1 (1996), pp. 267–288. DOI: `10.2307/2346178`.

[155] Su-in Lee et al. "Efficient L1 Regularized Logistic Regression". In: *AAAI* 6.1 (2006), pp. 401–408. DOI: `10.1.1.64.1993`.

[156] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. 3rd ed. Upper Saddle River, New Jersey: Prentice Hall, Inc., Feb. 2006.

[157] Terry K Koo and Mae Y Li. "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research". In: *Journal of Chiropractic Medicine* 15.2 (June 2016), pp. 155–163. DOI: 10.1016/j.jcm.2016.02.012.

[158] D. G. Altman and J. M. Bland. "Measurement in Medicine: the Analysis of Method Comparison Studies". In: *Statistician* 32.July 1981 (Sept. 1983), pp. 307–317. DOI: 10.2307/2987937.

[159] Douglas M Hawkins. "The Problem of Overfitting". In: *Journal of Chemical Information and Computer Sciences* 44.1 (2004), pp. 1–12. DOI: 10.1021/ci0342472.

[160] Sylvain Arlot and Alain Celisse. "A survey of cross-validation procedures for model selection". In: *Statistics Surveys* 4 (2009), pp. 40–79. DOI: 10.1214/09-SS054.

[161] Krzysztof Geras and Charles Sutton. "Multiple-source cross-validation". In: *Proceedings of the 30th International Conference on Machine Learning*. Ed. by Sanjoy Dasgupta and David McAllester. Vol. 28. Proceedings of Machine Learning Research 3. Atlanta, Georgia, USA: PMLR, 2013, pp. 1292–1300.

[162] Alireza Akhondi-Asl et al. "A logarithmic opinion pool based STAPLE algorithm for the fusion of segmentations with associated reliability weights." In: *IEEE Transactions on Medical Imaging* 33.10 (Oct. 2014), pp. 1997–2009. DOI: 10.1109/TMI.2014.2329603.

[163] D.L. Collins. "Design and construction of a realistic digital brain phantom". In: *IEEE Transactions on Medical Imaging* 17.3 (1998), pp. 463–468.

[164] Andriy Fedorov et al. "3D Slicer as an image computing platform for the Quantitative Imaging Network". In: *Magnetic Resonance Imaging* 30.9 (Nov. 2012), pp. 1323–1341. DOI: 10.1016/j.mri.2012.05.001.

Appendix A

Maths

This section presents various mathematical results which are not crucial to the thesis.

A.1 FLAIR MRI Intensity Modelling

While MR imaging is both complex and mutable, simulation of the expected signal intensities is possible using the relaxometry data in Table 1.1 and sequence signal equations – e.g. (1.3) and (1.4). In the clinic, this simulation helps select appropriate acquisition parameters $TE/TR/TI$ for the desired contrast; however, these contrast characteristics can also be later considered as covariates in performance analysis of segmentation tools. To this end, the acquisition parameters of the FLAIR image database (Table 3.2) were used with the data from Table 1.1 and Equation (1.4) to calculate the expected tissue contrasts for these data. The contrast between two graylevels was defined as

$$2\left|\frac{y_1 - y_2}{y_1 + y_2}\right|. \tag{A.1}$$

These results are summarized in Table A.1, and an example image is generated for each parameter set using the tissue maps from the BrainWeb database [163][1], as shown in Figure A.1.

A.2 Histogram Equalization Properties

Table A.1: Simulated FLAIR tissue contrasts using scan parameters from the experimental database.

Scanner	GM/WM	GM/CSF	GM/WMH	WM/CSF	WM/WMH	CSF/WMH
WMH 2017 (1)	0.19	1.28	0.32	1.16	0.50	1.45
WMH 2017 (2)	0.10	1.83	0.14	1.81	0.25	1.85
WMH 2017 (3)	0.13	1.95	0.27	1.94	0.39	1.96
MS 2016 (1)	0.28	2.00	0.88	2.00	1.09	2.00
MS 2016 (2)	0.25	2.00	0.82	2.00	1.01	2.00
MS 2016 (3)	0.32	1.77	0.99	1.69	1.21	1.92
MS 2008 CHB	—	—	—	—	—	—
MS 2008 UNC	0.19	0.78	0.32	0.62	0.50	1.04
ISBI MS 2015	0.12	1.19	0.14	1.11	0.26	1.27
In-House	0.19	0.78	0.32	0.62	0.50	1.04

