Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

jung diagnostics

Hochschule für Angewandte Wissenschaften Hamburg
Fakultät Life Sciences

Characterization of White Matter Damage in Multiple Sclerosis using Volumetry and Voxelbased Morphometry

Master Thesis

Biomedical Engineering

Presented by
Jainam Shah
2054321

HAW Bergedorf, Hamburg
31 August 2012

Supervisor: Prof. Dr. Ing. Friedrich Ueberle HAW Hamburg
Supervisor: Dr. Lothar Spies Jung Diagnostics GmbH

The thesis was supervised and created in the company Jung Diagnostics GmbH

Declaration

I hereby declare and confirm that this diploma thesis is entirely the result of my own work except where the references are mentioned. All used resources are explicitly referenced in the bibliography.

Hamburg, 31 August 2012

Jainam Shah

Executive Summary

Aim

The main objective of this thesis was to construct and validate an image processing framework which automatically detects and quantifies multiple sclerosis (MS) lesions in white matter based on high resolution T1-weighted magnetic resonance (MR) images.

Materials and Methods

A dataset consisting of artificially generated T1-weighted MR images of normal brains and brain images having artificial mild, moderate and severe MS lesions on them and their corresponding lesion masks was constructed. Clinical data of subjects with real MS lesions and three databases consisting of normal healthy subjects available at Jung Diagnostics were used in the experiments done during the thesis. The image processing framework was constructed using Statistical Parametric Mapping (SPM), a software suite designed for analyzing brain images. Two distinct biomarkers representing reduction and damage in white matter were investigated using this suite. A voxel-wise comparison of images between two groups of subjects, termed as voxel-based morphometry (VBM) was used for detecting these biomarkers (lesions) in brain images. VBM needs image preprocessing in order to make statistical comparisons between images efficiently. Certain image preprocessing modules of SPM were optimized for the same. Simulated and real subject brain images were used for optimizing segmentation routine by varying the parameter settings to find an optimal set of parameters. DARTEL, a high dimensional registration process, can be executed with two different setups, either by performing DARTEL on grey and white matter images individually or simultaneously, termed as independent and joint DARTEL respectively. An investigation was done to find out which setup is more robust. VBM was optimized by controlling the significance level of the statistical test (t-test) and varying the Gaussian smoothing kernel for DARTEL from 0 to 12mm. Volumetry (estimation of volumes) of artificially simulated normal brain images and brain images with MS lesions was done to estimate the effect of lesions on tissue classification. The optimized framework was validated using simulated brain image with severe lesion and severe lesion mask was used as a reference for interpretation. The statistical test (t-test) generates maps of t-values as an output based on

which quantification and visual interpretation can be done. Finally, the optimized framework was implemented on the clinical data of MS patients for lesion detection.

Results

A nice segmentation of tissues was attained for individual subjects at specific set of parameter settings, but no optimal set of parameters leading to a perfect segmentation which suits all the images was ascertained in this experiment. Joint DARTEL setup was found to be more robust. The statistical test demonstrated a good trade-off between specificity and sensitivity at 0.1 and 0.005 significance level for biomarkers representing reduction and damage in white matter respectively at a smoothing kernel size of 4 and 3 mm for grey and white matter images respectively. Volumetry of normal and lesioned brains showed misclassification of the same amount of white matter as grey matter due to lesion induced low contrast between grey and white matter tissues. Replicating this misclassification in the image processing framework in the form of two biomarkers, VBM performed with simulated brain image using all three databases showed an average accuracy and specificity of approximately 90% and 45 and 55% average sensitivity for biomarker mimicking reduction and damage in white matter respectively. The framework showed high specificity (>95%) with normal healthy subjects also.

Conclusion

The image processing framework was successfully implemented for detecting MS lesions. The framework when tested with simulated data, attained the benchmark of 90% set for accuracy and specificity but did not attain the benchmark of 60% set for sensitivity for reduction and damage in white matter. Thus, more work is needed to be done to make the current image processing framework more efficient in detecting MS lesions.

Table of Contents

List of Figures

List of Tables

Abbreviations

AD	Alzheimer's Dementia
CNS	Central Nervous System
CSF	Cerebrospinal Fluid
DARTEL	Diffeomorphic Anatomical Registration Through Exponential Lie algebra
EMS	Expectation Maximization Segmentation
FLAIR	Fluid Attenuated Inversion Recovery Sequences
FN	False Negative
FP	False Positive
FSL	FMRIB Software Library
FWHM	Full Width at Half Maximum
GM	Grey Matter
MNI	Montreal Neurological Institute
MP-RAGE	Magnetization Prepared Rapid Acquisition Gradient Echo
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
MSIF	Multiple Sclerosis International Federation
NMR	Nucleus Magnetic Resonance
PD	Positron Density
RF	Radio Frequency
ROI	Region Of Interest
SBD	Simulated Brain Database
SPM	Statistical Parametric Mapping
TE	Echo Time
TN	True Negative
TP	True Positive
TPM	Tissue Probability Map
TR	Repetition Time
VBM	Voxel Based Morphometry

WHO	World Health Organization
WM	White matter

1. Introduction

Motivation and Aim

Multiple sclerosis (MS) is considered as one of the most common neurological disorder [(1)]. It is a chronic autoimmune disease of the central nervous system which affects brain, spinal cord and optic nerves [(2)]. It usually affects people between the ages of 20 and 50 and is one of the most common causes of nontraumatic disability among young and middle-aged people. MS is twice as common in women as in men. Today, more than 2.1 million people are affected by multiple sclerosis worldwide [(3)].

Every case of MS is different from the other and so are the symptoms. Some of the symptoms of MS are blindness in one eye, double vision, painfulness, absence of coordination and fatigue [(4)]. In some cases the symptoms of MS are characterized by periods of relapse and remission while a progressive pattern is observed in others.

The World Health Organization (WHO) and the Multiple Sclerosis International Federation (MSIF) carried out a major research to determine the global epidemiology of MS. As a part of this research 112 countries were surveyed from 2005 to 2007 representing 87.8% of the world population [(5)].

None of the countries that participated in the survey was free of multiple sclerosis confirming MS to be a global disease. The prevalence estimates of MS are 30 per 100,000 globally. More white people are affected by MS worldwide than non-whites. Results of the survey confirmed that there are strong geographical patterns influencing the disease and that the frequency of MS varies by geographical region throughout the world, increasing with distance from the equator in both hemispheres [(5)].

Fig. 1.1 shows the geographical prevalence of multiple sclerosis worldwide as in 2007. Different color ranges indicates prevalence of MS in that country per 100,000 people. The median estimated prevalence of MS globally is 30 per 100,000 with a range of 5 to 80 [(5)].

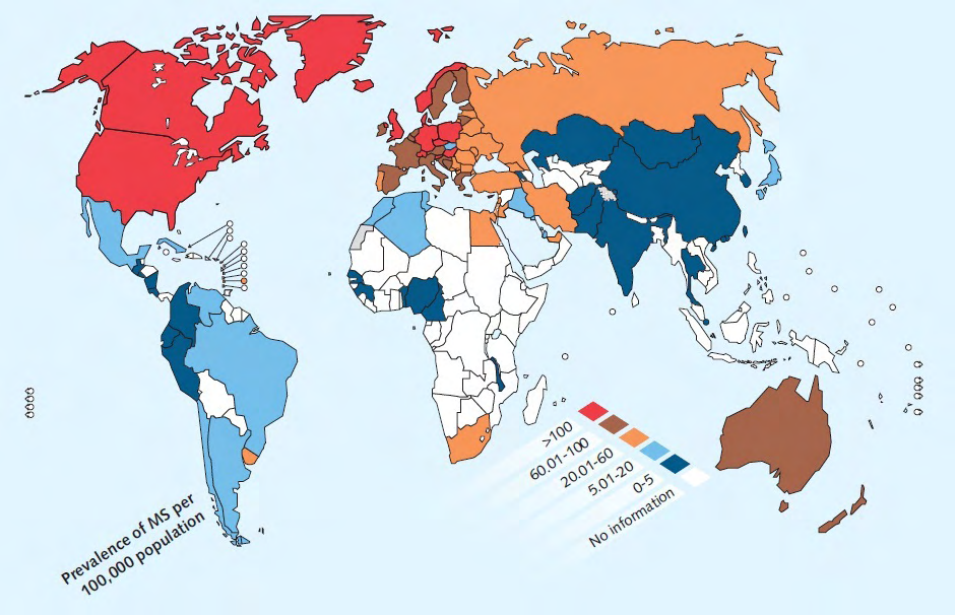


Fig. 1.1 Prevalence of MS globally [5]

Currently, there is no cure for multiple sclerosis, only treatment for MS is available. The treatment does not repair the destroyed nerve cells but slows down the MS process which successively reduces the neurodegeneration rate [6]. Steroids and drugs are available as a treatment for MS symptoms. Ongoing research on new drugs requires imaging techniques to get an insight about the progress of drug development and its effectiveness.

Magnetic resonance imaging (MRI) is an excellent imaging technique for examining parenchymal brain tissue. MRI is very effective in detecting lesions and their progression. Visual interpretation of MRI scans of brain alone doesn't provide concrete diagnosis about MS. A combination of MRI scans and computer aided techniques can prove to be a very helpful tool in detecting MS lesions more efficiently. The aim of this thesis work was to construct an automated image processing framework which can work efficaciously in characterizing MS lesions using MRI scans and computer aided techniques. This involves use of image processing tools and their optimization.

2. Multiple Sclerosis

2.1. Definition

Multiple sclerosis is a chronic disease of the immune system affecting the bodily functions depending on which area of the brain is affected [7]. MS mainly affects white matter of the brain destroying the myelin sheath. Myelin is a dielectric fatty material covering the axon which forms the white matter of brain due to its white appearance. Myelin sheath insulates and encases axons of nerve fibres in the central nervous system and helps in quick and smooth transmission of electrical impulses between the brain and rest of the body. When myelin is destroyed (as shown in *Fig. 2.1*) or in other words, demyelination occurs, electrical signals from the nerve fibres are sent slowly and less efficiently. Patches of scarred tissue form over the affected areas of the brain, further disrupting nerve communication. The sites where myelin is lost, i.e. the site of plaques or lesions appear as scarred or hardened tissues. In multiple sclerosis these scars appear at different areas of the brain and spinal cord. The term sclerosis literally means pathological hardening or scarring of tissue and thus, multiple sclerosis meaning many scars [7]. The symptoms of MS occur when the brain and spinal cord nerves no longer communicate properly with other parts of the body.

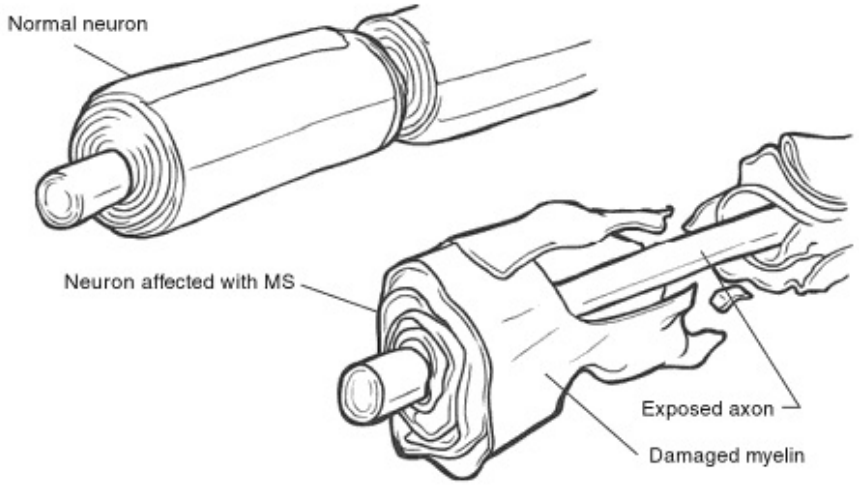


Fig. 2.1 Healthy neuron and neuron damaged by MS [2]

2.2. Etiology

Multiple sclerosis is an autoimmune disease i.e. cause of the disease is an attack of immune system of the body itself [8]. The immune system of the body is responsible for detecting and destroying foreign entities such as bacteria and viruses.

In MS, for unclear reasons immune cells of the body attacks and destroy the myelin sheath that insulates neurons in the brain and spinal cord assuming them to be foreign bodies. The immune system becomes too active and sends out specific type of white blood cells that attacks the myelin sheath as if it were a foreign entity [9]. The myelin sheath provides insulation between cells preventing short-circuiting of electrical signals between cells and speeds up transmission of these electrical signals. The damage in myelin sheath hinders passage of these electrical signals causing disruption of communication between the brain and other parts of the body leading to the symptoms of MS. These demyelinated regions appear as plaques.

The progression of symptoms in MS correlates to development of new plaques in the portion of the brain and/or spinal cord controlling the affected areas. The progression of MS can be unpredictable as there appears to be no pattern in the progression and appearance of new plaques.

2.3. Symptoms

Different areas of the central nervous system (CNS) are responsible for different bodily functions. As the areas affected by scarring can be scattered anywhere in the CNS, the symptoms can also vary with respect to the areas affected. As the symptoms are unpredictable depending on the location of brain tissue scarred, all the cases of MS are different from the others and the symptoms also vary from patient to patient [(7)].

Some of the initial symptoms include muscle weakness, loss of coordination and balance, blurred vision in one or both eyes, numbness and tingling. Later symptoms may include paralysis, vertigo, fatigue, speech and swallowing difficulty, loss of bowel and bladder control, cognitive changes such as memory problems, depression, personality changes, etc. [(10)].

Multiple sclerosis is divided into sub-classifications according to the symptom patterns, which are as follows [(11)],

- Relapsing-remitting – this pattern is characterized by acute attacks followed either by full or partial recovery with some symptoms still remaining. Approximately 85% of MS cases begin with this pattern.
- Secondary Progressive – this pattern of disease begins with relapsing-remitting course and enters into a progressive phase with a slow and steady overall worsening. Approximately 50% of patients with relapsing-remitting course enter in a progressive phase.
- Primary Progressive – in this pattern there is a slow progression of symptoms from the onset of the disease without attacks but with occasional and temporary minor improvements. This pattern is more prevalent in patients having an onset of MS after the age of 40 and more often in males. About 15% of patients are initially diagnosed with this pattern of MS.
- Progressive-relapsing – this pattern of disease shows a progression from the onset. The symptoms may or may not recover at all.

2.4. Diagnosis

A neurologist makes diagnosis of multiple sclerosis. A detailed medical history of the patient, symptoms and complaints is collected, followed by a physical and neurological examination. MS shares some of its symptoms with other diseases; therefore a search to exclude such symptoms is necessary. Certain laboratory tests in addition to taking a medical history and physical and neurological tests are done to confirm the diagnosis of MS. Evoked potential tests, lumber puncture and magnetic resonance imaging (MRI) scans are the tests which confirm the diagnosis of MS [(12)].

Evoked potential tests are measurements of electrical responses to stimulation of different sensory pathways. Three types of evoked potential tests were done in past [(13)].

- Sensory evoked potentials – by applying a small electrical charge on the arm or leg.
- Visual evoked potentials – the patient sits in front of a screen on which alternating check-board pattern is displayed.
- Brainstem auditory evoked potentials – the patient hears a series of clicks in both the ears.

In multiple sclerotic condition, levels of immune protein in cerebrospinal fluid (CSF) increases. This level of proteins is measured by lumber puncture or spinal tap technique. In this technique a needle is inserted below the end of spinal cord and a sample of CSF is extracted for clinical tests [(14)].

MRI is the most sophisticated diagnostic tool which offers a non-invasive imaging technique. MRI scans allow detecting structural and functional changes in brain matter over the course of disease. Diagnosis of MS can be confirmed with two or more lesions detected in white matter using MRI [(11)]. Number of lesions and their volumes can also be measured using MRI images. MRI scans also reveal the progression of disease by comparing new scans with the previous ones. Thus, MRI scans can be used to design a treatment for patients by modifying their drug dose at different stages of the disease [(15)].

Though, there is no test which definitively confirms the presence of MS, a second helpful aid in confirming the suspicion of MS is a group of tests. McDonald's criteria are used for confirming the suspicion of MS. McDonald criteria were introduced in 2001 by Ian McDonald et al. As there is no single clinical or diagnostic test which concludes to certain diagnosis of MS, McDonald criteria includes a combination of clinical and paraclinical tests [(16)].

These criteria include the use of magnetic resonance imaging techniques. A revised version of these criteria was published in 2010. The reasons for revision of the criteria were simplification of demonstration of CNS lesions detected in MRI scans and also the previous revision did not apply appropriately for non-Western Caucasian populations [(17)]. The criteria are included in *appendix*. Thus, MRI scans are used as imaging biomarkers. *Imaging Biomarkers* are described in detail in the section *MRI as Imaging Biomarker*.

2.5. Treatment

The cure for multiple sclerosis is not yet known to man. A cure for MS would mean a complete recovery of the damage occurred to the central nervous system. But there are treatments and approaches to reduce the symptoms of MS [(6)]. Though, the treatment does not repair the damaged nerve cells, but it slows down the disease process. Current treatments reduce the inflammation which in turn slows down the neurodegeneration rate which is responsible for long-term progression of disabilities.

Some of the drugs which can reduce the disease symptoms and progression are Avonex, Betaseron and Copaxone [(18) (19)]. Avonex and Betaseron are forms of interferon beta – 1a and 1b respectively, which are forms of immune system protein found in human body; and copaxone is glatiramer acetate [(20) (21) (22)]. Glatiramer acetate is a synthetic protein that simulates myelin basic protein. It is a component of the myelin that insulates nerve fibers in the brain and spinal cord [(22)]. These drugs have shown effective results in reducing the rate of relapses in the relapsing-remitting pattern of the disease. Avonex slows the progress of physical impairment, Betaseron aids reducing the severity of symptoms and Copaxone decreases disability.

2.6. MRI as Imaging Biomarker

The term *Biomarker* can be defined as a detectable biological phenomenon, which is either structural and/or functional [(23)]. An *Imaging Biomarker* is any physiological, anatomical or metabolic change/s detectable with one or more imaging techniques which helps in establishing the presence and/or severity of a disease. Imaging biomarkers provide a faster and more efficient way for diagnosing diseases and developing new drugs also [(24)].

Magnetic resonance imaging (MRI) is a well-known technique used to detect structural and functional changes in the body. Using MRI as an imaging biomarker structural changes occurring in the brain during the course of multiple sclerosis can be detected. Magnetic resonance (MR) images can also be used to detect the rate of neurodegeneration and brain matter volume changes.

Mainly four imaging modalities in MR imaging are used, T1-weighted, T2-weighted, fluid attenuated inversion recovery sequences (FLAIR) and proton density (PD) [(25)].

T1-weighted images are generated using gradient echo or spin echo sequences with short echo time (TE) and repetition time (TR). The signal acquisition depends on the fat content of the tissues. In T1 weighted images fluid appears darker than the fat. Subcutaneous fat appears brighter, thus white matter appears brighter than grey matter. Lesions in T1-weighted images appear darker than the surrounding tissue and are referred as T1-hypointense lesion or *black holes*. T1-weighted images are good for examining anatomical structures. Spin echo and gradient echo are different combinations of RF pulse sequence used in MRI to acquire data to form an image. TE and TR are explained in section *4.2*.

T2-weighted images are generated using gradient echo sequence with long TE and TR. The signal acquisition depends on the fluid contents in the tissues. CSF appears brighter and can be easily detected. T2-weighted images are good for evaluating pathology. Lesions in T2-weighted images appear brighter than the surrounding tissue and are referred as T2-hyperintense lesion.

Proton density or spin density weighted scans are generated using spin echo sequence usually and sometimes with gradient echo sequence also, with short TE and long TR. The signal intensity in PD-weighted images depends on the number of protons in per unit tissue. Tissues

with lower number of protons per unit appear darker than the tissues with higher number of protons per unit.

FLAIR images are similar to T2-weighted images. FLAIR scans are generated using inverse-recovery pulse sequence. In FLAIR images bright signals from CSF are suppressed unlike T2-weighted images. Thus, are good for evaluating regions near CSF filled spaces [25].

All the images used for the thesis are T1-weighted MR images as the normative database available to make comparisons also consists of T1-weighted images.

3. Technological Framework

Following chapter describes the technological framework used for experiments during the thesis. The technological framework used for the thesis work is adapted from the framework built for Alzheimer's Dementia patients as described in the thesis, *Implementation and Optimization of an Image Processing Framework to Evaluate PET/MRI Brain Data for Early Diagnosis of Alzheimer* [(26)]. This chapter gives an overview about the software suite and its modules used for the experiments in this thesis.

3.1. Statistical Parametric Mapping

Statistical Parametric Mapping (SPM) software is a suite of MATLAB functions and subroutines with some externally compiled C routines [(27)]. SPM is free but copyright software, distributed under the terms of the GNU General Public License. Current version of the suite, SPM8, released in April 2009 was used during the thesis. SPM is used as a MATLAB toolbox. SPM refers to the construction and assessment of statistical processes used to test hypotheses about functional imaging data.

The SPM software package has been designed for the analysis of brain imaging data sequences. The sequences can be a series of images from different cohorts, or time-series from the same subject. SPM follows a voxel based approach which consists of image pre-processing, statistical comparison and graphical representation. Images on which statistical analysis is to be done may be from different scanners and in different 3D systems, or may have noise. Thus, images are realigned, spatially normalized into a standard space and smoothed to remove any noise, if present. Using the general linear model (GLM), parametric statistical models are assumed at each voxel to describe the variability in the data in terms of experimental and confounding effects, and residual variability. The general linear model is explained in section *3.7*. Each voxel is assessed with univariate statistics for the hypotheses expressed in terms of the model parameters [(28)].

3.2. Voxel Based Morphometry (VBM)

Brain morphometry means measurement of brain structures and their changes. Voxel-based morphometry (VBM) is a voxel-wise comparison of brain images between two different groups of subjects by which structural changes can be investigated [29]. VBM requires pre-processing of MRI images before the voxel-wise investigation. First step to VBM is segmentation of MRI images and normalization to the same stereotactic space. Normalization is followed by smoothing after which the voxel-wise parametric statistical tests are performed. SPM consists of different modules using which VBM can be performed, which are discussed in further sections.

The process chain for performing voxel-based morphometry as shown in *Fig. 3.1* consists of pre-processing of MRI images. This step includes segmentation of images into grey matter, white matter and CSF. The segmented images are in their native space which should be converted to the standard space used for this thesis work i.e. MNI space. This is done by normalization process which is followed by smoothing. An alternative to normalization and smoothing processes is the DARTEL process. DARTEL is a high dimensional registration process performed to obtain a precise inter-subject mapping for the statistical test.

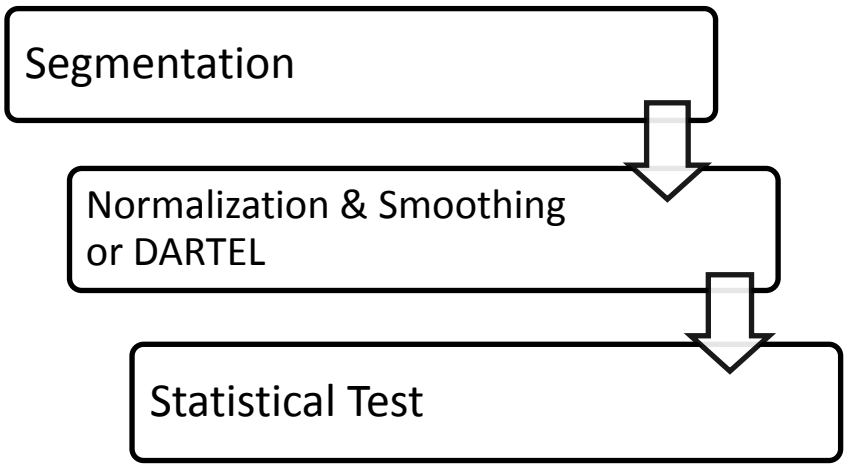


Fig. 3.1 VBM process chain

The major work of this thesis is to devise a framework for VBM of white matter for multiple sclerosis. Several methodologies were implemented and experiments were carried out to setup and optimize this framework using different modules of SPM. In order to setup a consistent methodology to detect lesions in white matter, simulated brain images were used for VBM for which the lesion areas are known in prior and quantified for detectability.

3.3. Segmentation

MR image of a healthy brain tissue can be classified into grey matter, white matter and CSF as shown in *Fig. 3.2*. Segmentation algorithm is used to classify these tissue types. This process is used to distinguish different brain tissue types from each other which make it easier for the statistical test to detect anomalies. Segmentation is based on a modified Gaussian Mixture Model clustering algorithm, which has been extended to include spatial maps of prior belonging probabilities, known as Tissue Probability Maps (TPMs), and also a correction for image intensity non-uniformity that arises in MR imaging [(30)]. Because the tissue classification is based on voxel intensities, partitions derived without the correction can be confounded by these smooth intensity variations.