(a) WMH 2017 (1)   (b) WMH 2017 (2)   (c) WMH 2017 (3)

(d) MS 2016 (1)   (e) MS 2016 (2)   (f) MS 2016 (3)

(g) MS 2008 UNC   (h) ISBI MS 2015   (i) In-House

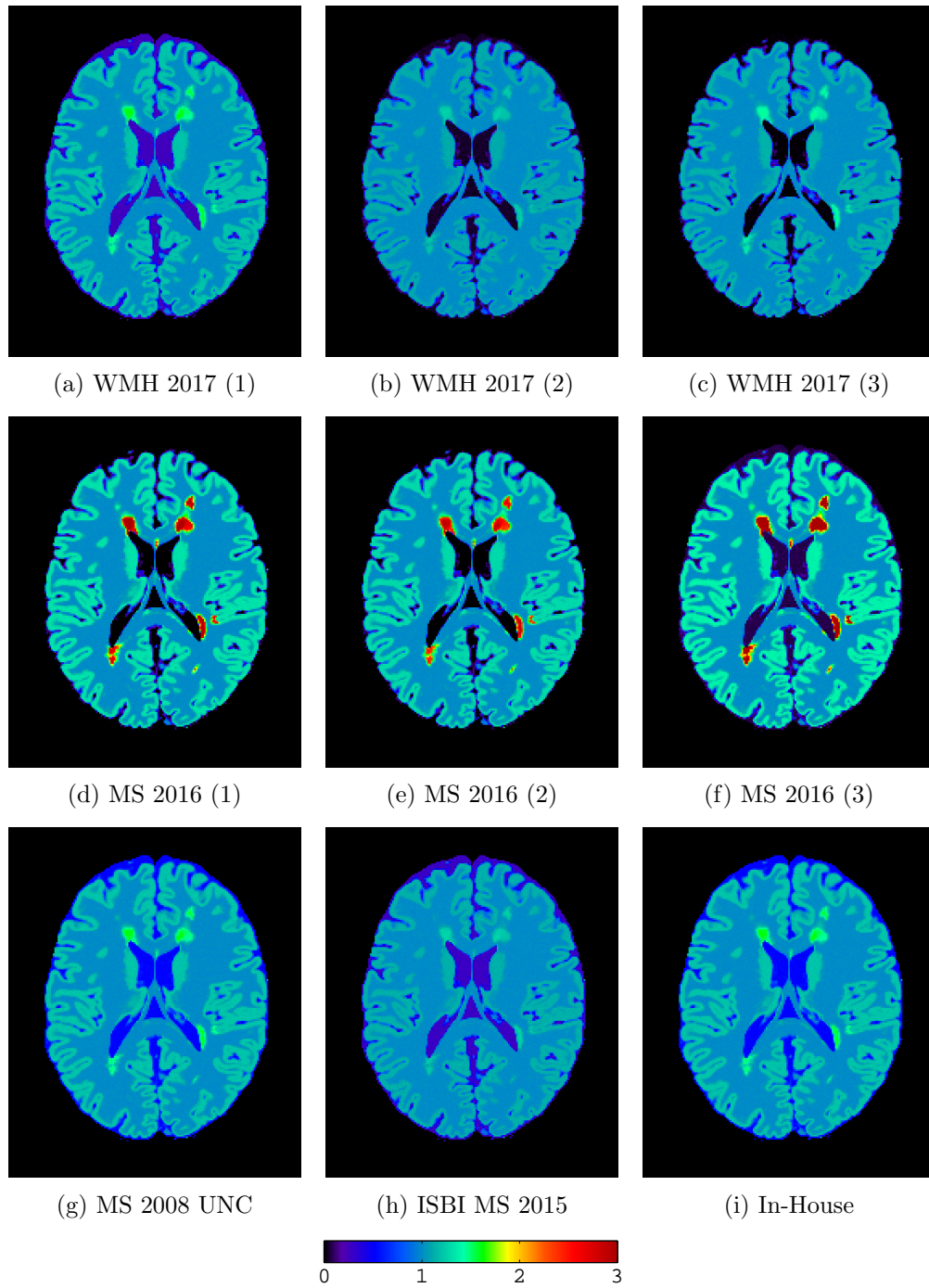Figure A.1: Simulated FLAIR images using scan parameters from the experimental database. Colourmap is arbitrary but consistent.