The model assumes that the MR image consists of a number of distinct tissue types (clusters) from which every voxel has been drawn. The model has approximate knowledge of the spatial distributions of these clusters, in the form of TPMs. Before using the current method for classifying an image, the image has to be in register with the TPMs. Registration is the process of transforming images from one coordinate system to other. Brains of subjects vary in shape and size, thus it is necessary to match them before making any comparisons. This is done using affine registration technique as described in next subsection. After segmentation, the grey matter, white matter and CSF images are transformed into MNI space preserving the amount of volume (explained further in next subtopic). MNI space was defined by Montreal Neurological Institute and thus the acronym MNI was coined. Its coordinates are derived from an average of 152 MRI structural images [(31)].

The names of these images have *'c1', 'c2' and 'c3'* appended to the name of the original image [(32)]. Where, *c1, c2 and c3* are grey matter, white matter and CSF images respectively. Images are in their native space after segmentation.

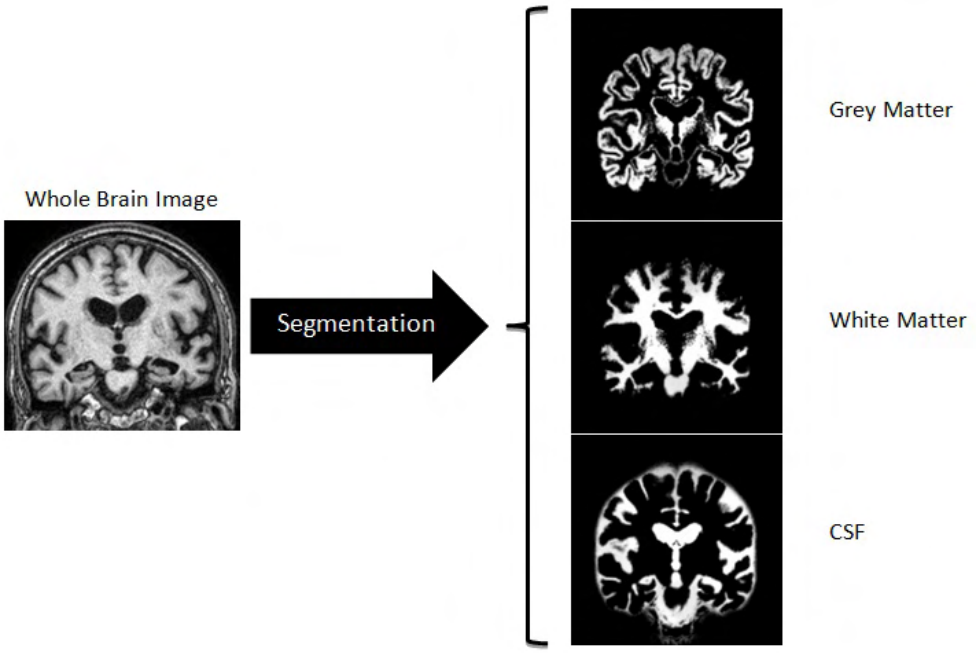


Fig. 3.2 Above figure shows a whole brain image and three different images, grey matter, white matter and CSF image obtained after segmentation

3.4. Normalization

After segmentation the images are still in their native space. Normalization warps the images to the same stereotactic space using a template image. As all the operations in SPM are done in MNI space normalization of images to be analyzed in same stereotactic space (here MNI) is necessary. An example of normalization is shown in *Fig. 3.3*. Cyceron templates were used for this thesis work. The Cyceron templates are computed over a sample of 662 healthy subjects aging from 63 to 75 years consisting of 331 males and females [(33)].

The normalization process of SPM adopts a two-step procedure to determine a transformation that minimizes the sum of squared differences between the voxel intensities of the template and image/s to be normalized [(34)].

The first step involves a linear registration or 12 parameter affine registration and the second step is non-linear transformation or warping. First step determines an affine transformation to match the size, shape and position of the images using translations, rotations, shears and zooms.

The second step involves non-linear registration for further correcting differences in size and shapes which are not accounted by affine transformation described by smooth discrete cosine transform functions. Non-linear transformation changes the volumes of brain regions. For preserving the volumes of each voxel a further processing step, modulation is introduced. Modulation multiplies the voxels of normalized image by their relative volumes before and after normalization.

The name of the normalized and warped image has *'mw'* appended to the name of its original name.

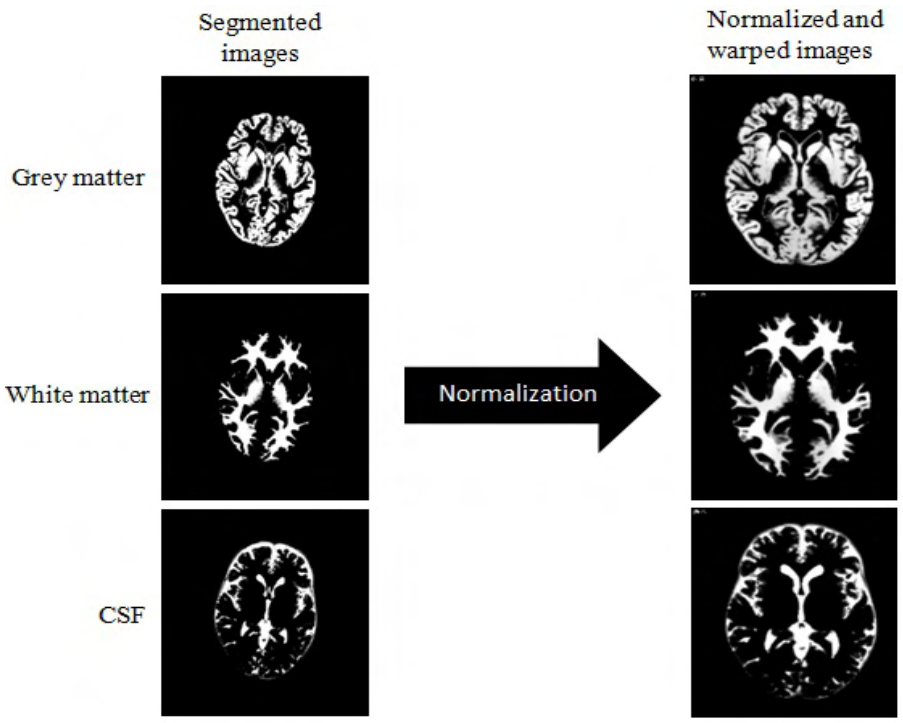


Fig. 3.3 Above fogure shows segmented images of grey matter, white matter and CSF and normalized and warped images.

For the thesis experiments *Unified Segmentation* approach is implemented. This approach enables image registration, tissue classification and spatial normalization in the same model [35]. Adapting unified segmentation algorithm allows a better segmentation and registration process which leads to more promising VBM results.

3.5. Smoothing

Smoothing makes voxel-wise analysis comparable by accounting the anatomical variability and increasing the signal-to-noise ratio [29]. Smoothing effect is the smearing of voxels into their neighboring voxels as shown in *Fig. 3.4*. The smearing effect of smoothing compensates for the registration errors. This in turn increases the sensitivity and specificity of the statistical test. The smoothing kernel is defined by the three dimensional full width at half maximum (FWHM). FWHM of the three dimensional Gaussian filter is a multiple of the voxel resolution. Each voxel of the smoothed image comprises average concentration of the surrounding voxels depending on the size of the smoothing kernel. The central limit theorem says that a distribution of a population will form a normal distribution with a large sample size. Following this principle, smoothing causes the voxel intensities to be normally distributed making the statistical tests more effective [29]. Smoothing also helps in compensating inaccurate normalization.

The name of the normalized and warped image has *'s'* appended to the name of its original name. So the segmented, normalized and warped and smoothed image has *'smwc'* appended to its original name.

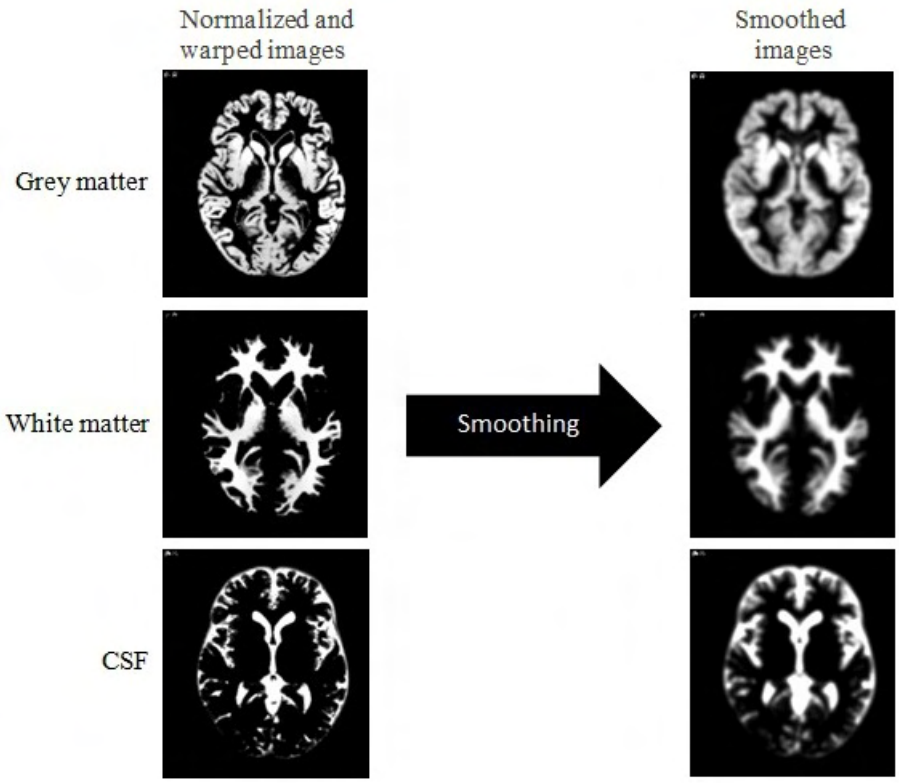


Fig. 3.4 Above figure shows a smoothing effect at 4mm FWHM of normalized and warped grey matter, white matter and CSF images

3.6. DARTEL

DARTEL is a high dimensional, nonlinear image registration procedure. It is the acronym for Diffeomorphic Anatomical Registration Through Exponential Lie algebra. In this procedure, first the mean of all the images to be darteled is created, which is used as an initial template [36]. In the next step, the deformations from this template to each of the individual images are computed and the template is then regenerated by applying the inverse of deformations to the images and averaging. This procedure is repeated a number of times. Finally, warped versions of the images can be generated by normalizing them into MNI space [36]. The step for conversion to MNI space in the toolbox also comprises of smoothing.

DARTEL process can be done independently or jointly. In independent DARTEL grey matter and white matter images are darteled one after the other i.e. the grey matter images are darteled and the output is a single image having grey matter intensities, then the white matter images are darteled for which the output is again a single image comprising of white matter intensity information. In joint DARTEL process the grey matter and white matter images are given as input to the algorithm together and the output is two images, one consisting grey matter information and the other for white matter.

An investigation was done to check which approach out of independent and joint DARTEL is more robust, as described in *appendix*. In this investigation joint DARTEL setup was found to be more robust. For all the experiments henceforth discussed, joint DARTEL setup was used.

3.7. Statistical Test

Once the images are pre-processed, a voxel-wise statistical test can be performed. The test is performed using general linear models (GLM) which allows multiple tests ranging from group comparisons in non-normally distributed data also and detection of regional differences [37]. GLM is a statistical linear model which can be used to implement parametric statistical tests. It is done in two steps, which includes analysis of variance of each voxel and a t statistic from the results for each voxel. The output of this test is an image file, statistical parametric mapping (SPM) or t-map showing the significantly affected regions.

Two sample t-test is used for voxel-wise comparison based on the studies made by M. Mühlau et al. [38]. T-test examines for the difference in means of two populations [39]. As we want to detect anomalies in an individual we compare a single subject to the control group. The image to be examined for anomalies is considered as one population and controls as the other [40] [26]. The null hypothesis H_0 is stated as means of two populations, here, voxel of image to be examined and mean of corresponding voxels of all control images to be equal ($\mu_1 = \mu_2$). Variances of both groups, subject and controls are assumed to be equal ($\sigma_1 = \sigma_2 = \sigma$). If the calculated t-value is smaller than the t-value at the particular selected significance level, the null hypothesis is not rejected and if the t-value is higher, the null hypothesis is rejected [26]. When a true null hypothesis is rejected such an error is called Type – I error. Type – I error is defined by its significance level (α). Some common values of α are 0.005, 0.05, 0.01, etc.

Equation 3.1 shows formula for t-value,

- $$t = \frac{\mu_1 - \mu_2}{\sigma^2 \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (3.1)$$

where, n_1 and n_2 represent the size of populations, subject and control group respectively. As the size of subject is one, $n_1 = 1$. Equation 3.2 shown the formula for t-value for $n_1 = 1$,

- $$t = \frac{\mu_1 - \mu_2}{\sigma^2 \times \sqrt{1 + \frac{1}{n_2}}} \qquad (3.2)$$

The statistical test generates maps of t-values as the output. These maps are images with t-values under null hypothesis of the statistical test.

4. Adaptation of Technological Framework to MS Patients

The technological framework described in previous chapter was created to analyze patients with Alzheimer's Dementia (AD). Age group of patients with AD is higher in comparison to multiple sclerosis patients. AD is a disease of grey matter and multiple sclerosis is a disease of white matter of the brain. Also, the template (Cyceron template as described in section *3.4*) used for normalization of images is defined using older healthy subjects. This technological framework is now being used for multiple sclerosis patients, who are a young group of patients compared to AD patients. Thus, necessary adaptations are made so that the framework can be used for multiple sclerosis patients. This process is described in this chapter.

4.1. Normative Database

Brain scans of normal healthy subjects are used to make statistical comparison. For the experimental work of this thesis, the statistical comparisons are made using one, two or all three of the following healthy subject databases from the databases available at Jung Diagnostics.

All the controls in the databases are healthy subjects. First database consists of 32 subjects, second of 46 and third of 42 subjects. The mean and maximum and minimum ages along with their standard deviations are given in the following table. All the scans are T1-weighted. *Table 1* shows details of the normative databases used during the thesis work,

Table 1 Details of Control Databases

	No. of subjects	**Min. age [years]**	**Max. age [years]**	**Mean**	**Standard deviation**
Database 1	32	22	80	44.58	13.12
Database 2	46	18	54.85	45.63	8.47
Database 3	42	25	74.75	57.20	12.76

4.2. Simulated Image Database

BrainWeb (McConnell Brain Imaging Centre in Montreal) provides ground truth images for the analysis of in vivo data acquired in the form of Simulated Brain Database (SBD). This database contains a set of realistic MR images generated using MRI Simulator [41]. The simulator implements a discrete-event simulation of nuclear magnetic resonance (NMR) signal production, and also models noise and partial volume effects of the image production process using models based on the Bloch equations [41].

MR images of a normal brain and brain with different severities of multiple sclerosis lesions and their corresponding lesion masks are available for download from BrainWeb (www.bic.mni.mcgill.ca/**brainweb**/). The lesions in the brain images with mild, moderate and severe MS lesions are exactly same as the lesions in mild, moderate and severe MS lesion mask images. The dataset consists of 11 images of each brain type i.e. normal brain, brain with mild lesion, moderate lesion and severe lesion. The ground truths for each brain type are also available from the same source. 'Ground truth' is a reference image on basis of which comparisons are made, in other words it can be said that ground truth images are the ideal images of simulation using which comparisons and inferences can be made.

MRI simulator provides several types of parameters to be specified in order to model various acquisition protocols. Following parameters we selected in order to simulate the image quality of MP-RAGE protocol. MP-RAGE is the acronym for Magnetization Prepared RApid Gradient Echo. It is a type of pulse sequence which uses very short TR allowing acquisition time to be less than one second and images almost free from motion artifacts [(42)]. The simulation parameters of the images are as in *Table 2*.

Table 2 Parameters of Simulated Images

Echo time	10ms
Flip angle	30˚
Image type	M
INU field	A
No. of echoes	1
% INU	20%
% Noise	3%
Random seed	0
Reference tissue	0
Scan technique	SFLASH
Slice thickness	1mm
TR	18ms

The simulation parameters are briefly described as below,

- Echo time (TE) is the time from the application of the pulse to the peak of the echo signal in pulse sequences [(43)]. It is defined in milliseconds.
- Flip angle is the angle at which the net magnetization is tilted on the application of external radio-frequency (RF) pulse [(44)]. It is measured in radians.
- Image type is the type of output image constructed by the simulator [(41)]. Either of the three images types is to be selected from imaginary (I), magnitude (M) and real (R).
- INU field and % INU is the intensity non-uniformity parameter. This field is based ton real MR scans to make them more realistic. The value ranges from -100 to 100, a negative value inverts the field [(41)].
- Number of echoes: it is the series of RF rephasing pulses.
- Random seed: it initializes the random number generator used to simulate noise. If the value is specified as zero a new pseudo-random seed is generated every time. The value ranges from 0 to 2147483647 [(41)].

- % Noise: standard deviation of the Gaussian noise applied to the image [41].
- Reference tissue is the tissue used as a reference for the percentage noise calculation [41].
- Scan technique defines the type of pulse sequence. The default type is used for the downloaded data.
- Slice thickness defines the thickness of slices. From 1 to 10mm.
- Repetition time (TR) is the time between consecutive excitation pulses [43]. It is defined in milliseconds.

Table 3 shows the nomenclature for normal, mild, moderate and severe MS lesion brains and their respective lesion masks is as follows,

Table 3 Simulated Database Nomenclature

Image	Assigned Name
Normal brain	t001-000001
Mild lesion brain	t001-000002
Moderate lesion brain	t001-000003
Severe lesion brain	t001-000004
Mild lesion mask	msles1
Moderate lesion mask	msles2
Severe lesion mask	msles3

Fig. 4.1 shows same slice of the images of normal brain, brain with mild, moderate and severe MS lesion. The lesion is highlighted in red.

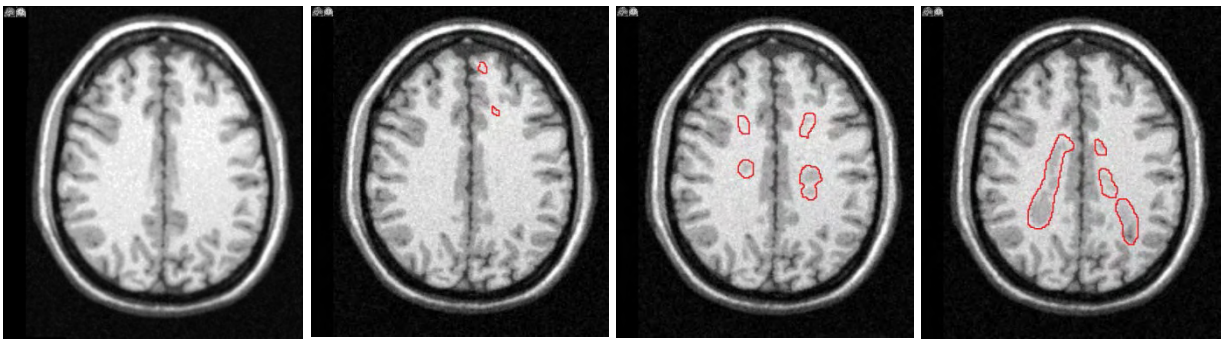


Fig. 4.1 Axial view of MS lesion brains (left to right – normal, mild, moderate and severe MS lesion brain)

Fig. 4.2 shows same slices of mild, moderate and severe lesion masks. Arrow heads show the mild lesions.

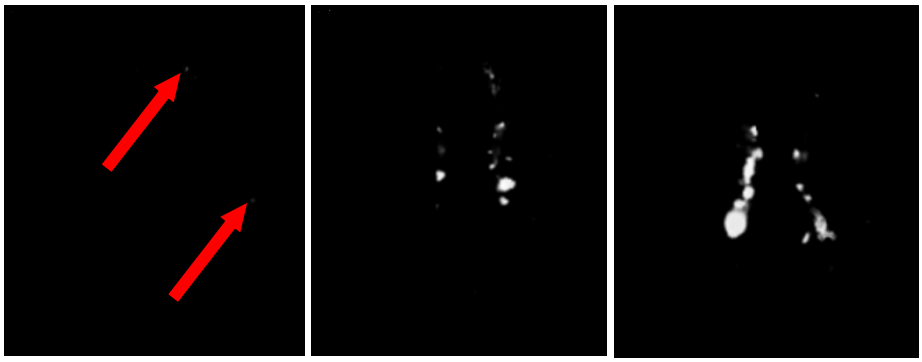


Fig. 4.2 Axial view of MS lesion masks (left to right – mild, moderate and severe MS lesion mask)

There is a variation in the brain images downloaded from BrainWeb database i.e., every time a job is ordered for the same simulated parameters there is a difference in the images due to simulated noise (3%). Therefore, to check the variation in volumes a dataset of 11 images of each brain type is constructed. The relative deviation of total intracranial volume for all brain type images from the ground truth is not greater than 0.21%.

4.3. Performance of Segmentation Engine on Normal Brains

4.3.1. Segmentation Optimization

Introduction

Segmentation engine of SPM classifies brain tissues into either of the three tissue types depending on their tissue probability maps. But this classification of tissues is not accurate, for some instances segmentation misclassifies some portion of tissue into another and/or scalp is classified into CSF. An assumption was made that by optimizing particular value/s of parameter settings there is a possibility that an accurate or near to accurate segmentation can be achieved. To optimize the performance of segmentation engine, volumetry was done on simulated brain images. Volumetry is the estimation of volumes of different brain regions or brain matters or lesions, for e.g. volume of white matter in a subject.

Materials and Methods

Simulated brain data for which ground truth was available and three real patient brain scans were used for optimization of the segmentation routine. For optimization of segmentation routine different parameters were varied from their default values to check which parameter value gives best results (volumes) near to ground truth.

Ground truth volumes of grey matter, white matter and CSF for simulated brain scans were available. Deviation in volumes of the segmented brain tissues of simulated data were compared to their ground truth volumes and for real patient data, volume deviations were compared using the volumes obtained at defaults segmentation parameter settings. A graphical presentation of comparison of volume deviations from their default parameter settings is made to check whether a same pattern in deviation is obtained for simulated and real patient data. A same pattern of volume deviations is expected for each and every brain image for all the parameter settings.

The real patient data consists of a female of 18 years *(s001-000002)*, a male of 43 years *(s001-000123)* and a male of 81 years *(s001-000129)*. The simulated image data is generated using the same protocol but with a variation in noise. *Table 4* shows the ground truth volumes for four simulated brain images.

Table 4 Ground truth (brain tissue volumes)

	Grey matter vol. [ml]	White matter vol. [ml]	CSF vol. [ml]
t000-000001	881.63	654.46	374.68
t000-000002	894.15	667.30	352.22
t000-000012	880.37	686.15	321.66
t000-000014	898.60	680.71	308.61

The parameters varied were warp frequency cutoff, bias FWHM, sampling distance and warp regularization.

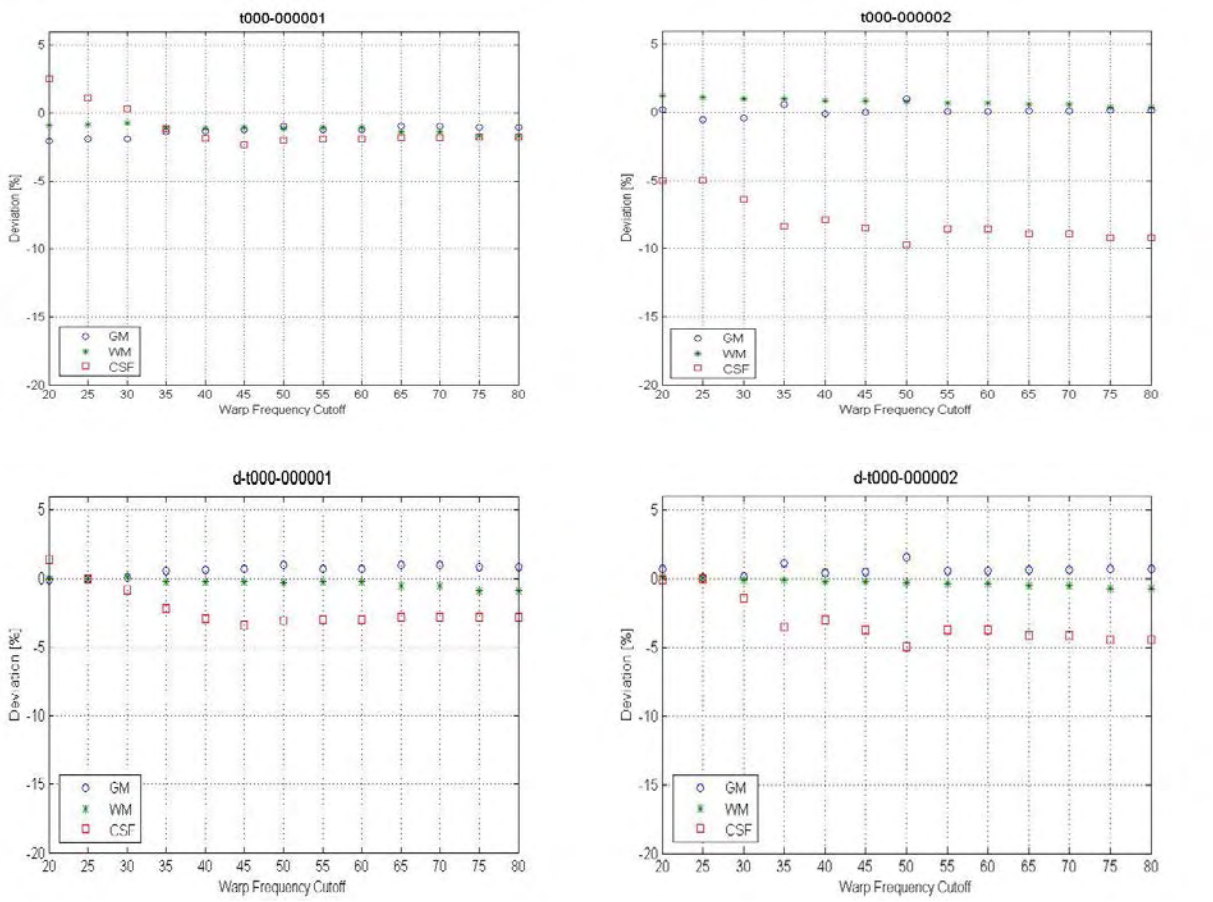
- Warp frequency cutoff. Discrete cosine transforms (DCT) of periods longer than the cutoff are used to describe the warps. A smaller cutoff frequency allows more image deformations to be modeled. Segmentation at smaller values of warp frequency cutoff consumes more time and computational power [(32)]. The default value of this parameter is 25. The experimental values were varied from 20 to 80 at an interval of 5.
- Bias FWHM. It is full width at half maximum of Gaussian smoothness. Bias FWHM models out intensity non-uniformity due to different tissue types. If the intensity non-uniformity is very smooth, then a larger value of bias FWHM is preferable [(32)]. The default value is 60mm. The experimental values were varied from 10 to 190 mm at an interval of 10.
- Sampling distance. “It is the approximate distance between sampled points when estimating the model parameters” [(32)]. Smaller values use more of the data, but the procedure comes with a computational time trade-off. The default value of this parameter is 3mm. The experimental values were varied from 1 to 4.5mm.
- Warp regularization. Intensity variations occur arises due MR physics and different tissue properties. Warping regularization controls the amount of these deformations to be modeled. More regularization allows smoother deformations to be modeled [(32)]. The default value of this parameter is 1. Experimental values were varied from 0.001 to 1000 with a factor of 10.

Results

All the graphs for *t000-000001 and t000-000002* show the deviation of volumes from the ground truth volumes. All the graphs with prefix 'd' and *s001-000002* and *s001-000129* show the deviation of volumes from their respective default values of parameter settings. Results for *t000-000012 and t000-000014* show a similar pattern as *t000-000002.*

Warp frequency cutoff

As shown in *Fig. 4.3*, with the increase in warp frequency cutoff, deviation of the volumes obtained from the ground truth decreases for grey matter, increases for white matter. The volume of cerebrospinal fluid gets stable at a frequency range of 50 to 60 for synthetic data. There are no similar patterns between t000-000001 and other synthetic data files.



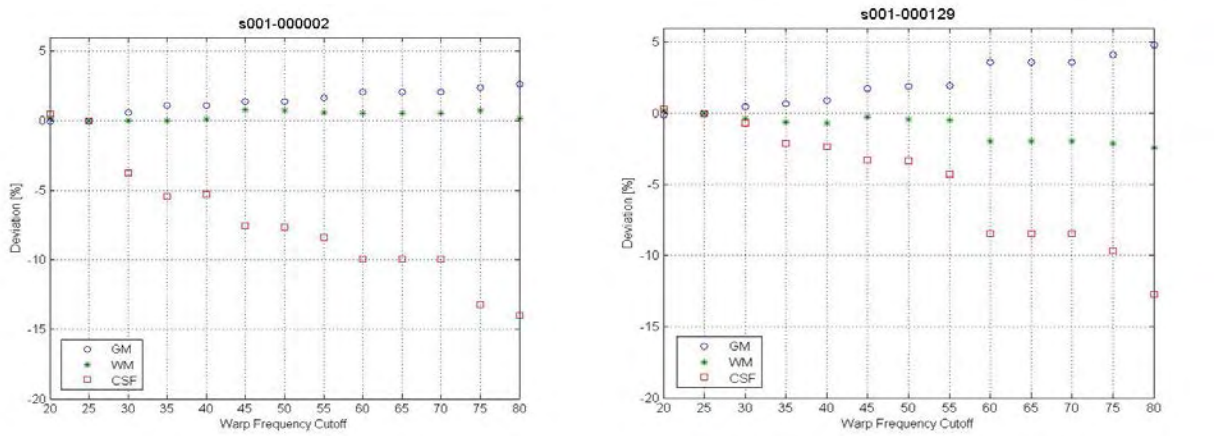
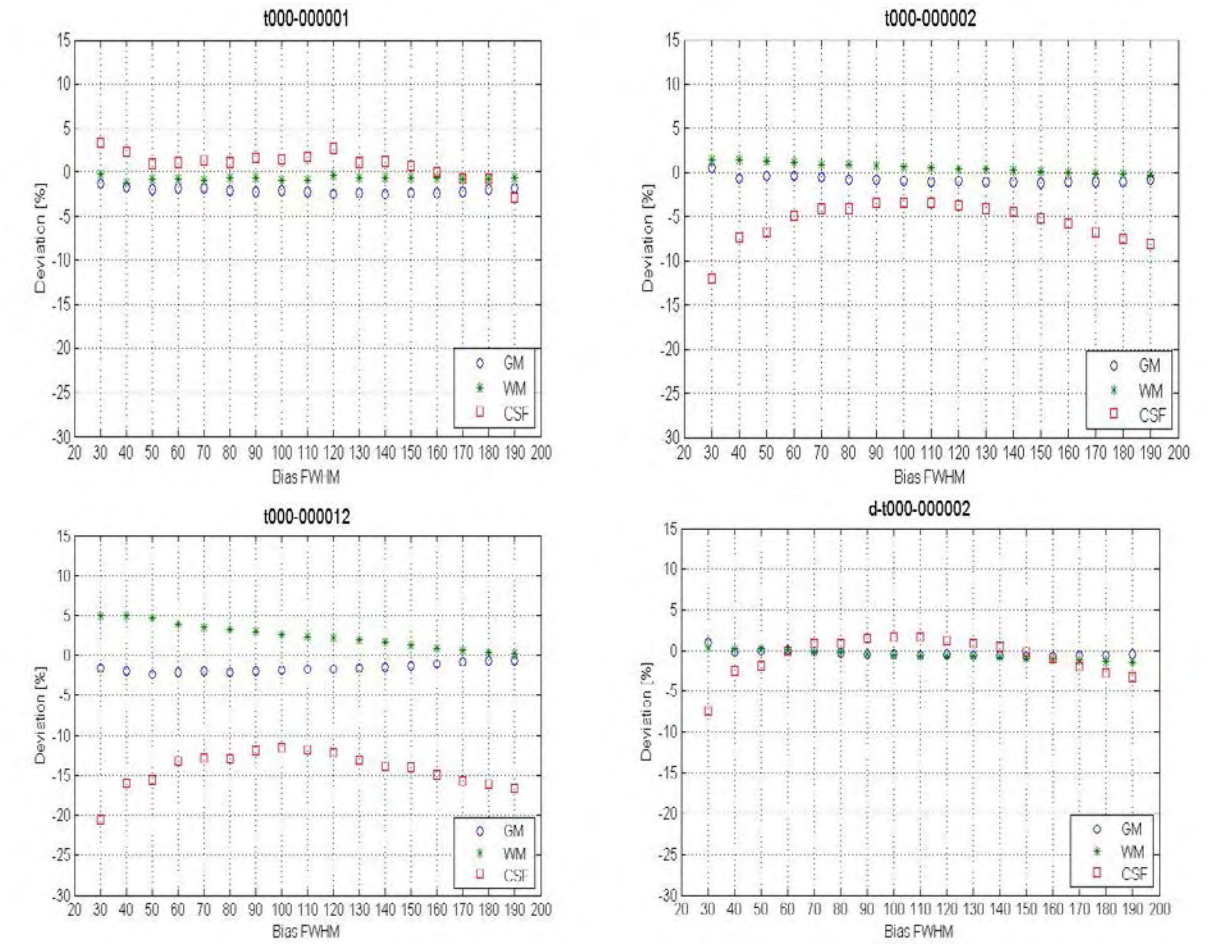


Fig. 4.3 Segmentation optimization using warp frequency cutoff parameter

The volumes of grey matter, white matter and cerebrospinal fluid deviates more from the default parameter setting with the increase in warp frequency cutoff. With deviation of parameters from their default values, the synthetic patient data shows a similar pattern in results, with the increase in the frequency the deviation in results also increases for grey matter, white matter and cerebrospinal fluid. Less deviation in the values of grey matter and white matter is noticed in younger patients *(s001-000002)* than older patients *(s001-000123 and s001-000129).* The patterns for *s001-000123 and s001-000129* are similar.

Bias FWHM

In simulated patient data, when comparing the deviation from ground truth, this parameter has an intense effect on the volumes of cerebrospinal fluid as shown in *Fig. 4.4*. When comparing with respect to the deviation from the default parameter settings, the synthetic patient data shows a similar pattern and a significant change in the cerebrospinal fluid volumes. In real patient data it has a significant effect on the volumes of white matter and CSF. There is no similarity between the patterns of simulated data and real patient data when comparing with deviation of parameters from their default values.



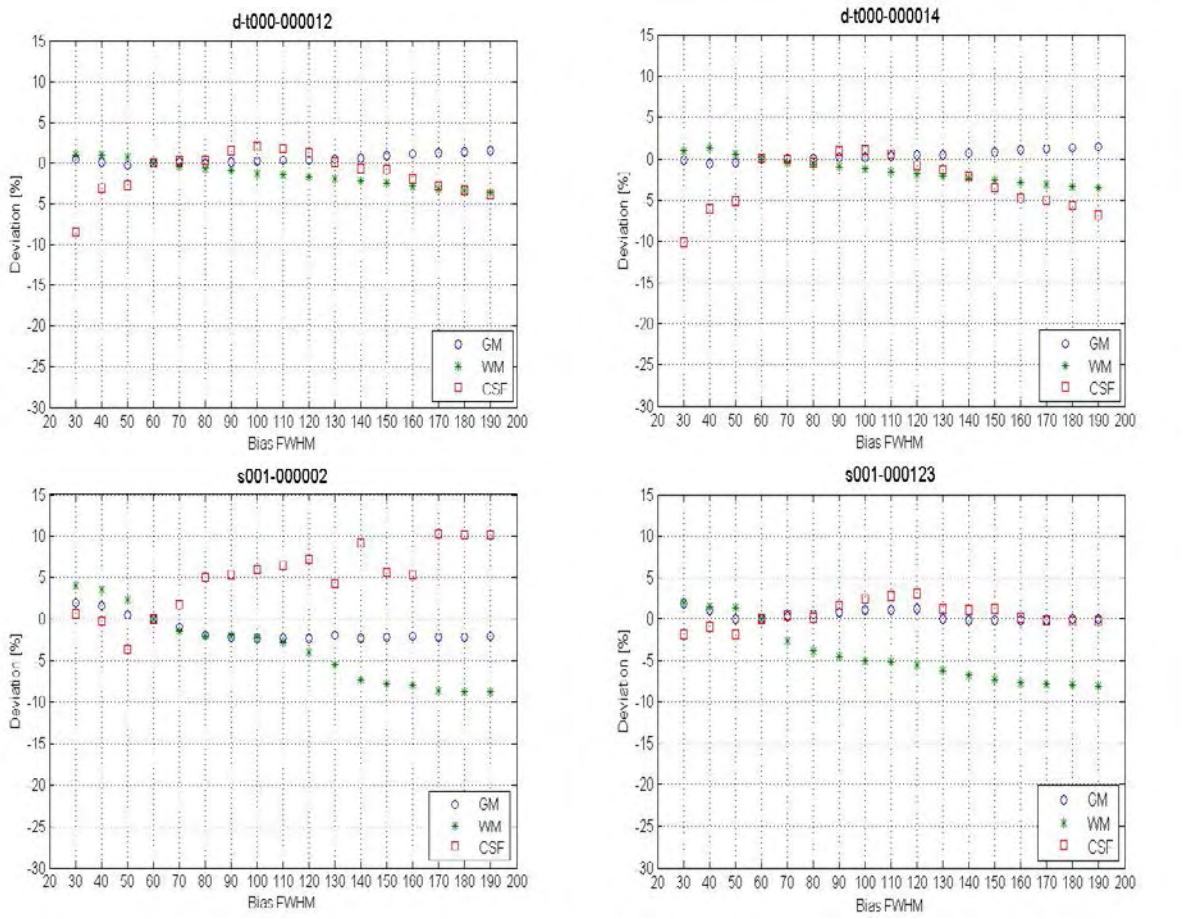


Fig. 4.4 Segmentation optimization using bias FWHM parameter

Sampling distance

It is evident from *Fig. 4.5* that changing sampling distance has an effect on the volumes of cerebrospinal fluid, but the effect on grey matter and white matter is insignificant. When comparing the results with respect to the deviation of parameter settings from default values the synthetic data show dissimilar patterns with major variations in cerebrospinal fluid.

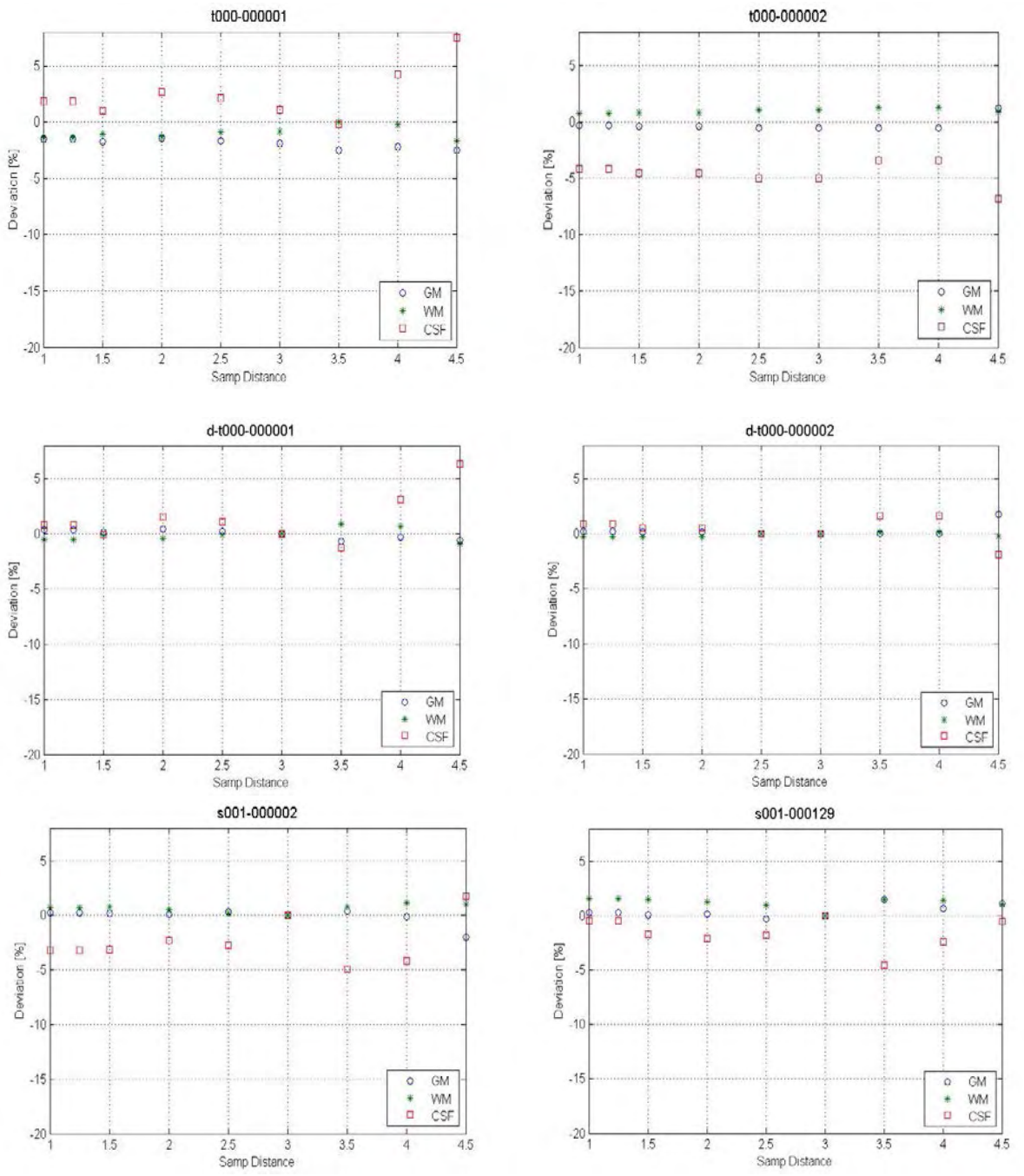


Fig. 4.5 Segmentation optimization using sampling distance parameter

Warp regularization

Altering warping regularization shows a similar pattern on all the patient data as shown in *Fig. 4.6*. It has a significant effect in the volumes of cerebrospinal fluid. Grey matter and white matter volumes do not vary much.

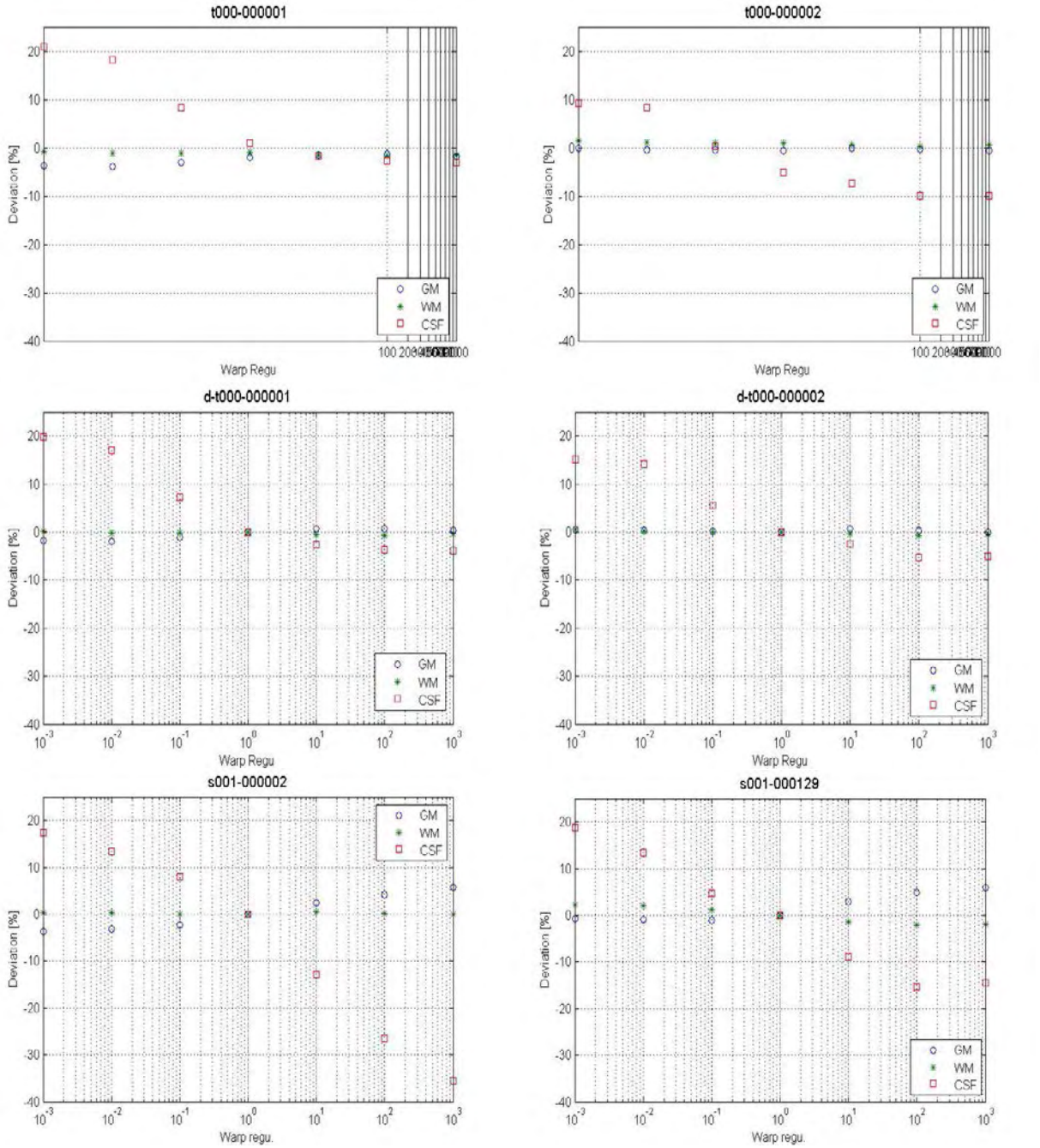


Fig. 4.6 Segmentation optimization using warp regularization parameter

Varying different parameters leads to more unsatisfying results, and thus indicates that the default parameter settings are the best suitable for segmentation. Though, there is a possibility of segmenting MR images perfectly by selecting customized parameter values for individual images, this process is time consuming but leads to better segmentation.

4.3.2. Segmentation Errors

Introduction

Segmentation inaccuracies cause misclassification of grey matter, white matter and CSF in the segmented images. This misclassification can be nullified up to an extent by optimizing the Gaussian kernel for smoothing of images if the inaccuracy in segmentation and registration is known.

Materials and Methods

In order to find the combined segmentation error a ground truth image (phantom image) for grey matter, white matter and CSF obtained from BrainWeb was used with the simulated data from the same source.

Three T1 images from BrainWeb with same simulation parameters are segmented using SPM. The phantom image and segmented images are first binarized and the segmented image is then reoriented in the same dimensions as phantom image using image calculator routine of SPM. A binary image has only two values of all voxels, either 0 or 1. Binarization of image refers to conversion of all voxel values greater than zero, to 1 and all remaining voxels to zero. Images can be binarized at thresholds other than zero also; this is further explained in section *4.4*.

On visually observing both the binarized images after reorientation of the segmented image it was evident that there is a shift in the segmented image. To find this shift a subtracted image is generated as shown in equation 4.1. Maximum shift can be defined by the maximum number of white pixels next to each other.

- $subtracted\ image = |\ phantom\ image - segmented\ image\ |$ *(4.1)*

First image in *Fig. 4.7* is the result of the reoriented and binarized segmented image overlaid on binarized phantom image shown in pink color. The white and red portions in the image are of phantom and segmented images respectively and the pink portion shows overlap of both the images. The white and red portion of the first image is the result of subtraction image in the second image.