Appendix B

Implementation

This appendix contains implementation details.

B.1 Computing

All computation in the current work was performed using the following workstation and software:

- **CPU:** Intel Core i7-6700K 4.00 GHz
- **RAM:** 16.0 GB DDR4
- **GPU:** NVIDIA GeForce GTX 980 Ti
- **OS:** Windows 10 Pro
- **Code:** MATLAB R2011a

B.2 Parallel Model Fitting

Speed of model fitting is a significant factor during development, particularly considering optimization of so-called hyperparameters. Faster training yields more model iterations, which inevitably bear improvements. While the estimation procedure outlined in § 2.2.1 must be repeated for all standardized voxels in the brain mask, it is possible to do this in parallel, since every estimation is independent. To do so, the training data must first be vectorized with respect to spatial location x, and matrix operations expanded explicitly to accommodate the new dimension.

To begin, the standardized training data from all subjects – features $\tilde{\boldsymbol{\mathcal{y}}}(x)$, with $\tilde{\mathcal{y}}^0 = 1$, and labels $\mathcal{C}(x)$ – are sampled from nonzero locations in the brain mask $M(x)$. These data are stored in two matrices $\mathbb{Y}$ and

$\mathbb{C}$, with dimensions $[V, N, K+1]$ and $[V, N, 1]$, respectively, where V is the total number of nonzero voxels in the brain mask, N is the number of subjects, and K is the number of features. A similar matrix is constructed for the initial parameters $\boldsymbol{\beta}^{(0)}(x)$, denoted $\mathbb{B}^{(0)}$, with dimensions $[V, 1, K+1]$. Let $\mathbb{Y}_n^k$ denote the vector of data from all voxels for the kth feature from the nth subject, and so on for $\mathbb{C}$ and $\mathbb{B}$.

In order to simplify subsequent calculations, the feature data are rectified according to the class labels, before the first iteration, as in

$$\mathbb{Y}_n^k = \begin{cases} +\mathbb{Y}_n^k, & \mathbb{C}_n \geq 0.5 \\ -\mathbb{Y}_n^k, & \mathbb{C}_n < 0.5 \end{cases}, \qquad \forall\, k \in \{1, \ldots, K\}. \tag{B.1}$$

Next, for a given iteration t, the following vector-compatible expansions of Equations (2.10), (2.11), and (2.13) yield the desired update matrix $\Delta \mathbb{B}^{(t)}$. These calculations are performed in `logreg.m`. Regarding notation: 1) the iteration index $^{(t)}$ is omitted for clarity, 2) element-wise multiplication is denoted by $\circ$, and 3) the variable K is now defined as 1, since this is essential to the simplification.

$$\mathbb{S} = \frac{1}{1+e^\eta}, \qquad \eta = \mathbb{B}^0 + \left(\mathbb{B}^1 \circ \mathbb{Y}^1\right) \tag{B.2}$$

$$\mathbb{A} = \mathbb{S} \circ (1 - \mathbb{S}) \tag{B.3}$$

$$\mathbb{G} = \nabla_{\mathbb{B}} \mathcal{L} - \lambda \mathbb{B}$$
$$= \begin{bmatrix} \mathbb{G}^1 \\ \mathbb{G}^2 \end{bmatrix}$$
$$= \begin{bmatrix} \sum_{n=1}^N \left(\mathbb{Y}_n^1 \circ \mathbb{S}\right) \\ \sum_{n=1}^N \left(\mathbb{Y}_n^2 \circ \mathbb{S}\right) \end{bmatrix} - \lambda \begin{bmatrix} \mathbb{B}^1 \\ \mathbb{B}^2 \end{bmatrix} \tag{B.4}$$