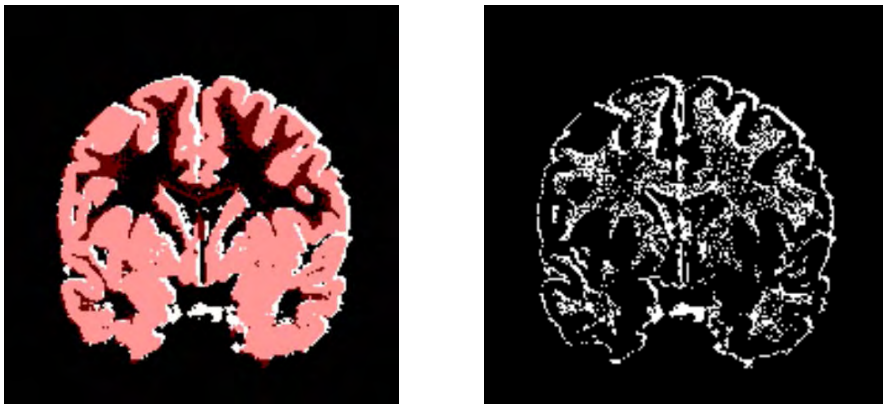


Fig. 4.7 Image on the left is the coronal view of an overlay of phantom and segmented image. Second image is the subtraction image

The voxels in white display a shift which is evident in all three, x, y and z dimensions. This shift is calculated for all planes for all three dimensions. For e.g. for an image with dimensions 181x217x181 i.e. a 3D matrix, if we consider X-axis, then all the voxels orthogonal to X-axis i.e. in the direction of Z-axis are scanned for maximum shift from the first till last coordinate of Y-axis. So, now we have a 2D matrix of Y×Z (217×181) elements with maximum shifts. A mean value is obtained from this 2D matrix which indicates an average shift in X-axis. Likewise, a scan is made for Y- and Z-axis.

Results

Table 5 shows the mean of maximum shift, denoted by '*MMS*', for three images in all three dimensions with their standard deviation, denoted by '*SD*',

Table 5 Registration error

	Image	MMS X [mm]	MMS Y [mm]	MMS Z [mm]	SD X	SD Y	SD Z
Grey Matter	t000-000001	3.56	3.9	3.06	4.41	4.88	3.84
	t000-000002	2.88	3.23	2.46	3.55	3.97	3.02
	t000-000012	2.88	3.23	2.51	3.49	3.96	3.04
White Matter	t000-000001	2.06	2.34	1.81	2.71	3.05	2.38
	t000-000002	2.04	2.3	1.84	2.62	2.9	2.35
	t000-000012	2.52	2.86	2.37	3.1	3.47	2.98
CSF	t000-000001	5.57	6.65	4.87	6.32	8.64	7.41
	t000-000002	5.03	5.99	4.57	5.50	7.58	6.8
	t000-000012	4.72	5.61	4.30	5.11	6.98	6.12

From the above table it can be estimated that there is approximately 4mm shift in grey matter, 3mm in white matter and 7mm in CSF (based on the mean of standard deviations). These segmentation errors can be compensated using a 4mm Gaussian kernel for smoothing grey matter images, 3mm Gaussian kernel for white matter images and 7mm Gaussian kernel for CSF images.

4.4. Effect of MS Lesions on Tissue Classification

Presence of MS lesion in the simulated brain images causes misclassification of brain tissues. The multiple sclerosis lesions in white matter are classified as grey matter due to the matching intensities of MS lesions and grey matter in T1-weighted images [45] [46]. Volumetry of simulated brain images, normal brain and lesion brains was done to check the extent of this misclassification of brain tissues due to mild, moderate and severe MS lesions.

The MS lesion masks, as introduced in section *4.2* are in Talairach space. Talairach space is defined by Talairach & Tournaux in 1988. This space is based on the dissection of a 60-year old French lady [47]. As all the statistical comparisons using SPM are done in MNI space, all lesion masks were first converted into MNI space using normalization algorithm from Unified Segmentation routine. MNI space was defined by Montreal Neurological Institute and thus the acronym MNI was coined. Its coordinates are derived from an average of MRI structural images from 152 individuals [31]. Normalization uses a transformation matrix obtained after segmentation which contains information about affine transformation. Transformation matrix generated after segmentation of one of the simulated normal brain from the dataset of 11 images was used to convert all three (mild, moderate and severe) MS lesions masks. Normalization was done at the default parameters used in unified segmentation other than the transformation matrix as mentioned above.

As there are many small clusters of lesion scattered all over the brain, the region of interest (ROI) was confined to simulate a more realistic lesion in white matter region. Binary masks were generated from the lesion masks which are now in MNI space, at a threshold of 20%, i.e. all voxels having intensity greater than 0.2 in the lesion mask are up scaled to 1 and less than or equal to 0.2 are down scaled to 0. *Table 6* shows the volumes of lesions before and after thresholding.

- $lesion\ mask > 0.2 \rightarrow 1$ *(4.2)*
- $lesion\ mask \leq 0.2 \rightarrow 0$ *(4.3)*

Image Calculator routine from SPM is used to generate the binary masks.

Table 6 Volumes of lesion masks in MNI space before and after thresholding at 20%

Lesion type	Lesion vol. before thresholding [ml]	Lesion vol. after thresholding [ml]
Mild	1.09	1.46
Moderate	6.19	8.81
Severe	12.20	18.18

After thersholding at 20% and generating binary masks, the voxels which were before of low intensities are now up scaled to value 1 making borders of the lesion definite. As a consequence, the lesion volumes increase after thresholding.

The segmentation routine classifies brain image into either of the three brain tissues, grey matter *(c1)*, white matter *(c2)* or CSF *(c3)* depending on their respective tissue probability maps (TPMs). As there is no TPM for MS lesion in the segmentation routine, the intensity of voxels in the lesion area corresponding to that of grey matter up to an extent are classified as grey matter voxels. Therefore, with white matter, some portion of grey matter is also classified as a MS lesion. Therefore, to check the size of this misclassification, volumes inside the lesion area for all brain matters are estimated.

For calculating the volumes inside the lesion area, segmented images of all 11 normal brains were masked with mild, moderate and severe MS lesions and the average of white matter, grey matter and CSF proportions was calculated.

Normal brain images are masked with mild, moderate and severe MS lesion as shown in *Fig. 4.8*,

- $c_n .* (\text{msles1} > 0.2)$, $c_n .* (\text{msles2} > 0.2)$, $c_n .* (\text{msles3} > 0.2)$ *(4.4)*

  where, c_n ($c_n = c_1, c_2, c_3$) are segmented images of **normal brain**

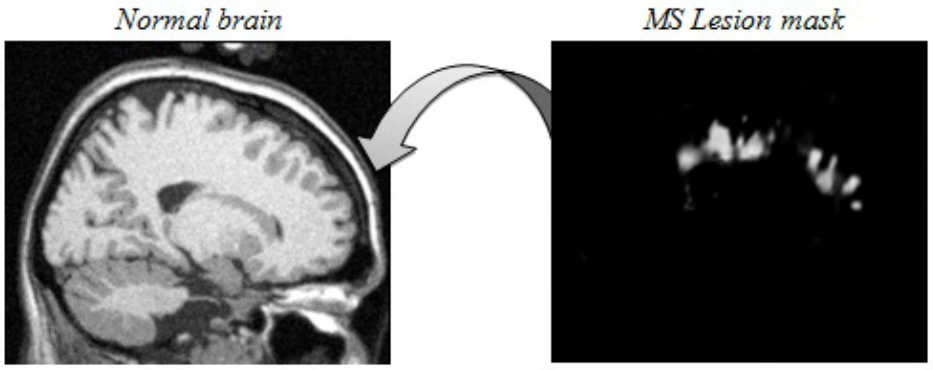


Fig. 4.8 MS lesion masked on a normal brain image

Volumetry of mild, moderate and severe MS lesion brain images is also done masking mild, moderate and severe MS lesion masks respectively as shown in *Fig. 4.9*.

- c_{mild} .* (msles1 > 0.2), c_{mod} .* (msles2 > 0.2), c_{sev} .* (msles3 > 0.2) *(4.5)*

  where, c_{mild} (c_{mild} = c_1, c_2, c_3) are segmented images of **mild lesion brain**
  c_{mod} (c_{mod} = c_1, c_2, c_3) are segmented images of **moderate lesion brain**
  c_{sev} (c_{sev} = c_1, c_2, c_3) are segmented images of **severe lesion brain**

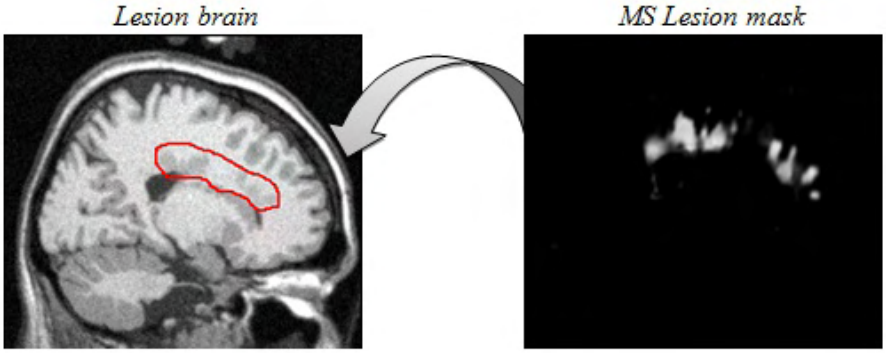


Fig. 4.9 MS lesion mask on a lesion brain image

The region highlighted in red is the lesion area.

In the above equations, '.*' means voxel-wise multiplication of two images. *'msles_x > 0.2'* indicates that only the part of lesion mask image which has intensity more than *0.2* will be taken into account, the remaining lesion is ignored in masking.

Table 7 shows average proportions of white matter, grey matter and CSF volumes inside the lesion region for normal brain when masked with MS lesions.

Table 7 Volumes inside the lesion

Substance / Brain type	White matter [%]	Grey matter [%]	CSF [%]
c_n .* (msles1 > 0.2)	69.80	23.79	6.41
c_n .* (msles2 > 0.2)	80.93	14.44	4.63
c_n .* (msles3 > 0.2)	90.20	6.18	3.62

Table 8 shows average proportions of white matter, grey matter and CSF volumes inside the lesion region for mild, moderate and severe lesion brains when masked with their respective lesion masks.

Table 8 Volumes inside the lesion

Substance / Brain type	White matter [%]	Grey matter [%]	CSF [%]
c_{mild} .* (msles1 > 0.2)	58.68	35.01	6.31
c_{mod} .* (msles2 > 0.2)	65.97	29.37	4.66
c_{sev} .* (msles3 > 0.2)	74.06	22.26	3.68

The values in the above tables indicate proportion of each brain matter calculated from the average of 11 images of their respective type of brain.

Ideally, none of the brain matters should be detected inside the lesion area when a respective lesion mask is applied on the lesion brain, but due to misclassification of lesion all the matters are detected.

The graphs in *Fig. 4.10* demonstrate the absolute difference of white and grey matter volume inside the lesion area; for e.g., for white matter,

$$|\,(\text{vol. of WM for } c_n \,.*\, (\text{msles1} > 0.2)) - (\text{vol. of WM for } c_{mild}.*\, (\text{msles1} > 0.2))\,| \qquad (4.6)$$

The graph on the left is for white matter and on the right is for grey matter. Lesion volumes are 1.46, 8.81 and 18.18 ml representing mild, moderate and severe MS lesions respectively. Standard error of estimate for the fits of grey matter and white matter are 5.80 and 5.77 respectively.

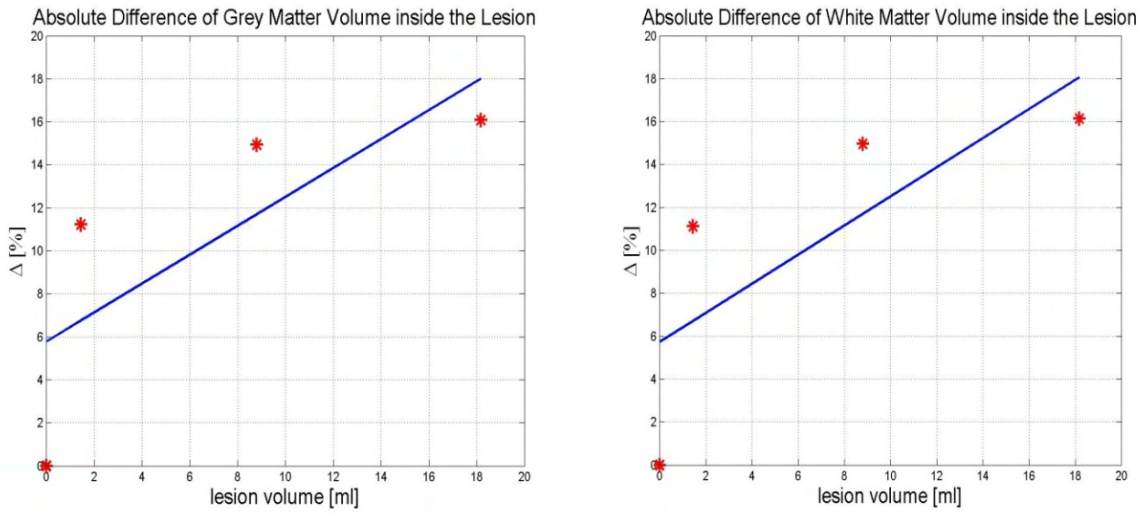


Fig. 4.10 Absolute difference of white matter and grey matter volumes inside the lesion

Fig. 4.11 show similar graphs, but here the linear regression fits are forced to pass through origin as at zero lesion volume the difference in misclassification should also be zero. Standard error of estimate for the fits of grey matter and white matter are 8.21 and 8.15 respectively. Although, standard error of estimate for the fits not intercepting zero are smaller, the later fits give a more realistic impression about the behavior of lesion misclassifications.

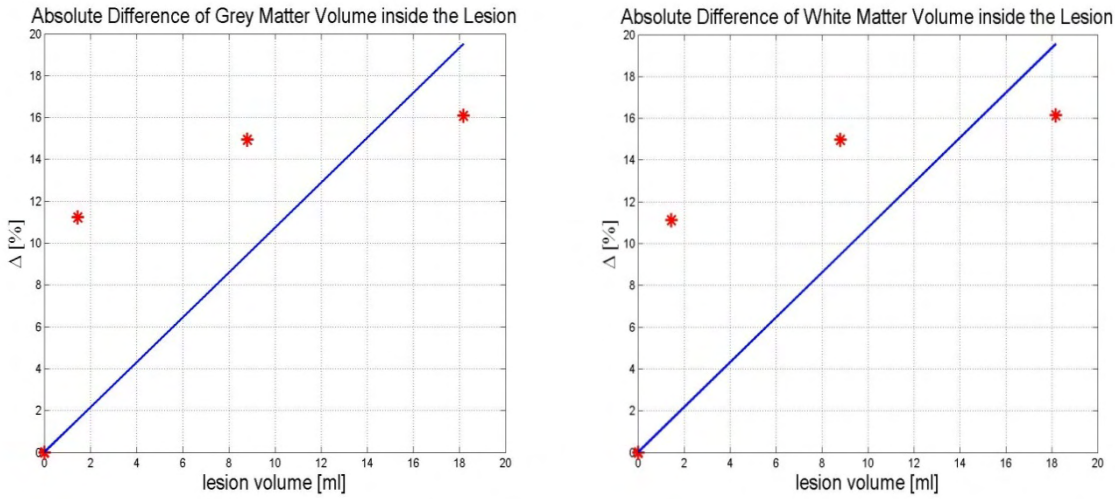


Fig. 4.11 Absolute difference of white matter and grey matter volumes inside the lesion

From the results of volumetry it can be inferred that the effect size (Δ) for white matter is 11.12%, 14.96% and 16.14% for mild, moderate and severe lesion respectively, and 11.22%, 14.93% and 16.08% for mild, moderate and severe lesion respectively for grey matter.

By volumetry of the MS lesion brain images it is evident that there is a misclassification of the lesion due to the segmentation algorithm as there is no lesion template present in segmentation and normalization routine. Based on the linear regression analysis it is observed that the effect size scales with the lesion volume, i.e. greater the lesion volume, bigger is the effect size. Thus, approximately same volume of white matter is misclassified as grey matter. As the effect size scales with the lesion volume, all further investigations are limited to only one MS lesion mask. Severe lesion mask was chosen for later findings as a representative of MS lesion.

4.5. Simulation of Artificial Lesion

4.5.1. Introduction

Segmentation process depends on the intensity differences in order to segment brain tissues. Intensity differences between grey matter and white matter decrease in presence of lesions contributing to misclassification of lesion as grey matter [46]. Therefore, lesions are not detected completely by the statistical test described in section *3.7*. In order to make the lesions more detectable, both white matter and grey matter needs to be examined. Thus, to see the effect of lesion in both white and grey matter, statistical test is run for both brain tissues. Therefore, t-maps are generated for white matter and grey matter components, named as hypo and hyper intensity maps respectively. For this, we attenuate white matter and amplify it in the grey matter with an artificial lesion.

The above described simulation of artificial lesions can be implemented in VBM with two different approaches. In the first approach simulation of lesions is done on pre-processed i.e. normalized and smoothed images and the other in which lesion is simulated on normalized image and then smoothing is done. The first approach is an artificial kind of setup for testing the efficiency of lesion detection by the framework. Whereas, the second approach is more realistic as real patient images already have a lesion on them and then they are pre-processed.

4.5.2. Materials and Methods

In order to generate t-maps (hypo- and hyper-intensity maps), the lesion masks (mild, moderate and severe MS lesion masks) are applied on pre-processed images. Attenuation of lesion on white matter images which means reduction in white matter and amplification of lesion on grey matter images, elevation in withe matter is done.

As multiple sclerosis is a disease of white matter, the lesion should be in white matter region only. Therefore, only white matter should be detected in the area of lesion. But a small volume of grey matter is also detected due to misclassification of the lesion as grey matter. As observed from the results based on the studies made by Renske Boer et al. [45] and the experiment described in section *4.4* (effect of MS lesions on tissue misclassification), the amount of misclassification of grey matter scales to the effect size. Thus, in order to detect the lesion fully in the grey matter we add the same portion of the white matter in the lesion that is missing in the

grey matter image, from the white matter image. Equation 4.8 for amplification mimics the misclassification as explained above.

Simulation of artificial lesions can be implemented by two different approaches in VBM. In the first approach, lesion is applied after normalization and smoothing, simulates very accurate delineation of lesions on the brain.

$$(normalized\ and\ smoothed\ image\ *\ simulated\ lesion)\ \rightarrow\ Statistical\ test$$

As the lesion is masked on the smoothed image, there is a very accurate division of lesion from the normal brain tissue which can be detected easily by the statistical test.

The second approach is slightly different from the first. In this method, the segmented and normalized brain image is attenuated with MS lesion mask and then the normalized image is smoothed, followed by statistical test.

$$(normalized\ image\ *\ simulated\ lesion)\ \rightarrow\ smoothing\ \rightarrow\ Statistical\ test$$

As the brain images are smoothed after the application of lesion mask, the voxels of lesion also gets smeared out with the neighboring voxels and so does the neighboring voxels in the lesion area. As a result of this smearing effect of lesion voxels after smoothing, there is a less accurate delineation of lesion in the test. This method simulates a more real brain image.

Fig. 4.12 shows smoothed (8mm Gaussian kernel) images generated at 90% attenuation with severe lesion mask with both the above described approaches. The lesion in the left image is clearly visible than in the right although both the images are smoothed with 8mm filter size.

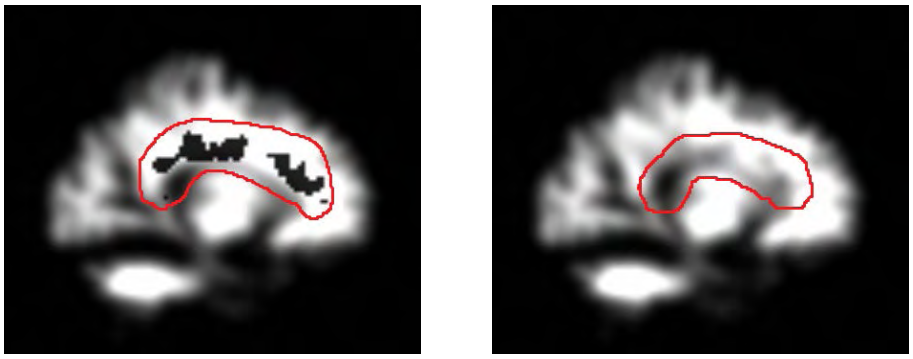


Fig. 4.12 Sagittal view of same brain image simulated with 90% MS lesion. On the left is the image generated with model 1 and on the right with model 2. Region highlighted in red is one of the lesion areas.

Following biomarkers are implemented for both VBM approaches discussed above.

- Biomarker (a): Expression for **attenuation** in Image Calculator is as follows,

$$f = i2 .* (1 - \text{effect size} .* (i3 > 0.2)) \quad (4.7)$$

- Biomarker (b): Expression for **amplification** in Image Calculator is as follows,

$$f = i1 .* (i3 \leq 0.2) + (i1 + (\text{effect size} .* i2)) .* (i3 > 0.2) \quad (4.8)$$

where, i1 = mwc1 image file (normalized grey matter image) or
smwc1 image file (smoothed and normalized grey matter image)

i2 = mwc2 image file (normalized white matter image) or
smwc2 image file (smoothed and normalized white matter image)

i3 = severe lesion mask

i1 and i2 are selected according the approach implemented for the statistical test.

Effect size is the percentage volume of white matter in the lesion. In the experiments for this thesis work the effect size is varied from 0% to 100% with 10% increment.

Equation for biomarker *a,* indicates voxel wise multiplication of smoothed and normalized white matter image and the resulting image generated by the inversion of percentage of MS lesions equivalent to effect size. MS lesion is thresholded at 20%. The equation is explained graphically step-by-step as follows,

Fig. 4.13 shows same slices of white matter of normal brain *(i2)* and severe lesion mask *(i3)*.

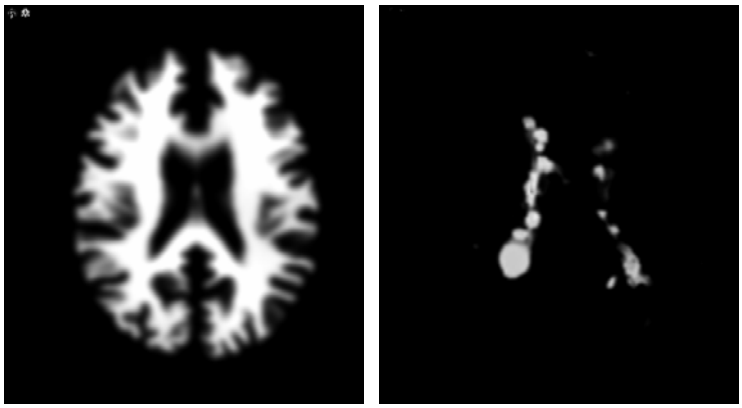


Fig. 4.13 Slice of normal brain image (left) and severe lesion mask (right)

Fig. 4.14 shows same slice of severe lesion mask thresholded at 20% *(i3 > 0.2)* and the same image at an effect size of 50% attenuation *(effect size .* (i3 > 0.2))*, i.e. intensity of the image reduced by 50%.

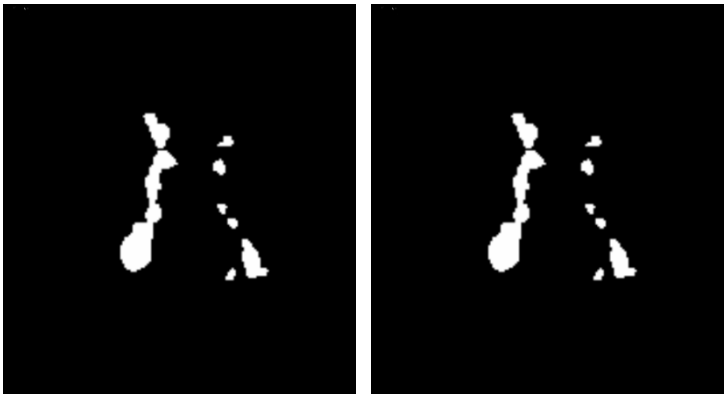


Fig. 4.14 Slice of severe lesion mask at 20% thresholding (left) and 50% attenuation (right)

Fig. 4.15 shows same slice of inverted severe lesion mask after thresholding and attenuation *(1 - effect size .* (i3 > 0.2))* and this image masked with white matter image *(i2 .* (1 - effect size .* (i3 > 0.2)))*.

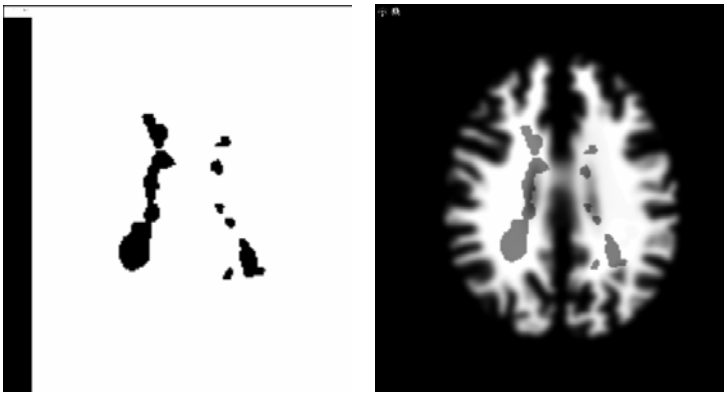


Fig. 4.15 Slices of inversion image of thresholded and attenuated severe lesion mask (left) and white matter masked with the left image (right)

Likewise, equation for biomarker *b*, indicates summation of voxel wise multiplication of smoothed and normalized grey matter image and voxel wise multiplication of MS lesion at 20% thresholding and the resulting image generated by the summation of percentage of white matter image equivalent to effect size and grey matter image.

Fig. 4.16 shows an example of attenuation and amplification of severe MS lesion on white matter and grey matter images of the normal brain at an effect size of 50% each as explained in equations above,

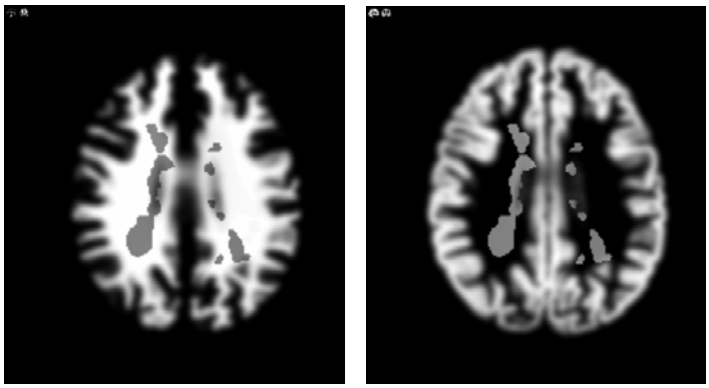


Fig. 4.16 Reduction in white matter (left) and elevation in white matter (right)

After pre-processing and simulation of artificial lesion, statistical test is done. In the t-maps generated as an output of the statistical test, every voxel that is highlighted is not necessarily an actual lesion. There are some voxels which are not in reality lesions but are highlighted, and vice-versa. Thus, accuracy, sensitivity and specificity are the measures used to quantify the efficiency of the statistical test. Quantification is explained in detail in *appendix*. All three measures of quantification should be 100 percent which will indicate the framework to be ideal for detection of MS lesions.

It can be assumed that first approach is more sensitive than the second approach. The assumption was made on the basis of the fact that there is a more accurate delineation of MS lesion in the first model than in the second. To support the above assumption, voxel-based morphometry using database – 1 was done with both methods. Severe MS lesion was used to simulate the lesion on the normal brain image at 0.005 significance level. Smoothing increases signal-to-noise ratio which in turn improves the results of statistical test. So, the second model was also tested at different smoothing kernel sizes to check the best suitable size of Gaussian smoothing kernel that can be used in further experiments.

The above mentioned VBM procedure was performed using the MATLAB program available at Jung Diagnostics on the normal subject database. This program performs a statistical test on the database files against each other, one after the other to characterize group differences. For e.g., consider a database with 10 controls in it, from *controls_001* to *controls_010*. First image data from the database i.e. *controls_001* is considered as the subject file and remaining files *(from controls_002 to controls_010)* as controls for the statistical test and voxel-wise investigation is executed. In the next sequence, second image data from the database *(controls_020)* is considered as the subject file and remaining files *(controls_010 and controls_030 to controls_010)* are considered as controls and statistical test is performed. Thus, each image data is statistically tested against the remaining data. Thus, a mean of quantification measures i.e. accuracy, sensitivity and specificity can be calculated for the individual results of each statistical test. These mean values give an overview about the performance of the implemented VBM approach.

4.5.3. Results

The graphs for mean accuracy, mean sensitivity and mean specificity in the results below and in other experiments henceforth indicate the mean of respective quantity calculated from the results of quantification after statistical test of each image in the database. Attenuation is the effect size (refer section *4.5.2*) at which lesion was simulated. The legend in the mean sensitivity graphs indicates the area under curve.

Fig. 4.17 shows the results of quantification of VBM done at 12mm smoothing kernel. With the second approach 100% accuracy, sensitivity and specificity is not achieved even at 90% attenuation, whereas nearly almost 100% accuracy and specificity, and 100% sensitivity is observed at 40% attenuation with first approach. In the second approach of VBM as the lesion boundaries are smoothed out, statistical test doesn't detect the lesion accurately and thus, low sensitivity is achieved. With the increase in effect size (from 0% to 100%) of simulated lesion the smearing effect of lesion voxels also increases. Therefore, the lesion voxels are smeared out into non-affected white matter region simulating new clusters of lesion. Thus, increases false positives inducing a downward slope in the results of accuracy and specificity in the graphs of second model. Similar observations can be made from the graphs in *Fig. 4.18* which shows the results for both the models at 8mm smoothing kernel.

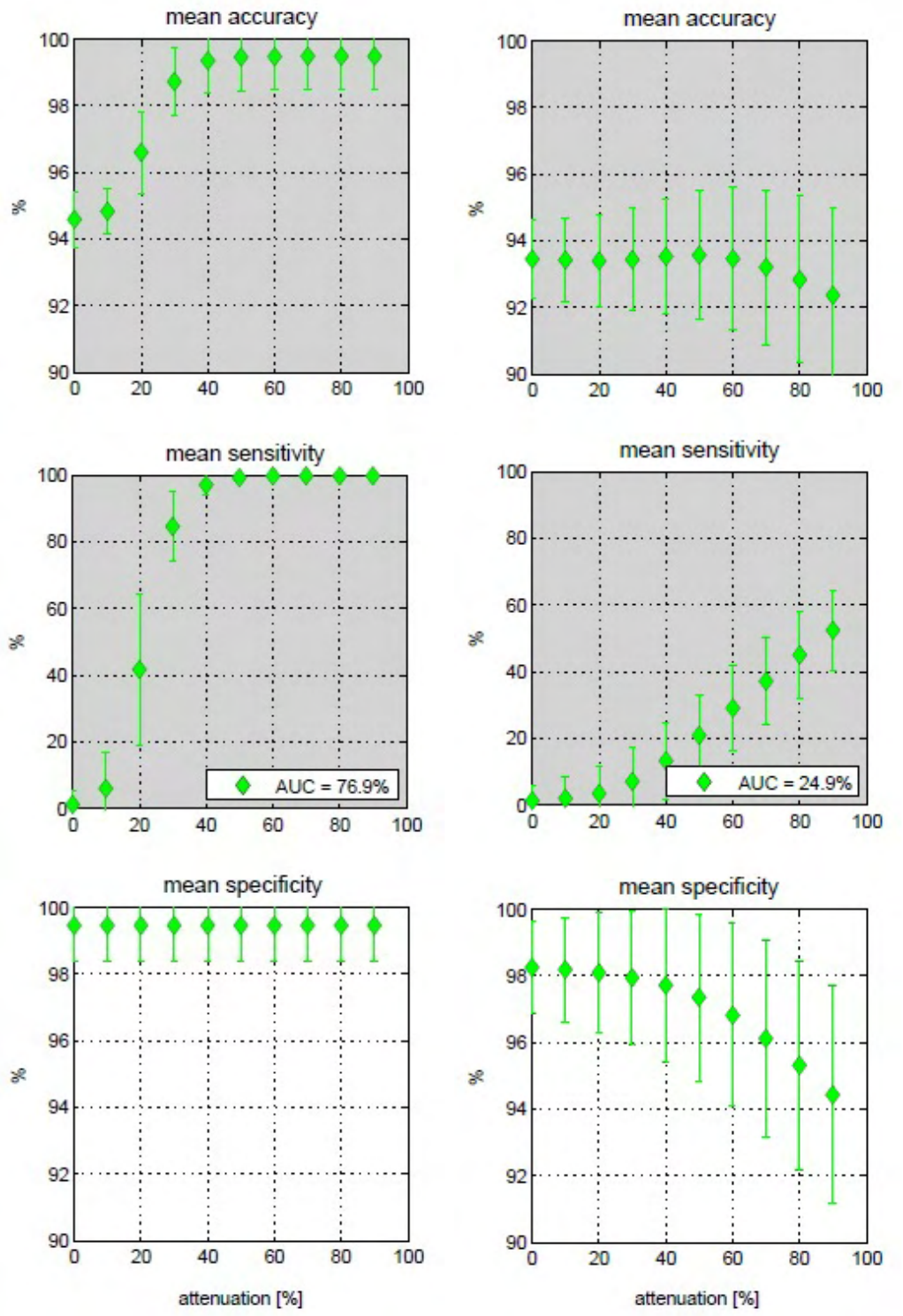


Fig. 4.17 Graphs shows mean sensitivity and specificity and accuracy for both the models at 12mm filter size. Left graph represents results of model 1 and right of model 2

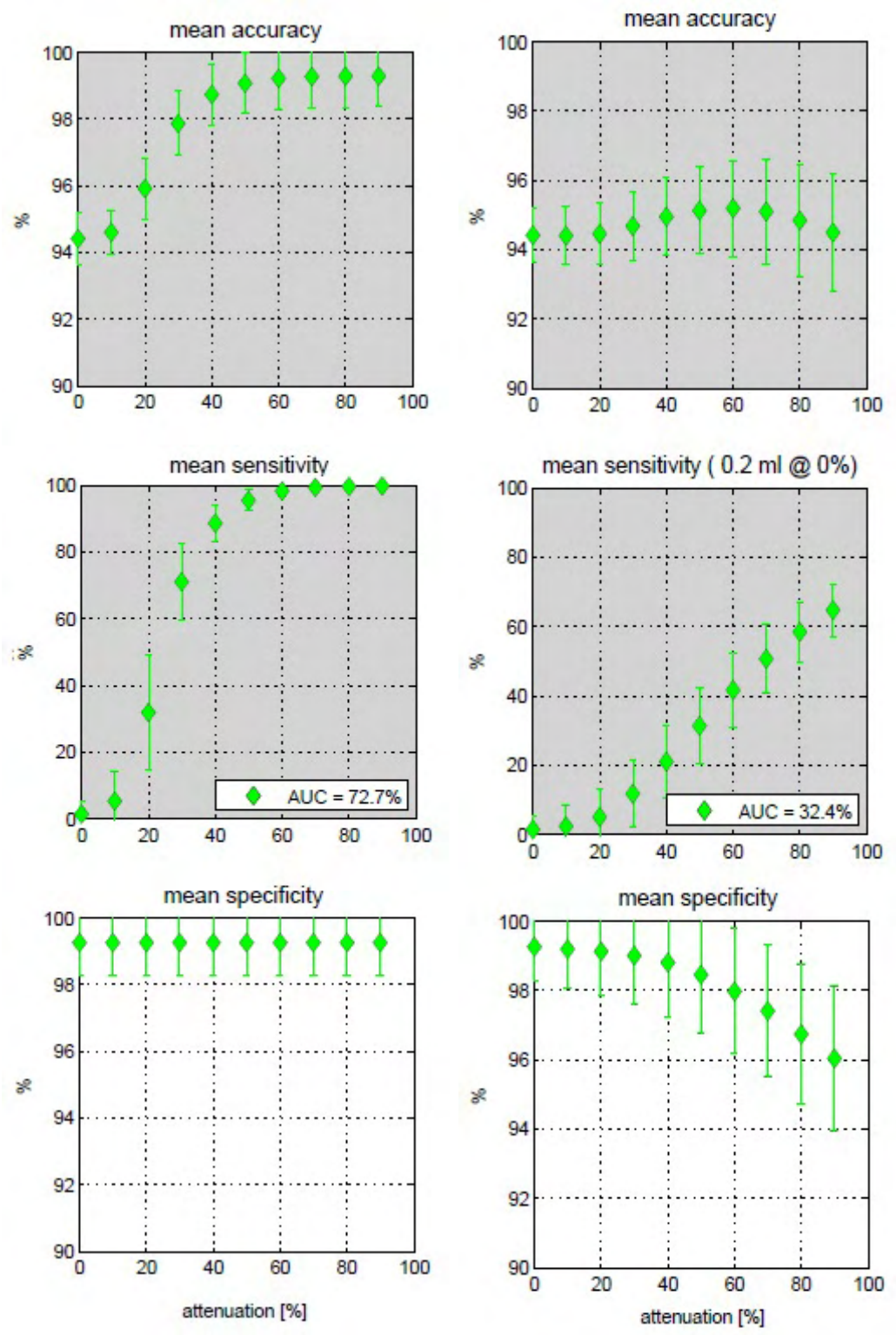


Fig. 4.18 Graphs shows mean sensitivity and specificity and accuracy for both the models at 8mm filter size. Left graph represents results of model 1 and right of model 2

Due to a large filter size the smearing effect in the voxels is also higher. This effect can be controlled by reducing the Gaussian kernel to smaller sizes. *Fig. 4.19* and *Fig. 4.20* shows results obtained by performing morphometry with second model on images smoothed at 2mm, 4mm, 8mm and 12mm, p-value 0.005 and lesion type severe,

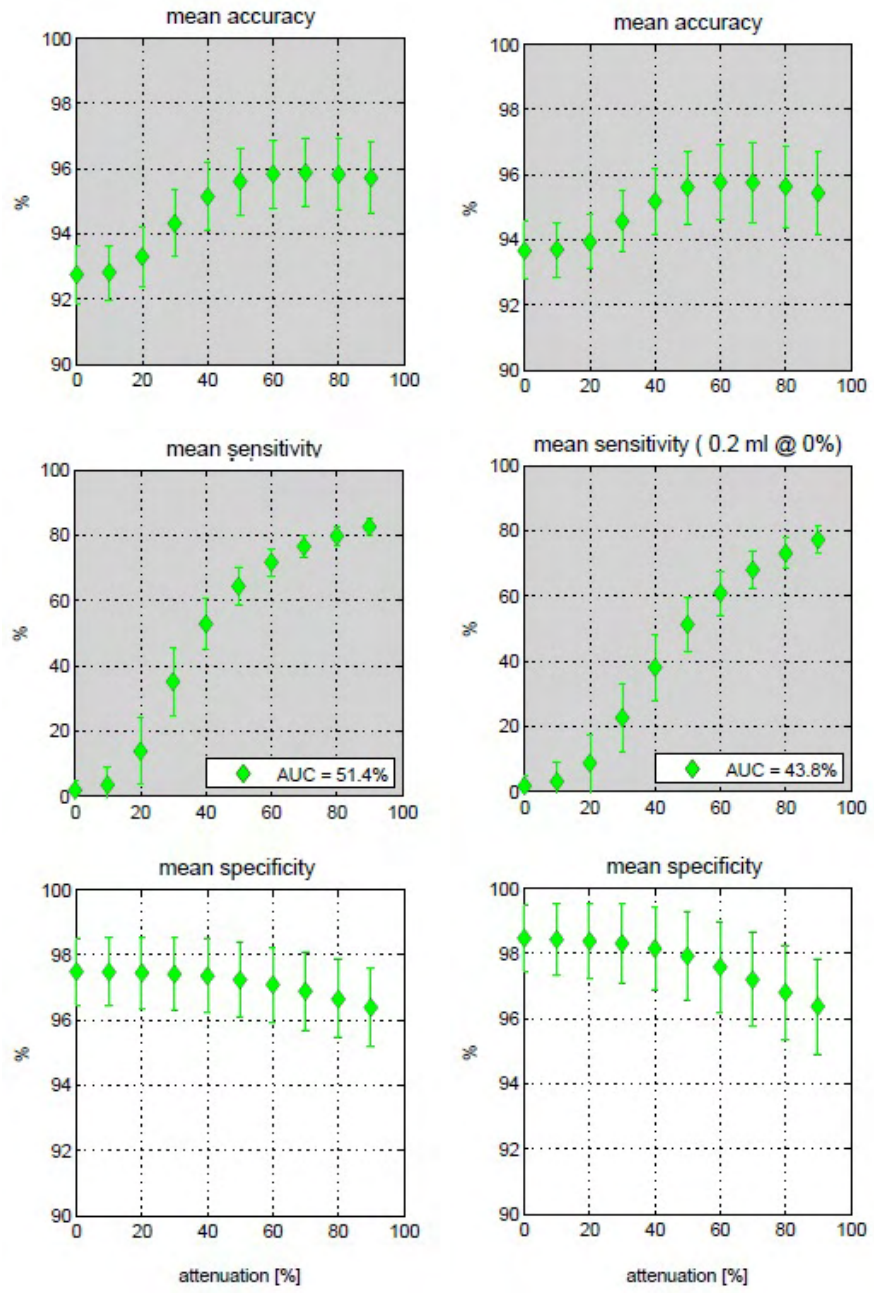


Fig. 4.19 Above graphs shows mean sensitivity and specificity and accuracy. Graphs from left to right represent results obtained with 2mm and 4mm respectively

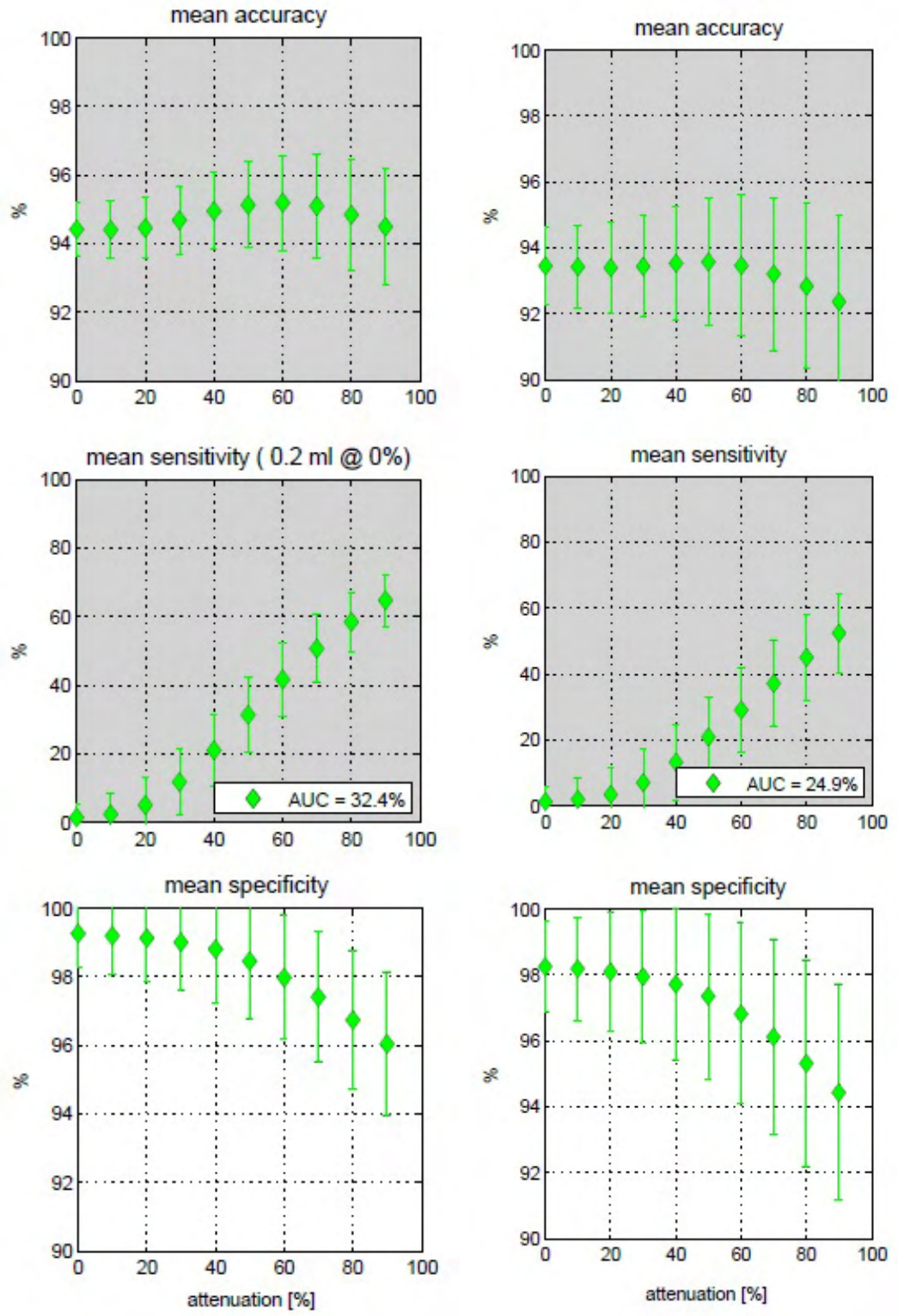


Fig. 4.20 Above graphs shows mean sensitivity and specificity and accuracy. Graphs from left to right represent results obtained with 8mm and 12mm respectively

It is apparent that model – 1 is more efficient in detecting the simulated lesion due to clear delineation of the boundaries of lesion. The measures of accuracy, sensitivity and specificity are higher for first approach in comparison to the second. However, second approach is more realistic for lesion detection. It is also evident from the later part of this experiment that smaller the Gaussian kernel more efficient the second approach of VBM becomes. A smaller Gaussian kernel causes less smearing effect, thus there are more true positives and less false positives, which in consequence improves the performance of the model.

5. Implementation of VBM on Simulated Data for Optimization of Framework

5.1. Introduction

From the results of the previous section *4.5* (simulation of artificial lesion), it is evident that when a lesion is applied before pre-processing the image, better results are obtained for statistical tests using images generated with smaller smoothing filter size. Same inference can be made about smoothing kernel from the results of section *4.3.2* (segmentation error) also. Using previous results and observations, VBM is further optimized varying significance level in the statistical test and smoothing kernel size for joint DARTEL implementation to make it more robust for detection of MS lesions. For the framework built during this thesis work, a benchmark of 90% accuracy and specificity and 60% sensitivity was set. Threshold for sensitivity was kept low taking into account the fact that misclassification of lesion voxels as grey matter leads to poor detection of the lesion clusters.

5.2. Materials and Methods

As the volume and region distribution of MS lesion available from BrainWeb is known, it can be used as a reference to optimize the performance of VBM model. Brain image with severe MS lesion was used as the patient image and severe MS lesion mask as a reference for optimization of VBM model. Normative database – 1, 2 and 3 were used as controls for the statistical test.

VBM was performed on brain image with severe MS lesion at three different significance levels, 0.05, 0.01 and 0.1 for hypo intense maps and two significance levels, 0.005 and 0.05 for hyper intense maps. The significance levels selected for optimization were based on the results of previous experiments which show different sensitivities in detecting lesions when investigating reduction in white matter (hyper intense maps) and damage in white matter (hypo intense maps). Normalized and smoothed images of grey matter and white matter generated by joint DARTEL implementation were used in the statistical test. The smoothing kernel for DARTEL was varied from 0 to 12mm (0, 2, 3, 4, 5, 6, 8 and 12mm).

The model set-up to detect MS lesions in patients affected by MS should detect the lesions efficiently but at the same time it should be efficient enough not to detect lesions in normal (with no MS lesions) healthy subjects. Therefore, the optimized model is tested for normal subject also. Three normal subject data were used to verify the efficiency of the framework for normal brains. These subjects were clinically diagnosed with no disease. The age of these subjects is from 25 to 64 years. The details of these normal healthy subjects are shown in *Table 9.*

Table 9 Normal Subjects

Subject	Gender	Age
1	Not available	25
2	F	44
3	F	64

Images of VBM results for subject – 1 are shown in the results sub-section.

An overlay of detected lesion and contour of the ground truth image of severe MS lesion was used for visual interpretation. A graphical representation of accuracy, sensitivity and specificity was used for numeric interpretation of VBM results.

5.3. Results

Fig. 5.1 and *Fig. 5.2* show graphs for accuracy, sensitivity and specificity at different significance levels and smoothing kernels for VBM performed on simulated brain image (severe MS lesion brain image) using database – 1 as the control database for statistical comparison.