$$\mathbb{H} = \nabla_{\mathbb{B}}^2 \mathcal{L} - \lambda \mathbb{I}$$
$$= \begin{bmatrix} \mathbb{H}^{1,1} & \mathbb{H}^{1,2} \\ \mathbb{H}^{2,1} & \mathbb{H}^{2,2} \end{bmatrix}$$
$$= \begin{bmatrix} \sum_{n=1}^N \left(\mathbb{A} \circ \mathbb{Y}_n^1 \circ \mathbb{Y}_n^1\right) & \sum_{n=1}^N \left(\mathbb{A} \circ \mathbb{Y}_n^2 \circ \mathbb{Y}_n^1\right) \\ \sum_{n=1}^N \left(\mathbb{A} \circ \mathbb{Y}_n^1 \circ \mathbb{Y}_n^2\right) & \sum_{n=1}^N \left(\mathbb{A} \circ \mathbb{Y}_n^2 \circ \mathbb{Y}_n^2\right) \end{bmatrix} - \lambda \begin{bmatrix} 1 & \\ & 1 \end{bmatrix} \tag{B.5}$$

$$\mathbb{D} = \det \mathbb{H}$$
$$= \left(\mathbb{H}^{1,1} \circ \mathbb{H}^{2,2}\right) - \left(\mathbb{H}^{1,2} \circ \mathbb{H}^{2,1}\right) \tag{B.6}$$

$$\Delta \mathbb{B} = -\mathbb{H}^{-1} \mathbb{G}$$
$$= \frac{1}{\mathbb{D}} \left[\left(\mathbb{H}^{2,2} \circ \mathbb{G}^1 - \mathbb{H}^{2,1} \circ \mathbb{G}^2\right) \quad \left(\mathbb{H}^{1,2} \circ \mathbb{G}^2 - \mathbb{H}^{1,1} \circ \mathbb{G}^1\right) \right]^T \tag{B.7}$$

B.3   Manual Segmentations

It was necessary to create and edit a small number of binary segmentation masks during this work. To do this, the Editor module from the 3D Slicer imaging platform [164] was used[1], including the Wand, Paint, and Erase functions. Figure B.1 shows the user interface during a lesion segmentation.
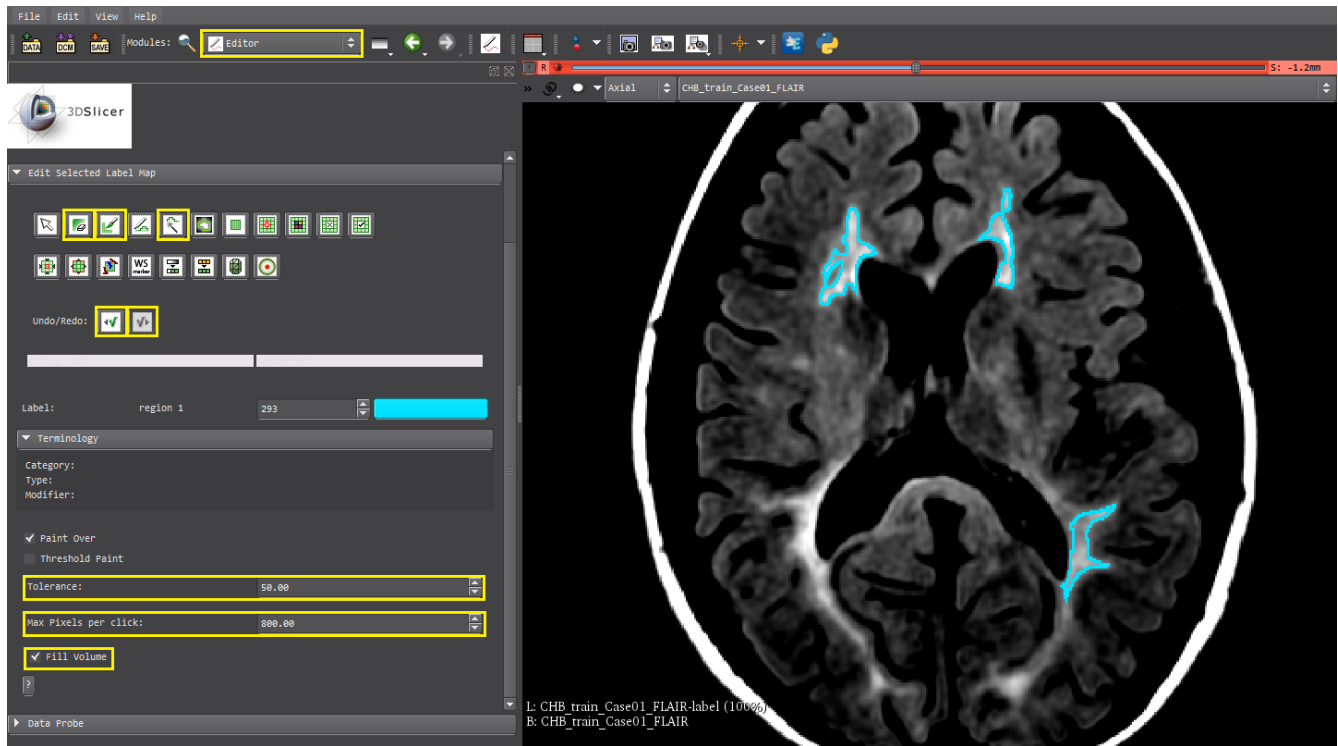