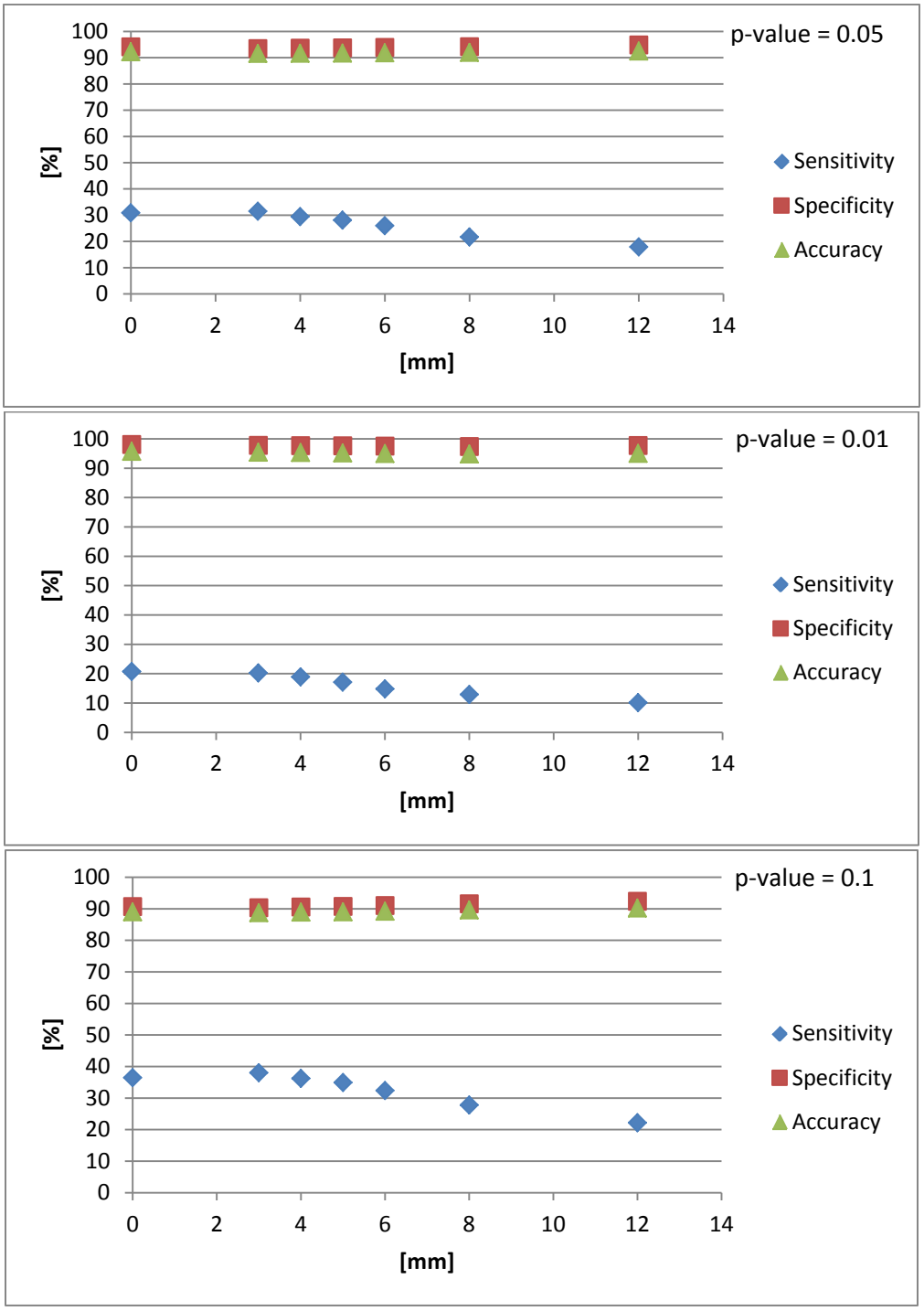


Fig. 5.1 Accuracy, sensitivity and specificity for hypo intense maps at 0.05, 0.01 and 0.1 significance levels

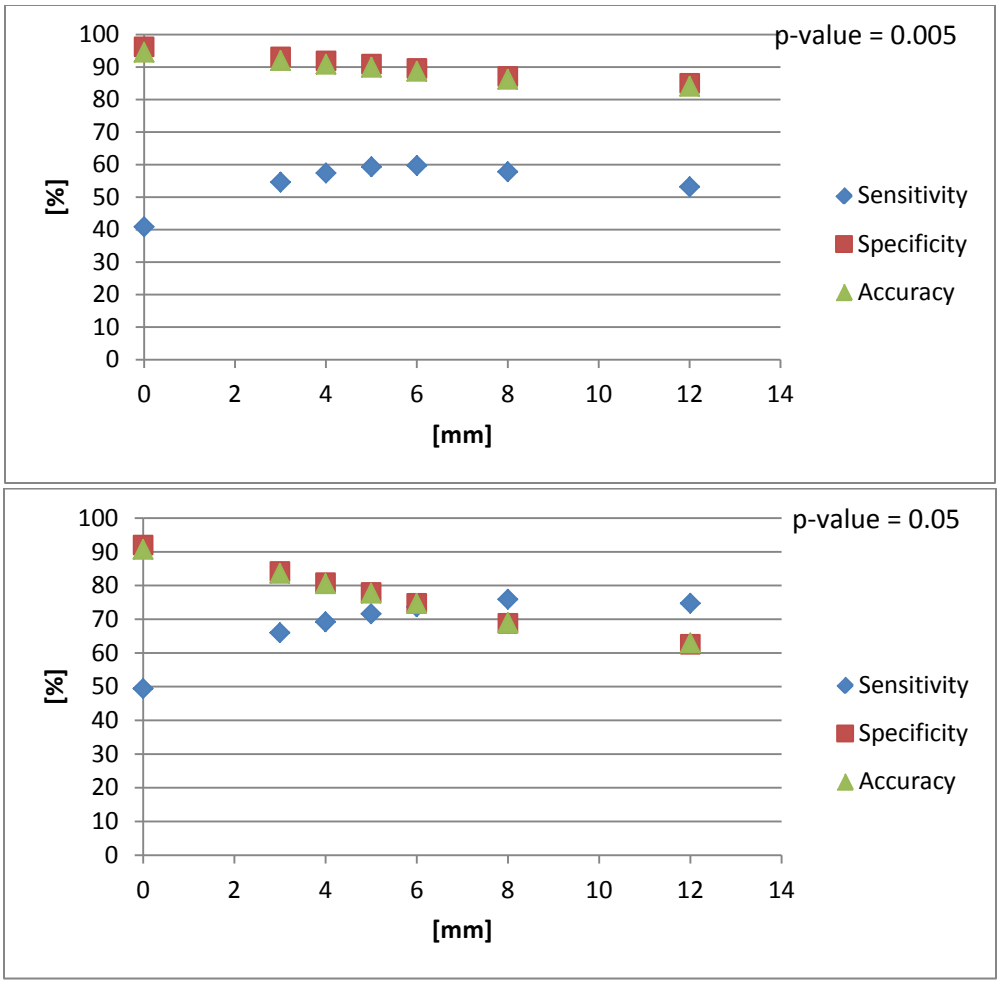


Fig. 5.2 Accuracy, sensitivity and specificity for hyper intense maps at 0.005 and 0.05 significance levels

High accuracy and specificity are obtained for both hypo and hyper intense maps for VBM done using images smoothed at lower smoothing kernel sizes. Very low sensitivity is obtained for hypo intense maps. In performance of VBM model for hyper intense maps there is a trade-off between accuracy and specificity, and sensitivity with increasing smoothing kernel size. As observed in section *4.5*, it is evident in this experiment also, that images pre-processed with smaller smoothing filter sizes leads to better performance of VBM.

High specificity is obtained for lesions detected as reduction in white matter (hypo intense maps) at all three significance levels but with low sensitivity. Significance level of 0.1 offers the best trade-off between sensitivity and specificity values at lower smoothing kernel sizes. A higher sensitivity is obtained for lesions detected as damage in white matter (hyper intense maps) at small smoothing kernel size, with the significance level of 0.05 in the statistical test but

simultaneously low specificity is achieved. A good trade-off at significance level of 0.005 is observed. Approximately 90% accuracy and specificity for both biomarkers and approximately 40% and 55% sensitivity for lesions detected as reduction and damage in white matter respectively was attained for database – 1. For database – 2, the values for accuracy and specificity for biomarker *a* (hypo intense map) and *b* (hyper intense map) are approximately 90 and 95%, while sensitivity is approximately 49 and 53% for biomarker *a* and *b* respectively. For database – 3, accuracy and specificity for biomarker *a* (hypo intense map) and *b* (hyper intense map) are approximately 93 and 95%, while sensitivity is approximately 45 and 58% for biomarker *a* and *b* respectively. The average of accuracy and specificity for all three databases is more than 90% and average of sensitivity for biomarker *a* is 45% and for biomarker *b* is 55% Graphs for database – 2 and 3 can be found in *appendix*.

All the figures with t-maps henceforth show axial slices of brain. Design of the t-maps used in the entire thesis work is a copyright of Jung Diagnostics (Copyright © 2012 Jung Diagnostics). Blue contour indicates actual (ground truth) severe MS lesion. The intensities indicate severity of lesion at that particular region. Color bar indicates severity of the lesion. A yellow/blue region demonstrates presence of more severe lesion and red/purple region demonstrates presence of less severe lesion. Numbers on the left of color bar corresponds to the t-values and on the right shows p-value.

Following two images indicate elevated grey matter intensities in white matter region of brain.

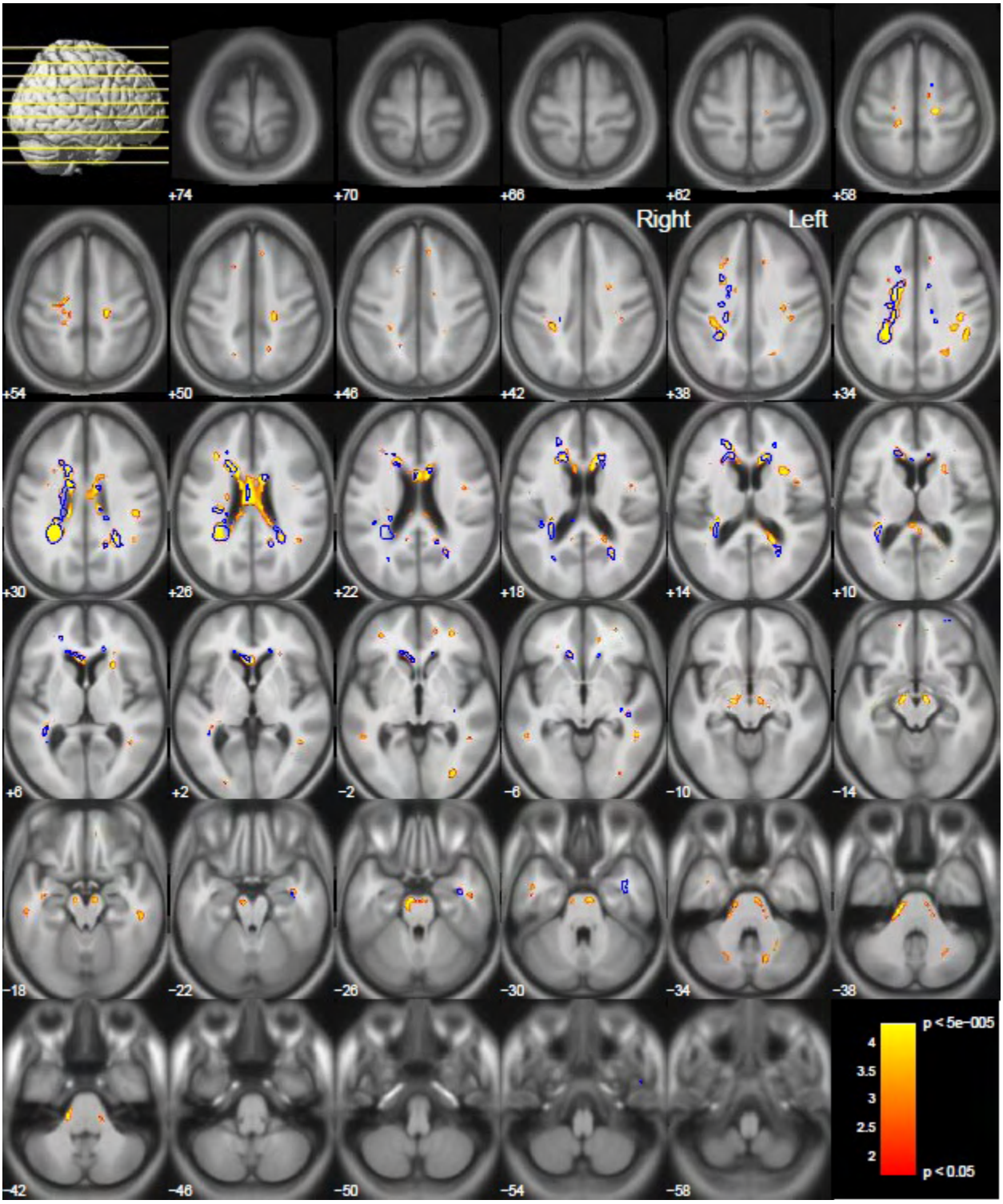


Fig. 5.3 t-map (hpyermap, significance level 0.005 and smoothing filter 4mm)

Pixels with color intensities (from the color bar) inside blue contour indicate true positives and outside the contour indicate false positives.

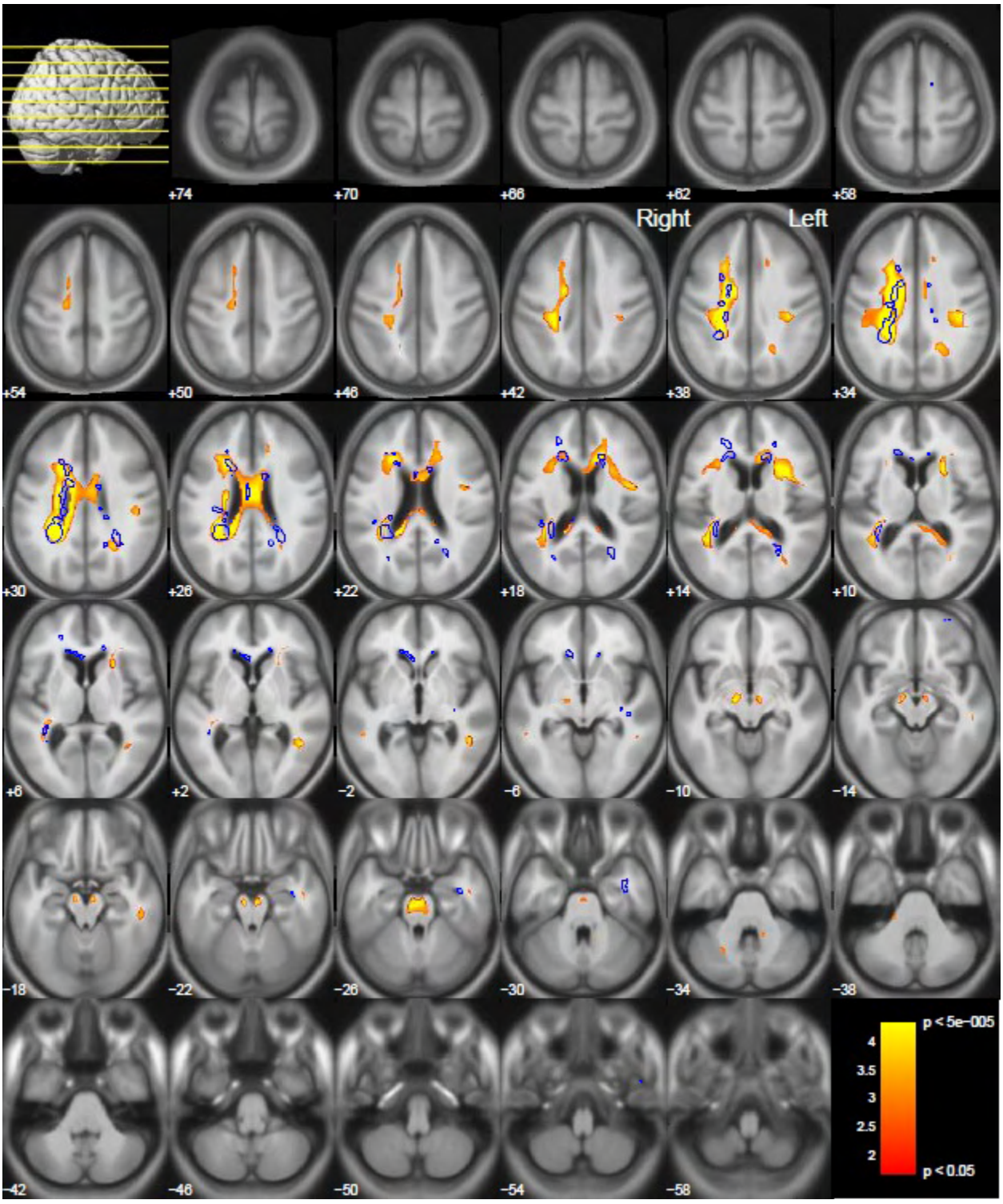


Fig. 5.4 t-map (hpyermap, significance level 0.005 and smoothing filter 12mm)

From *Fig. 5.3* and *Fig. 5.4* it is evident that less false positive lesion regions are detected using 4mm filter size for smoothing kernel compared to regions detected using 12mm filter size.

Following two images indicate reduced white matter in white matter region of the brain.

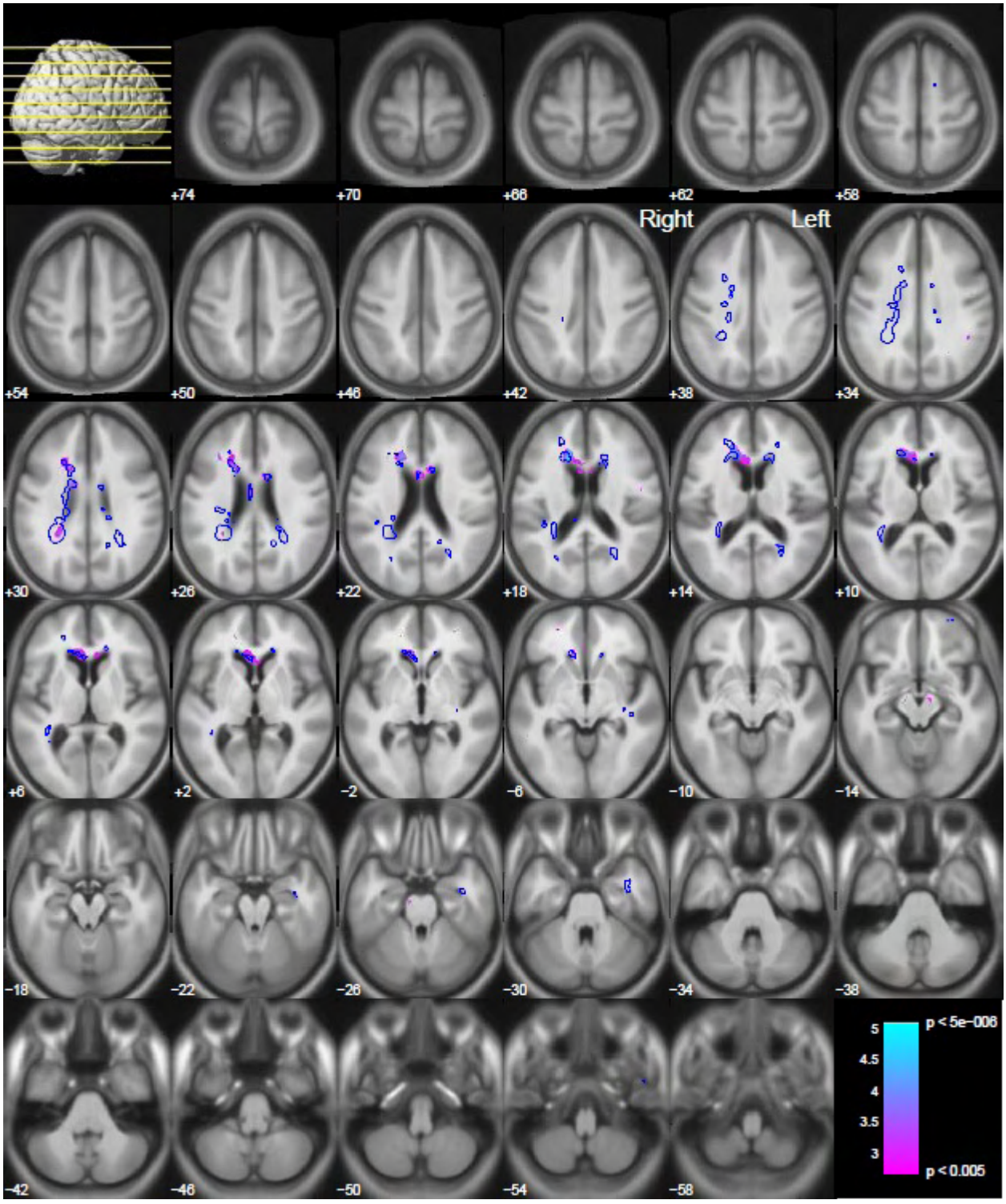


Fig. 5.5 t-map (hpyomap, significance level 0.1 and smoothing filter 3mm)

Pixels with color intensities (from the color bar) inside blue contour indicate true positives and outside the contour indicate false positives.

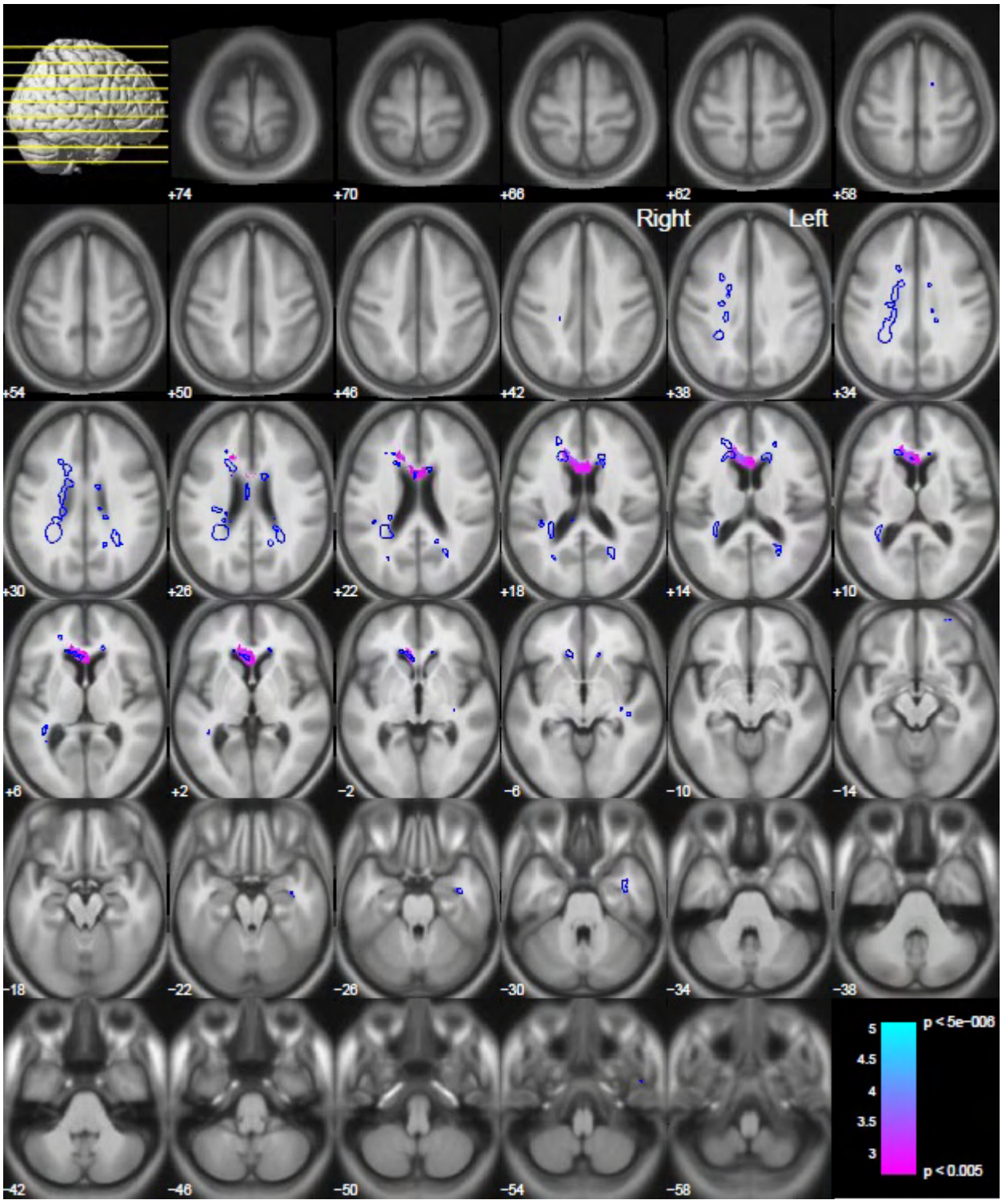


Fig. 5.6 t-map (hpyomap, significance level 0.1 and smoothing filter 12mm)

From *Fig. 5.5* and *Fig. 5.6* is evident that more true positive and less false positive lesion regions are detected using 4mm filter size for smoothing kernel compared to regions detected using 12mm filter size.

A good trade-off between accuracy and specificity, and sensitivity is obtained at significance level 0.1 and smoothing kernel size of 3mm for DARTEL for detecting reduction in white matter and at 0.005 significance level and 4mm smoothing kernel for detecting damage in white matter.

On performing VBM on normal healthy subjects, very small or no clusters of pixels were classified as MS lesions. The clusters classified as MS lesions are due to partial volume effects in most instances. Partial volume effect is a conflict of contrast between brain tissues at the borders. It is seen when more than one tissue type shares intensity at the same voxel [(48)].

Considering these subjects disease free false positives (FP), true negatives (TN) and specificity were calculated. *Table 10* shows FP, TN and specificity calculated with respect to white matter as region of interest (ROI) and *Table 11* shows FP, TN and specificity calculated with respect to severe MS lesion as ROI. It is evident that high specificity for both ROIs indicates the model to be robust for normal healthy subjects also.

Table 10 Specificity for normal subject data (ROI – white matter mask)

	False positives vol. [ml]		True negatives vol. [ml]		Specificity [%]	
	Hypomap	**Hypermap**	**Hypomap**	**Hypermap**	**Hypomap**	**Hypermap**
Simulated normal brain image	3.12	1.44	272.80	274.48	98.87	99.48
Normal subject 1	0.78	3.45	275.14	272.47	99.72	98.75
Normal subject 2	0.82	4.48	275.10	271.44	99.70	98.38
Normal subject 3	0.64	11.12	275.28	264.80	99.77	95.97

Table 11 Specificity for normal subject data (ROI – severe MS lesion mask)

	False positives vol. [ml]		True negatives vol. [ml]		Specificity [%]	
	Hypomap	**Hypermap**	**Hypomap**	**Hypermap**	**Hypomap**	**Hypermap**
Simulated normal brain image	0.38	0.05	11.98	12.31	96.93	99.60
Normal subject 1	0.02	0.04	12.34	12.32	99.84	99.68
Normal subject 2	0.05	0.35	12.31	12.01	99.60	97.17
Normal subject 3	0	1.28	12.36	11.08	100	89.64

6. Implementation of Optimized Framework on Clinical Data

6.1. Introduction

Previous chapters described the modules of used in SPM, different approaches of implementing VBM and their optimization. The technological framework built for older group of patients for detecting AD was adapted to younger group of patients for detecting multiple sclerosis lesions. Further, crucial parameters were optimized to increase the detectability of lesions in white matter region. Performance of the customized framework was then tested using simulated data available from BrainWeb. This chapter describes the implementation of customized and optimized framework to detect lesions in real MS patients.

6.2. Materials and Methods

Voxel-based morphometry was used to detect MS lesions in real patients. The framework described in previous section, constructed adapting special modules for pre-processing was implemented on the clinical data. Optimized parameters of different modules were implemented in the framework. Joint DARTEL was used for registration of images after segmentation. Grey and white matter images were smoothed with 4mm and 3mm smoothing kernel size respectively. Significance level of 0.1 was used for detecting reduction in white matter (for generating hypo intense maps) and a significance level of 0.005 was used to detect damage in white matter (hyper intense maps) for the statistical test.

Images of six patients were tested for detecting multiple sclerosis lesions using the current framework. All the patients show a relapsing remitting pattern of multiple sclerosis. Age of patients is from 30 years to 68 years and two patients are males and four are females. *Table 10* shows the details of patients,

Table 12 Clinical Data

Sr. No.	Gender	Age
1	M	30
2	F	41
3	F	55
4	M	69
5	F	32
6	F	35

6.3. Results

In the figures below, red contour indicates boundaries of white matter mask. White matter contour helps in concluding that whether a detected lesion is really in white matter region or not. The white matter mask was generated thresholding the Cyceron template for white matter at 85% (refer to section *4.4* for thresholding of image). The intensities indicate severity of lesion at that particular region. Color bar indicates severity of the lesion. A yellow/blue region demonstrates more severe lesion and red/purple region demonstrates a less severity of lesion. Numbers on the left of color bar corresponds to the t-values and on the right shows p-value.

Following image indicates elevated grey matter intensities in white matter region in the brain of healthy subject – 1.

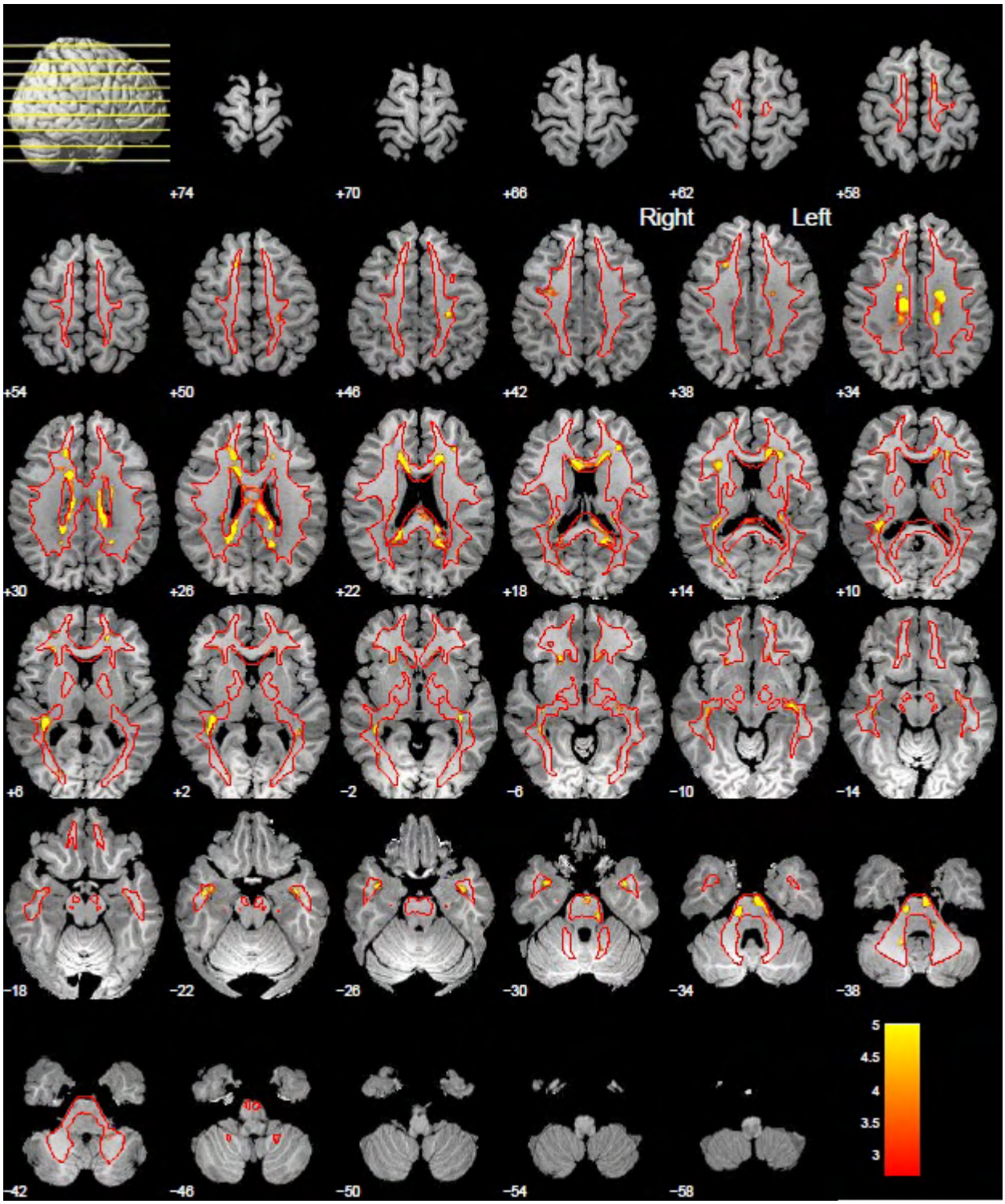


Fig. 6.1 t-map – subject 1 (hpyermap, significance level 0.005 and smoothing filter 4mm)

Following image indicates reduced white matter in white matter region in the brain of healthy subject – 1.

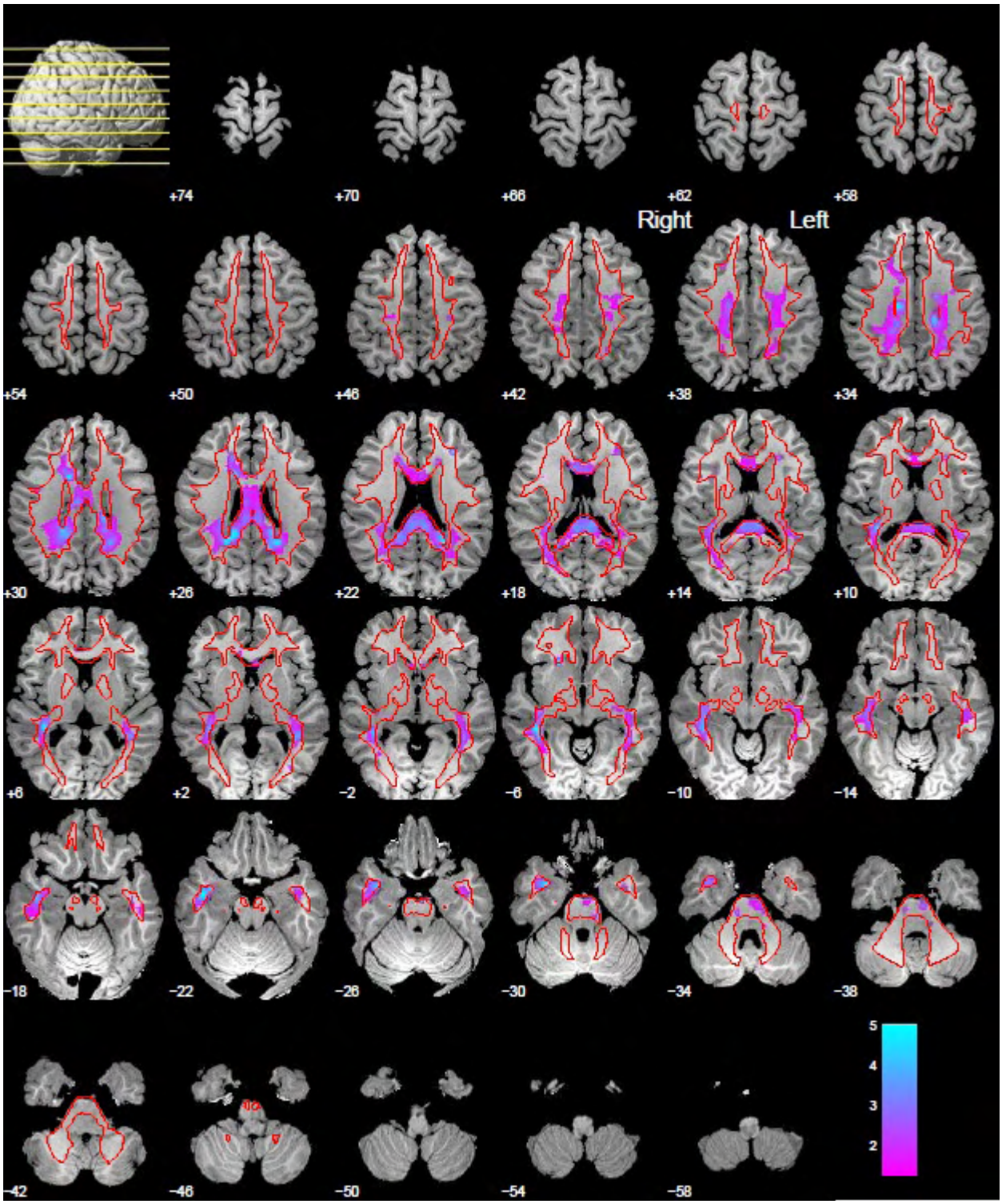


Fig. 6.2 t-map – subject 1 (hpyomap, significance level 0.1 and smoothing filter 3mm)

Following image indicates elevated grey matter intensities in white matter region in the brain of healthy subject – 4.

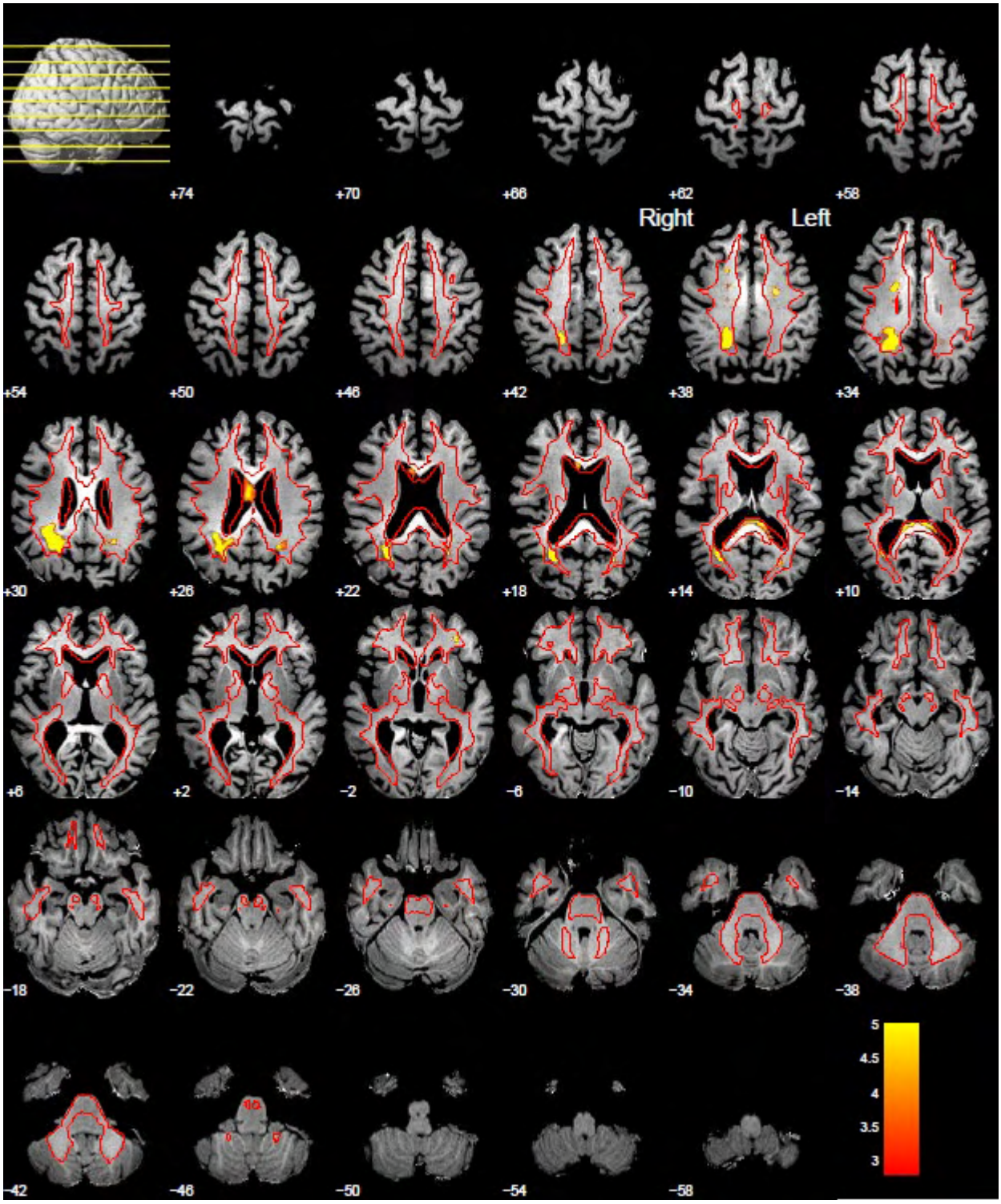


Fig. 6.3 t-map – subject 4 (hpyermap, significance level 0.005 and smoothing filter 4mm)

Following image indicates reduced white matter in white matter region in the brain of healthy subject – 4.

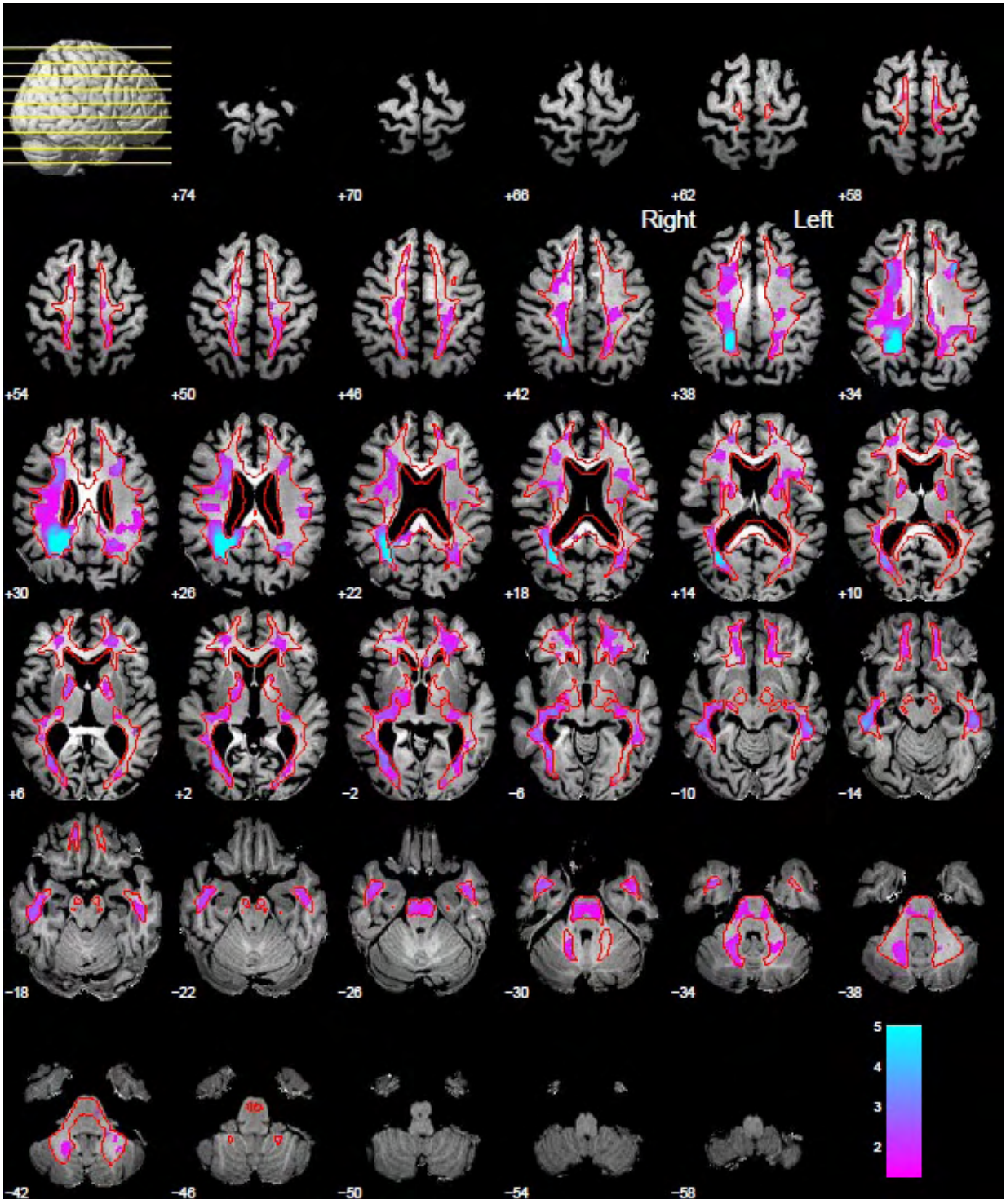


Fig. 6.4 t-map – subject 4 (hpyomap, significance level 0.1 and smoothing filter 3mm)

Similar results for other subjects (subject – 2, 3, 5 and 6) can be found in *appendix*.

7. Discussion, Conclusion and Future Aspects

Discussion

Segmentation is one of the most important steps of image pre-processing. Performance of segmentation routine has an influence on performance of normalization and statistical comparison also. A comment about optimal segmentation can be made by visual interpretation of individual patient data using anatomical landmarks. In one study done by a group, four popular segmentation algorithms from SPM99, SPM2, EMS (Expectation Maximization Segmentation) and FSL (FMRIB Software Library) were tested [46]. According to the findings of this study, all the algorithms failed to segment T1-weighted images with lesions correctly so they used skull stripped unsegmented images in their statistical comparisons. In this thesis work instead of using skull stripped unsegmented images; an experiment to optimize the segmentation routine was done as described in section *4.3.1*, segmentation optimization. A nice segmentation of tissues was attained for individual subjects at specific set of parameter settings but no optimal set of parameters leading to a perfect segmentation which suits all the images was found in this experiment. This is due to the fact that the presence of lesions causes low contrast between grey matter and white matter in T1-weighted images [46].

In the next section *4.3.2* an average of misclassification due to segmentation is found out using a set of simulated brain images. The mean of these misclassification errors in grey matter, white matter and CSF were approximately 4mm, 3mm and 7mm respectively. An assumption was made that these misclassifications can be compensated by using equivalent smoothing kernel sizes for grey matter, white matter and CSF respectively. Further, to support this assumption an experiment was done to check the best suitable Gaussian smoothing kernel size which can compensate segmentation errors as described in section *4.5*. Better sensitivity and specificity were obtained in detecting lesions at smaller smoothing kernel sizes. Research done by S.M.D. Henley et al. also supports these results. In this research, they used three different smoothing kernel sizes (4, 6 and 8mm) and found out that smaller size (4 or 6mm) is sufficient [49].

Section *4.4* describes the effect of misclassification of MS lesions. As lesion and grey matter intensities are matching there is a misclassification of approximately same volume of grey matter. This is evident for all three MS lesions that the misclassification scales according to the volumes of lesions and is also described in the studies made by Renske de Boer et al. [(45)] and Emmanuel Stamatakis et al. [(46)].

A model to simulate lesions replicating the behavior of misclassification was built to optimally detect lesions in the statistical comparisons implementing two different approaches of VBM as described in section *4.5*. Optimization of the framework was done by performing VBM at different significance levels and smoothing kernel sizes. From the experiments described in section *5* it was evident that lesions for reduction in white matter are detected more efficiently at higher significance levels, whereas damage in white matter can be detected at lower significance levels also, both at lower smoothing kernel size. A good significance level for the statistical test for detecting reduction and damage in white matter is 0.1 and 0.005 respectively, with the smoothing filter size of 4mm for grey matter and 3mm for white matter images. After segmentation, the white matter image consists of voxels in the lesion region. These voxels are actually lesion voxel but due to inaccurate segmentation are classified as white matter. Thus, the statistical test detect fewer voxels as lesions in the voxel-wise comparison. This could be a possible reason for lower sensitivity when investigating white matter for lesions and thus a higher significance level is required in the statistical test to detect white matter reduction. The benchmark of 90% accuracy and specificity and 60% sensitivity as discussed in section *5* was almost achieved, with more than 90% accuracy and specificity for both biomarkers and approximately 45% and 55% sensitivity for lesions detected as reduction and damage in white matter respectively. In the studies done by Sonya Mehta et al., voxel-based morphometry was evaluated for focal lesion detection using T1-weighted images. In their findings also it is evident that the lesions are incorrectly classified as grey matter and thus, VBM fails to detect all the lesions [(50)]. In contrast to this thesis work, in the studies done by Emmanuel Stamatakis et al. quantification measures for true positives and true negatives as high as 99 and 95% respectively were obtained, with false positives and false negatives as low as 1 and 6% respectively [(46)]. In this work T1-weighted images were used to detect artificially simulated lesions and this technique is based on statistical comparison of skull stripped and unsegmented images. The use

of unsegmented images was adapted in their technique as the segmentation engine failed to segment the brain tissues correctly due to the presence of lesions.

Conclusion

In conclusion, this thesis work describes the implementation and optimization of an image processing framework for detection of multiple sclerosis lesions. The framework constructed to detect lesions in AD patients was successfully adapted to detect MS lesions optimizing SPM modules and VBM methodology. The segmentation module of SPM is quite good in classifying grey matter and white matter tissues but at the same time misclassifies MS lesions as grey matter. No optimal set of parameters leading to a perfect segmentation which can be adapted in general was found in the experiment done for optimizing segmentation. Smoothing kernel sizes equivalent to the segmentation errors (*4.3.2*), when used were found to be effective in compensating the errors. Two biomarkers defined to simulate artificial lesions proved to be effective providing a foundation for implementing the image processing framework. Using these biomarkers the framework was successfully optimized to detect MS lesions. The framework when tested with simulated data, attained the benchmark of 90% set for accuracy and specificity but did not attain the benchmark of 60% set for sensitivity for reduction and damage in white matter. Thus, more work is needed to be done to make the current image processing framework more efficient in detecting MS lesions. The framework detected MS lesions fairly in real clinical data also. Though, a comment regarding false positives cannot be made as there is no reference ground truth image (for MS lesions) for either of the real clinical subjects. But a confirm evaluation about the lesions detected can be made by a medical specialist. Voxel-based morphometry can be further developed to detect MS lesions more efficiently by optimizing currently used techniques and/or implementing new techniques for image preprocessing and statistical comparison.