Figure B.1: 3D Slicer user interface for performing in-house manual segmentations and revisions. The tools used are highlighted in yellow, while the segmentation (in-progress) is shown in blue.

B.3.1   MS 2008 WMH Masks

Since the reported performance of an automatic segmentation algorithm depends on the manual segmentations to which it is compared, it is important to obtain good manual segmentations. Unfortunately, the original manuals in the MS 2008 Segmentation Challenge contained obvious artifacts and inconsistencies, as shown at left in Figure B.2. Therefore, it was deemed necessary to redo these manuals. The resulting revisions are shown at right in Figure B.2.

---

[1] 3D Slicer Editor tool documentation is available here: https://www.slicer.org/wiki/Documentation/4.6/Modules/Editor.
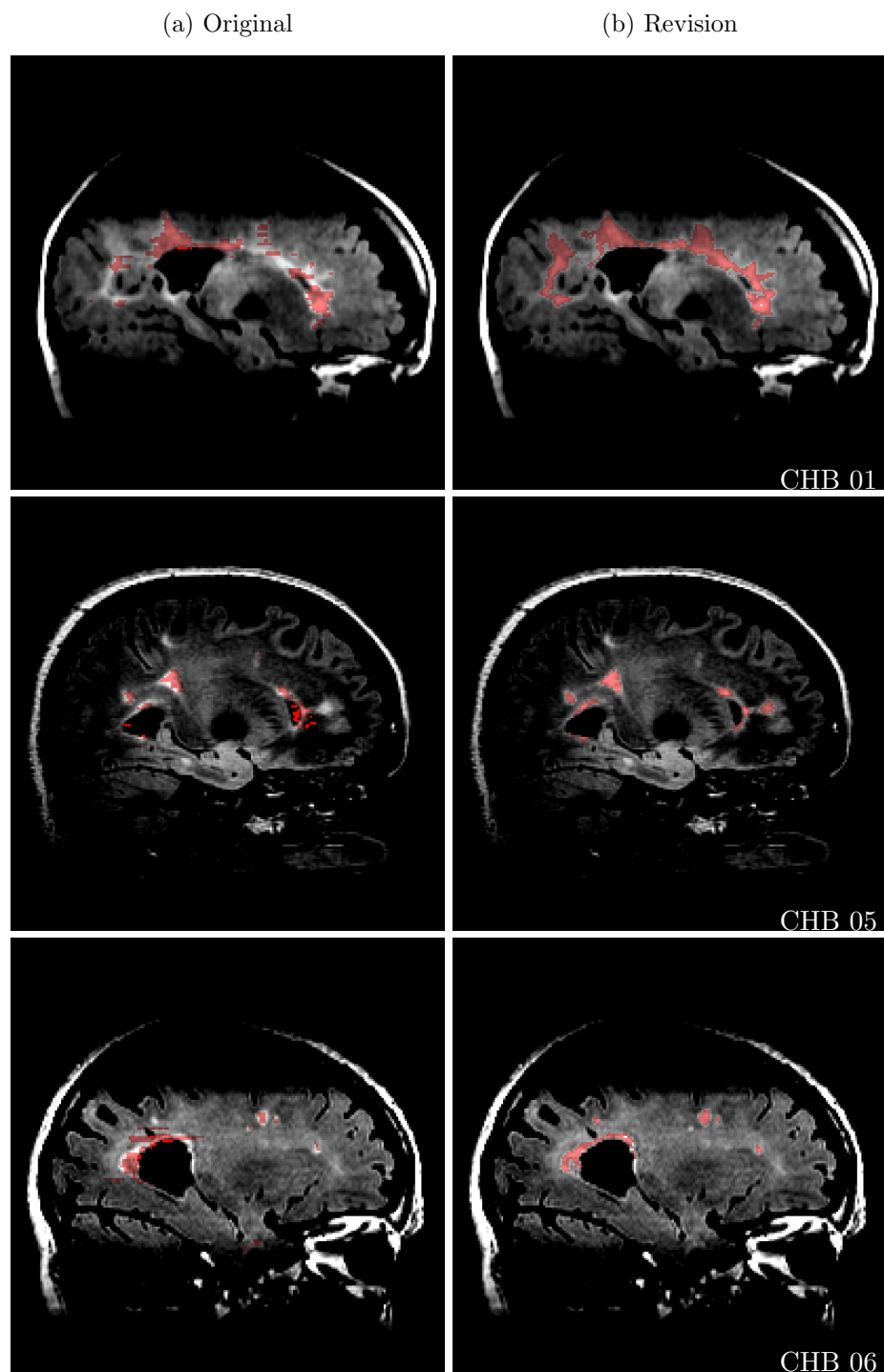
Figure B.2: Example revisions to the manual segmentations for the MS 2008 challenge dataset.
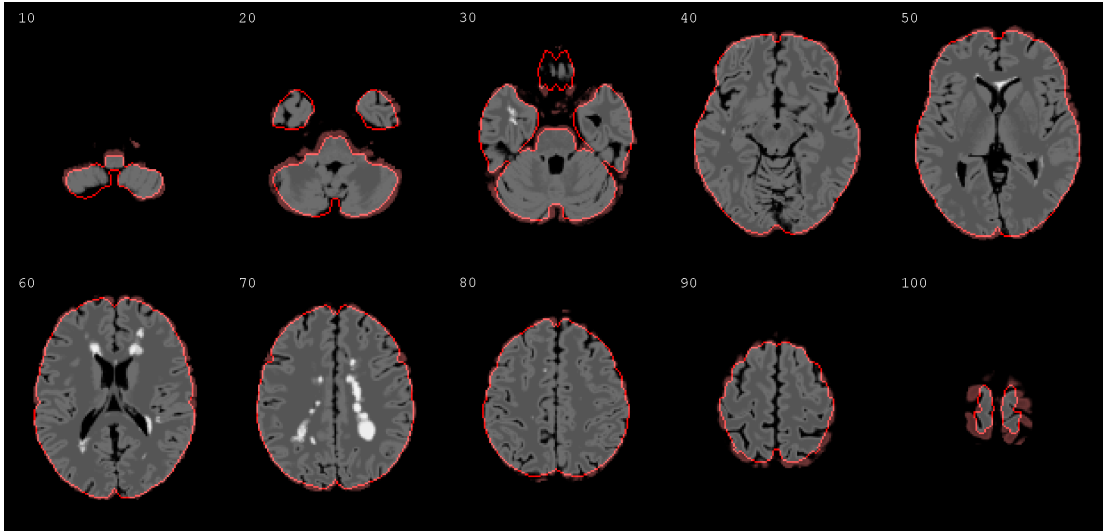
Figure B.3: Manually refined brain mask in MNI space. Mask outline is highlighted in red; inclusions are shown in grayscale; exclusions are tinted red.

B.3.2   Brain Mask

In order to vectorize image data for parallel processing, a binary mask selecting voxels of interest in standardized space is also required. Since only voxels in the brain are of interest, this mask is called a "brain mask". The brain mask used here was derived from the ICBM tissue prior images [131] in MNI space: after initial thresholding of the combined GM + WM + CSF probabilities at 0.5, manual refinements were completed and symmetry was enforced. The mask is slightly small on purpose, since tissues outside the brain are frequently bright in FLAIR images, and can be mistaken for lesions by naive models. The resulting mask is shown in Figure B.3.

Appendix C

Code

```
>> hypdef.m
    function [h] = hypdef(h)
    % flag-like names
    h.name.key      = 'v5-WMH';   % h.name.key      = 'LPA';
    h.name.data     = 'mni96';    % h.name.data     = 'mni109';
    h.name.cv       = 'nocv';     % h.name.cv       = 'nocv';
    % scanner parameters
    h.cmap          = NIH3;
    h.Ni            = [];
    h.scan.idx      = [1,2,3,4,5,6,9]; % h.scan.idx      = [1,2,3,4,5,6,9,7,8,10];
    h.scan.clr      = rainbow7;        % h.scan.clr      = rainbow10;
    % sampling parameters
    h.sam.fresh     = 1;
    h.sam.resize    = 0.5;
    h.sam.dx        = kernelshifts(binsphere(1));
    h.sam.flip      = 1;
    % grey standardization parameters
    h.std.type      = 'pdfeq';
    h.std.pdf       = 'gauss';
    h.std.mask      = 'brain';
    % logistic regression parameters
    h.lr.Nit        = 30;
    h.lr.B          = [0,0];
    h.lr.alpha      = 1;
    h.lr.reg.la     = 1e-3;
    h.lr.reg.py     = [0,1];
    h.lr.reg.pc     = [0,1];
    h.lr.pad        = [-1.5;1];
    h.lr.pp.filter= @(B)(op23(B,@(B)medfilt2(B,[3,3])));
    % post processing parameters
    h.pp.saveles    = 'lpa';
    h.pp.thr.def    = 0.5;
    h.pp.thr.Nit    = 30;
    h.pp.minmm3     = 5;
    h.pp.mrf.Nit    = 0;
    h.pp.mrf.w      = 0.5;
    h.pp.mrf.q      = 0.8;
    % cross validation and
    h = hypfill(h);
```

```matlab
>> arbiter.m
function [] = arbiter(h)
% =================================================
statusupdate(80);
if nargin == 0, [h] = hypdef(); end
statusupdate([h.name.full]); statusupdate();
statusupdate(80);
% -------------------------------------------------
statusupdate('loading training data...');
[h,Y,C] = gettrainingdata(h); t = [];
% =================================================
for c = 1:numel(h.cv.N)
  statusupdate(80);
  % -------------------------------------------------
  statusupdate(c,numel(h.cv.N)); statusupdate();
  [idx] = isplit(h,c);
  % -------------------------------------------------
  statusupdate('computing regression...'); statusupdate();
  [o.B{c},o.thr(c)] = trainlogreg(h,idx,Y,C);
  % -------------------------------------------------
  statusupdate('measuring test performance...');
  [o,t] = performancebat(h,o,t,idx);
  %[o] = performancetest(h,'loso',o,idx);
  % -------------------------------------------------
  statusupdate('saving...'); statusupdate();
  save(h.save.name,'h','o','t','-v7.3');
end
statusupdate(80);
% =================================================
statusupdate('summarizing results...'); statusupdate();
summarizeresults(h, o, t);
% -------------------------------------------------
statusupdate('done');statusupdate();
statusupdate(80);
% =================================================

function [h,Y,C] = gettrainingdata(h)
if h.sam.fresh
  [h,Y,C]   = loadtraining(h);
  ho.M      = h.M;
  ho.std.M  = h.std.M;
  ho.sam.Mr = h.sam.Mr;
  ho.sam.i  = h.sam.i;
  save(h.save.sam,'-v7.3','Y','C','ho');
else
  statusbar(h.Ni,1,h.Ni/3,1);
  load(h.save.sam,'Y','C','ho');
  h.M      = ho.M;
  h.std.M  = ho.std.M;
  h.sam.Mr = ho.sam.Mr;
  h.sam.i  = ho.sam.i;
  statusbar(h.Ni,h.Ni,h.Ni/3,1);
end

function [idx] = isplit(h,c)
idx.c       = c;
idx.i.train = h.cv.i ~= c; % training    (img)
idx.i.valid = h.cv.i == c; % validation (img)
idx.s.train = any(bsxfun(@eq,h.sam.i, find(idx.i.train)')); % training    (data)
idx.s.valid = any(bsxfun(@eq,h.sam.i, find(idx.i.valid)')); % validation (data)
```

58

```matlab
  % special case: full batch
  if sum(idx.i.train) == 0
    idx.i.train = idx.i.valid;
    idx.s.train = idx.s.valid;
  end

  function [B,thr] = trainlogreg(h,idx,Y,C)
  [Y,C,idx] = dataregfun(h.lr.reg.py,h.lr.reg.pc,Y,C,idx);
  % compute the regression
  [b] = logreg(h, Y(:,idx.s.train), C(:,idx.s.train));
  % reconstruct the parameter images
  [B] = reconparams(h, b);
  %thr = h.pp.thr.def;
  % find the optimal threshold
  statusupdate('optimizing threshold...');
  [thr] = thropt(h,Y,C,b);
```

>> logreg.m

```matlab
  function [B] = logreg(h,Y,C)
  [V,N,K] = size(Y);
  % expand the parameters to volumes if not done already
  B = h.lr.B; % size(B) nees to be [V, 1, 2];
  if size(B,1) == 1;
    B = padarray(shiftdim(B,-1),[V-1,0,0],'post','replicate');
  end
  Y = padarray(Y,[0,0,1],1,'pre');
  S = C; S(C>=0.5) = +1; S(C< 0.5) = -1;
  for k = 1:K+1
    YS(:,:,k) = Y(:,:,k).*S;
  end
  clearvars('S','Y');
  % computing max batch size for GPU
  GPU = gpuDevice(1);
  Nf  = 9;                                  % determined empirically
  Nb  = ceil(4*(V*N*Nf)/GPU.FreeMemory); % 4-byte x [size] x Nf vars
  Nd  = ceil(V/Nb);
  % for each batch
  for b = 1:Nb
    % select the batch indices for GPU
    bi  = ((b-1)*Nd)+1 : min(b*Nd,V);
    % transfer to GPU
    Yb = gpuArray(single(YS(bi,:,:))); % features
    Bb = gpuArray(single( B(bi,:,:))); % beta
    % run the optimization
    statusupdate(b,Nb);
    for t = 1:h.lr.Nit
      Bb = update(Yb,Bb,h.lr.alpha,h.lr.reg.la);
      % statusbar
      statusbar(h.lr.Nit,t,h.Ni/3,1);
    end
    % store back from GPU
    B(bi,:,:) = gather(Bb);
    reset(GPU);
  end
  B = squeeze(B);

  function [Bout] = update(Y,B0,alpha,la)
  % compute the activation
```

59

```matlab
B = padarray(B0,[0,size(Y,2)-1,0],'replicate','post');
S = B.*Y; S = S(:,:,1) + S(:,:,2); % could be sum(B.*Y,3) but CUDA error
S = 1./(1+exp(S));
clear('B');
% compute the Hessian elements
A    = S.*(1-S);
% fuck fuck fuck
%H11 = sum(A.*Y(:,:,1).*Y(:,:,1), 2);% + la.*B0(:,:,1);
%H22 = sum(A.*Y(:,:,2).*Y(:,:,2), 2) + la.*B0(:,:,2);
% fuck fuck fuck
H11 = sum(A.*Y(:,:,1).*Y(:,:,1), 2);% + la;
H22 = sum(A.*Y(:,:,2).*Y(:,:,2), 2) + la;
% fuck fuck fuck
H12 = sum(A.*Y(:,:,1).*Y(:,:,2), 2);
Hd  = H11.*H22 - H12.*H12; % determinate for inversion
clear('A');
% compute the gradient
G1  = sum(Y(:,:,1).*S,2);% - la.*B0(:,:,1);
G2  = sum(Y(:,:,2).*S,2) - la.*B0(:,:,2);
clear('S');
% compute the update
dB = reshape([(H22.*G1 - H12.*G2)./Hd, (H11.*G2 - H12.*G1)./Hd],size(B0));
% apply the update
Bout = B0 + alpha*dB;
```

Appendix D

Resources

Here are some helpful links for WMH segmentation research.

D.1   Toolboxes

D.2   Neuroimaging Toolboxes

D.3   WMH Segmentation Toolboxes

LST: Lesion Segmentation Tool [107, 143] for SPM – http://www.statistical-modelling.de/lst.html

D.4   Prior Work

D.4.1   Competitions & Open Access Data

Name	Conference	Ref.	Link
MS LSC	MICCAI 2008	[79]	http://www.ia.unc.edu/MSseg/
Longitudinal MS LSC	ISBI 2015	[57]	http://iacl.ece.jhu.edu/index.php/MSChallenge
MS LSC	MICCAI 2016	[80]	https://portal.fli-iam.irisa.fr/msseg-challenge
WMH LSC	MICCAI 2017	[12]	http://wmh.isi.uu.nl/

"LSC" – Lesion Segmentation Challenge

D.4.2   Prior Work Table

www.uoguelph.ca/~jknigh04/wmlseg/table.html