Future Aspects

The current version of framework works efficiently for detecting multiple sclerosis lesions. This framework can be further optimized for a better detectability of lesions customizing certain modules of SPM. Segmentation with a set of optimized parameters leading to minimum misclassification errors in tissue segmentation can be implemented for every subject to be examined. Templates using MRI scans of younger group of subjects, generated specifically for multiple sclerosis disease can be used for registration and normalization for a better match of brains for statistical tests. Further optimization of segmentation routine is possible if a large dataset of simulated brains is available with their ground truth images so that they can be compared individually for volume deviations and visual interpretation. VBM technology can be further optimized implementing other statistical tests; for e.g. ANOVA. The current image processing framework needs to be clinically evaluated by neurologists.

Bibliography

1. Multiple Sclerosis International Federation. [Online] [Cited: June 11, 2012.] http://www.msif.org/en/about_ms/index.html.

2. **Charles Smith.** Multiple Sclerosis An Introduction to the Disease. [book auth.] Rosalind Kalb. *Multiple Sclerosis.*

3. National Multiple Sclerosis Society. [Online] [Cited: June 7, 2012.] http://www.nationalmssociety.org/about-multiple-sclerosis/what-we-know-about-ms/faqs-about-ms/index.aspx.

4. **Bolaños Jiménez R et al.** *Multiple sclerosis: An overview of the disease and current concepts of its pathophysiology.* 4, s.l. : Journal of Neuroscience and Behavioural Health, April 2011, Vol. 3.

5. *Atlas of Multiple Sclerosis.* s.l. : World Health Organization, 2008.

6. **Ian Robinson David Rog.** Principles of Management in MS. [book auth.] *Multiple Sclerosis.* 2010.

7. **Schapiro, Randall T.** What is Multiple Sclerosis. [book] *Managing the Symptoms of Multiple Sclerosis.*

8. **Marci Catanzaro George Kraft.** Disease Distribution. [book] *Living with Multiple Sclerosis.* 2000.

9. **Marci Catanzaro George Kraft.**Causes of Multiple Sclerosis. [book] *Living with Multiple Sclerosis.* 2000.

10. **Randall Schapiro.** Part II: Managing MS Symptoms. [book] *Managing the Symptoms of Multiple Sclerosis.* 2007.

11. **Jock Murray, Stephen Reingold Nancy Holland.** What is Multiple Scelrosis and How is it Diagnosed. [book] *Multiple Sclerosis: A Guide for the Newly Diagnosed.* 2007.

12. **Marci Catanzaro George Kraft.** Diagnosis. [book] *Living with Multiple Sclerosis.* 2000.

13. Evoked Potentials. *National Multiple Sclerosis Society.* [Online] [Cited: July 22, 2012.] http://www.nationalmssociety.org/about-multiple-sclerosis/what-we-know-about-ms/diagnosing-ms/evoked-potentials/index.aspx.

14. **Jock Murray, Stephen Reingold Nancy Holland.** The Diagnosis of Multiple Sclerosis. [book] *Multiple Sclerosis: A Guide for the Newly Diagnosed.* 2007.

15. **Charles Smith.** The Diagnostic Process. [book] Rosalind Kalb. *Multiple Sclerosis.* 2012.

16. **Ian McDonald et al.** *Recommended Diagnostic Criteria for Multiple Sclerosis: Guidelines from the International Panel on the Diagnosis of Multiple Sclerosis.* s.l. : Wiley Liss Inc., 2001.

17. **Chris Polman et al.** *Diagnostic Criteria for Multiple Sclerosis: 2010 Revisions to the McDonald Criteria.* s.l. : American Neurological Association, 2011.

18. **Jock Murray, Stephen Reingold Nancy Holland.** What treatment is available? [book] *Multiple Sclerosis: A Guide for the Newly Diagnosed.* 2007.

19. Treatments. *National Multiple Sclerosis Society.* [Online] [Cited: July 22, 2012.] http://www.nationalmssociety.org/about-multiple-sclerosis/what-we-know-about-ms/treatments/index.aspx.

20. Avonex (interferon beta 1a). *National Multiple Sclerosis Society.* [Online] [Cited: July 22, 2012.] http://www.nationalmssociety.org/about-multiple-sclerosis/what-we-know-about-ms/treatments/medications/interferon-beta-1a-avonex/index.aspx.

21. Betaseron; Extavia (interferon beta 1b). *National Multiple Sclerosis Society.* [Online] [Cited: July 22, 2012.] http://www.nationalmssociety.org/about-multiple-sclerosis/what-we-know-about-ms/treatments/medications/interferon-beta-1b/index.aspx.

22. Copaxone (Glatiramer Acetate). *National Multiple Sclerosis Society.* [Online] [Cited: July 22, 2012.] http://www.nationalmssociety.org/about-multiple-sclerosis/what-we-know-about-ms/treatments/medications/glatiramer-acetate/index.aspx.

23. **John Smith, Gregory Sorensen, James Thrall.** *Biomarkers in Imaging: Realizing Radiology's Future.* s.l. : RSNA, 2003.

24. **Terry McCormick, Kathleen Martin and Michael Hehenberger**. Advancing the Utility of Imaging Biomarkers. s.l. : IBM, 2009.

25. **Iain Wilson, Christopher Rofe, Graham Lloyd Rajat Chowdhury.** [book] *Radiology at a Glance.* 2010.

26. **Per Suppa.** *Implementation and Optimization of an Image Processing Framework to Evaluate PET/MRI Brain Data for Early Diagnosis of Alzheimer.* Hamburg : s.n., 2010.

27. Statistical Parametric Mapping. *Wellcome Trust Centre for Neuroimaging.* [Online] [Cited: May 12, 2012.] http://www.fil.ion.ucl.ac.uk/spm/.

28. The SPM approach in brief. [Online] [Cited: May 28, 2012.] http://www.fil.ion.ucl.ac.uk/spm/.

29. **Karl Friston, John Ashburner.** Voxel-based morphometry. [book auth.] Karl J. Friston et al. *Statistical Parametric Mapping - The Analysis of Functional Brain Images.* 2007.

30. **John Ashburner, Karl Friston.** Image Segmentation. [book auth.] Karl Friston, William Penny John Ashburner. *Human Brain Function.*

31. The MNI brain and the Talairach atlas. *http://imaging.mrc-cbu.cam.ac.uk/.* [Online] [Cited: May 20, 2012.] http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach.

32. **John Ashburner et al.** Segmentation. [book] *SPM8 Manual. 2012.*

33. **Hervé Lemaître et al.** *Age- and sex-related effects on the neuroanatomy of healthy elderly.* s.l. : Elsevier, 2005, Vol. 26.

34. **John Ashburner, Karl Friston.** Spatial Normalization using Basis Functions. [book auth.] Karl Friston, William Penny John Ashburner. *Human Brain Function.*

35. **John Ashburner, Karl Friston.** *Unified Segmentation.* 2005.

36. **John Ashburner et al.** DARTEL Tools. [book] *SPM8 Manual.* 2011.

37. **John Ashburner, Karl Friston.** Statistical Modelling and Inference. [book auth.] Karl Friston et al. *Statistical Parametric Mapping.* 2007.

38. **M. Mühlau et al.** *Voxel-Based Morphometry in Individual Patients: A Pilot Study in Early Huntington Disease.* 2009.

39. **Saul Teukolsky, William Vetterling, Brian Flannery, William Press.** Statistical Description of Data. [book] *Numerical Recipes in C.* 1992.

40. **Rif. Henson.** *Comparing a single patient versus a group of controls (and SPM).* 2006.

41. *BrainWeb.* [Online] [Cited: July 27, 2012.] http://brainweb.bic.mni.mcgill.ca/brainweb/.

42. **de Lange EE et al.** *Magnetization prepared rapid gradient-echo (MP-RAGE) MR imaging of the liver: comparison with spin-echo imaging.* 1991.

43. **Richard Semelka, Mark Brown.**Pulse Sequences. [book] *MRI Basic Princilples and Applications.* 2010.

44. **Kevin King, Xiaohong Zhou Matt Bernstein.** Radiofrequency Pulses. [book] *Handbook of MRI Pulse Sequences.*

45. **Renske de Boer et al.** W*hite matter lesion extension to automatic brain tissue segmentation on MRI.* s.l. : Neuroimage, 2009, Vol. 45.

46. **Emmanuel Stamatakis et al.** *Identifying lesions on structural brain images—Validation of the method and application to neuropsychological patients.* 2005.

47. **Wilkin Chau, Anthony McIntosh.** *The Talairach coordinate of a point in the MNI space.* s.l. : Elsevier, 2005, Vol. 25.

48. **John Ashburner, Karl Friston.** Segmentation. [book auth.] Karl Friston et al. *Statistical Parametric Mapping - The Analysis of Functional Brain Images.* 2007.

49. **S.M.D. Henley et al.** *Pitfalls in the Use of Voxel-Based Morphometry as a Biomarker: Examples from Huntington Disease.* s.l. : AJNR, 2010, Vol. 31.

50. **Sonya Mehta et al.** *Evaluation of Voxel-based Morphometry in Focal Lesion Detection in Individuals.* s.l. : Elsevier, 2003.

Appendices

McDonald's Criteria for Multiple Sclerosis [17]

Clinical Presentation	Additional Data Needed for MS Diagnosis
≥2 attacks[a]; objective clinical evidence of ≥2 lesions or objective clinical evidence of 1 lesion with reasonable historical evidence of a prior attack	None
≥2 attacks[a]; objective clinical evidence of 1 lesion	Dissemination in space, demonstrated by: ≥1 T2 lesion in at least 2 of 4 MS-typical regions of the CNS (periventricular, juxtacortical, infratentorial, or spinal cord)[d]; or Await a further clinical attack[a] implicating a different CNS site
1 attack[a]; objective clinical evidence of ≥2 lesions	Dissemination in time, demonstrated by: Simultaneous presence of asymptomatic gadolinium-enhancing and non-enhancing lesions at any time; or A new T2 and/or gadolinium-enhancing lesion(s) on follow-up MRI, irrespective of its timing with reference to a baseline scan; or Await a second clinical attack[a]
1 attack[a]; objective clinical evidence of 1 lesion (clinically isolated syndrome)	Dissemination in space and time, demonstrated by: For DIS: ≥1 T2 lesion in at least 2 of 4 MS-typical regions of the CNS (periventricular, juxtacortical, infratentorial, or spinal cord)[d]; or Await a second clinical attack[a] implicating a different CNS site; and For DIT: Simultaneous presence of asymptomatic gadolinium-enhancing and non-enhancing lesions at any time; or A new T2 and/or gadolinium-enhancing lesion(s) on follow-up MRI, irrespective of its timing with reference to a baseline scan; or await a second clinical attack[a]
Insidious neurological progression suggestive of MS (PPMS)	1 year of disease progression (retrospectively or prospectively determined) plus 2 of 3 of the following criteriad: 1. Evidence for DIS in the brain based on ≥1 T2 lesions in the MS-characteristic (periventricular, juxtacortical, or infratentorial) regions 2. Evidence for DIS in the spinal cord based on ≥2 T2 lesions in the cord 3. Positive CSF (isoelectric focusing evidence of oligoclonal bands and/or elevated IgG index)

If the Criteria are fulfilled and there is no better explanation for the clinical presentation, the diagnosis is ‘‘MS’’; if suspicious, but the Criteria are not completely met, the diagnosis is ‘‘possible MS’’; if another diagnosis arises during the evaluation that better explains the clinical presentation, then the diagnosis is ‘‘not MS.’’

[a]An attack (relapse; exacerbation) is defined as patient-reported or objectively observed events typical of an acute inflammatory demyelinating event in the CNS, current or historical, with duration of at least 24 hours, in the absence of fever or infection. It should be documented by contemporaneous neurological examination, but some historical events with symptoms and evolution characteristic for MS, but for which no objective neurological findings are documented, can provide reasonable evidence of a prior demyelinating event. Reports of paroxysmal symptoms (historical or current) should, however, consist of multiple episodes occurring over not less than 24 hours. Before a definite diagnosis of MS can be made, at least 1 attack must be corroborated by findings on neurological examination, visual evoked potential response in patients reporting prior visual disturbance, or MRI consistent with demyelination in the area of the CNS implicated in the historical report of neurological symptoms.

[b]Clinical diagnosis based on objective clinical findings for 2 attacks is most secure. Reasonable historical evidence for 1 past attack, in the absence of documented objective neurological findings, can include historical events with symptoms and evolution characteristics for a prior inflammatory demyelinating event; at least 1 attack, however, must be supported by objective findings.

[c]No additional tests are required. However, it is desirable that any diagnosis of MS be made with access to imaging based on these Criteria. If imaging or other tests (for instance, CSF) are undertaken and are negative, extreme caution needs to be taken before making a diagnosis of MS, and alternative diagnoses must be considered. There must be no better explanation for the clinical presentation, and objective evidence must be present to support a diagnosis of MS.

[d]Gadolinium-enhancing lesions are not required; symptomatic lesions are excluded from consideration in subjects with brainstem or spinal cord syndromes.

MS = multiple sclerosis; CNS = central nervous system; MRI = magnetic resonance imaging;
DIS = dissemination in space;DIT = dissemination in time; PPMS = primary progressive multiple sclerosis; CSF = cerebrospinal fluid; IgG = immunoglobulin G.

Evaluation of Independent and Joint DARTEL Performance

DARTEL is a high dimensional, non-linear image registration process. DARTEL process can be done by two different setups, independently and jointly as discussed in section *3.6*. An investigation was done to evaluate the performance of independent and joint DARTEL process.

Voxel-based morphometry using normative databases was done to check the performance of both DARTEL setups. VBM procedure was performed using the MATLAB program available at Jung Diagnostics. Working of this program is explained at the end of section *4.5.2*. Artificial lesions implementing both the biomarkers (biomarker *a* and *b*, as discussed in section *4.5*) were simulated on the normalized and smoothed grey matter and white matter images of normal brains using the severe MS lesion mask. Normalized and smoothed grey matter and white matter images were generated by both, independent and joint DARTEL.

Statistical test was done at significance level of 0.005 and both, grey matter and white matter images were smoothed at 8mm. Normative database – 1 and 2 were used for statistical comparison in the VBM process.

The graphs below show the mean of sensitivity obtained from the results of quantification after statistical test of each image in the database. Attenuation indicates the effect size. Legend in the graphs indicates the area under curve.

For database – 2 when investigating white matter to detect lesion simulated by biomarker *a* (attenuation), higher sensitivity is obtained by joint DARTEL process compared to independent process. Almost 100% lesion was detected at 40% effect size using images for statistical comparison generated by joint DARTEL setup; whereas approximately 85% of lesion is detected using images generated by independent DARTEL setup.

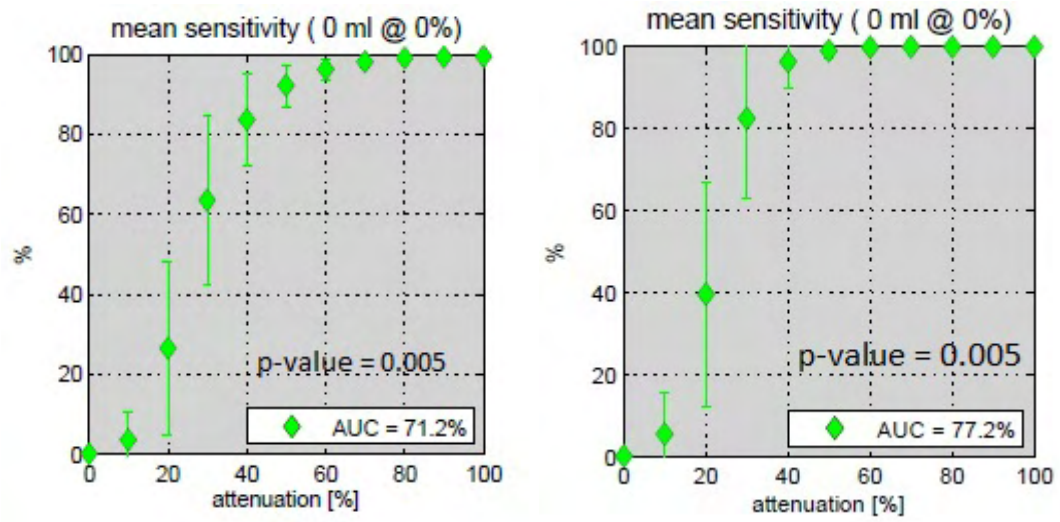


Fig. 0.1 Mean sensitivity curves of independent (left) and joint (right) DARTEL

When investigating grey matter to detect lesion simulated by biomarker *b* (amplification), almost 100% lesion was detected at 30% effect size using images for statistical comparison generated by joint DARTEL setup, whereas 100% lesion was never detected using images generated by individual DARTEL setup.

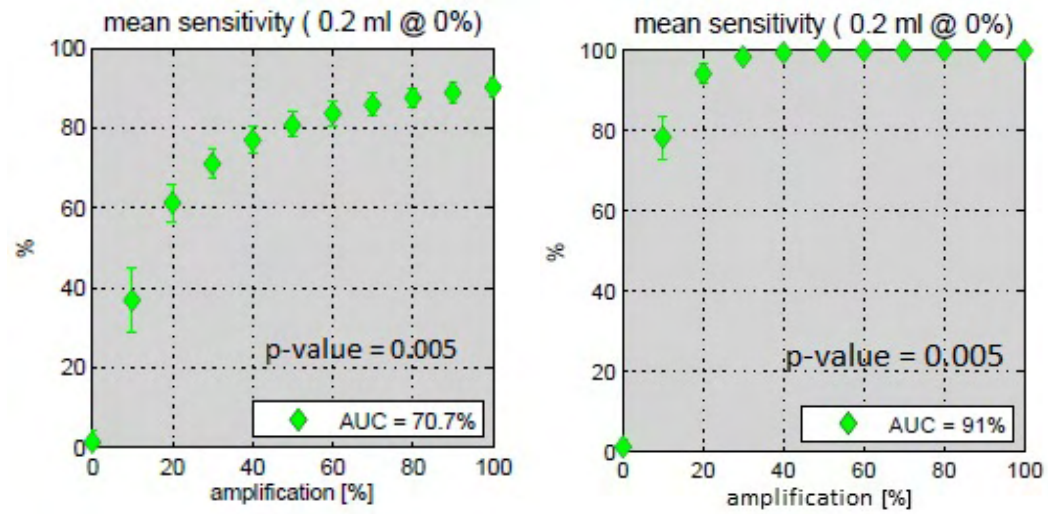


Fig. 0.2 Mean sensitivity curves of independent (left) and joint (right) DARTEL

Similar results to database – 2 were obtained for database – 1 also. When investigating white matter to detect lesion simulated by biomarker *a* (attenuation), almost 95% lesion was detected at 40% effect size using images for statistical comparison generated by joint DARTEL setup; whereas approximately 85% of lesion was detected using images generated using independent DARTEL setup.

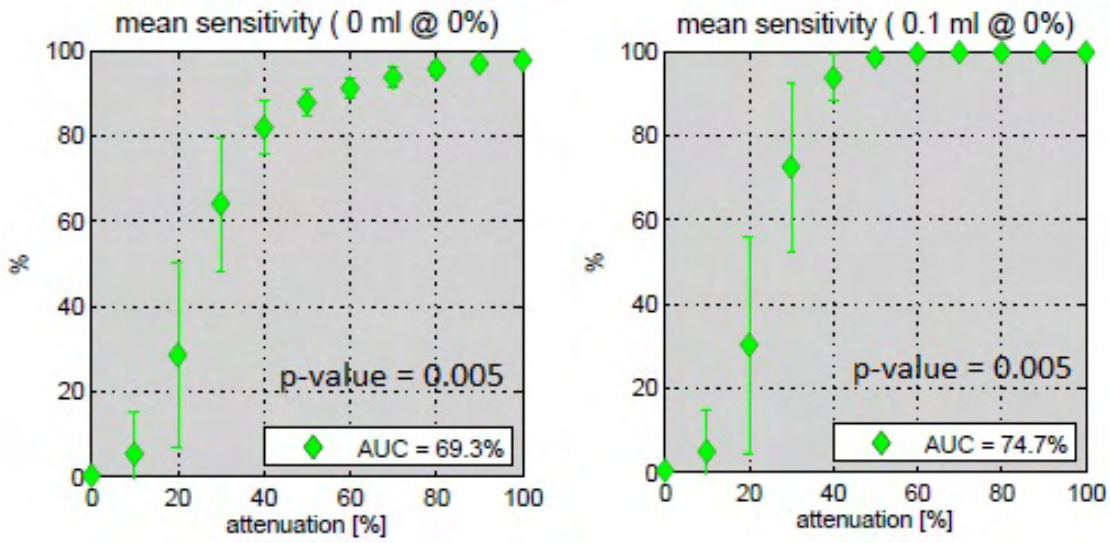


Fig. 0.3 Mean sensitivity curves of independent (left) and joint (right) DARTEL

When investigating grey matter to detect lesion simulated by biomarker *b* (amplification), almost 100% lesion in grey matter was detected at 30% effect size using joint DARTEL image files, whereas 100% lesion is not detected at 100% effect size also using DARTEL files generated by individual DARTEL process.

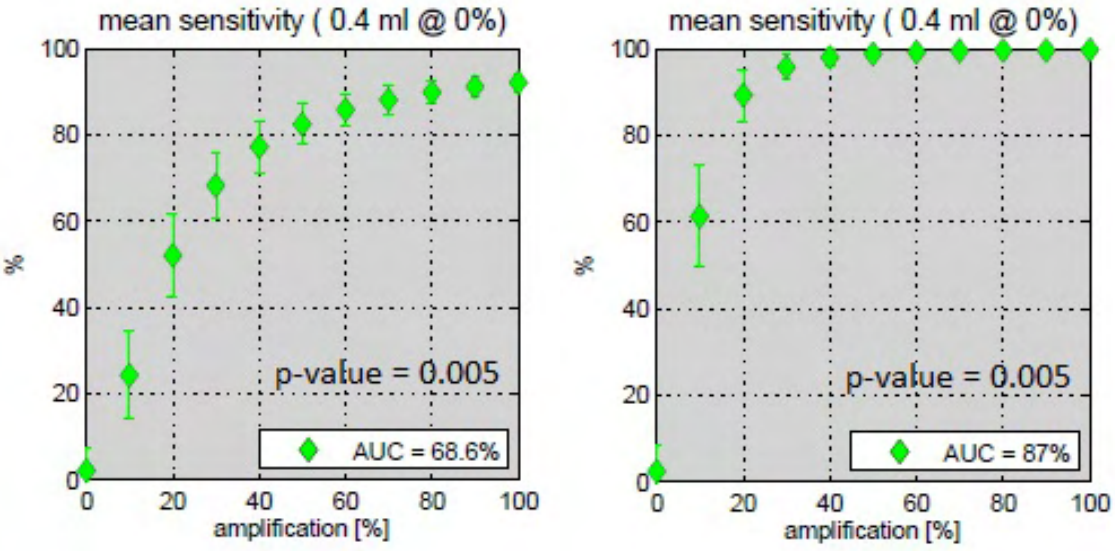


Fig. 0.4 Mean sensitivity curves of independent (left) and joint (right) DARTEL

DARTEL process when done independently consumes more computational power and time, as two DARTEL processes are run consecutively for grey and white matter, whereas the process consumes half the time when done jointly. Also 100% sensitivity in detecting hyper-intensity voxels is not achieved when DARTEL files generated by independent DARTEL process are used. This exhibits that joint DARTEL process is more efficient and produces better results compared to independent process.

Quantification

Accuracy, sensitivity and specificity are the measures used to quantify the efficiency of the statistical test. Accuracy is the measure of correctness of the statistical test in identifying and excluding a given condition. Sensitivity gives the proportion of actually diseased regions which are correctly identified by the test. Specificity gives the proportion of negatives that are correctly identified by the test.

Accuracy, sensitivity and specificity are defined by true positives, false positives true negatives and false negatives. True positives are the voxels in t-map which are highlighted and are really lesion voxels. False positives are those voxels which are not actually lesion voxels but highlighted due to inaccuracy in the test model. True negatives are the voxels which are not diseased and not detected as lesion in the statistical test also. False negatives are the voxels which are actually diseased regions but not classified as lesions by the test.

Table 13 Terms used to define accuracy, sensitivity and specificity

Outcome of statistical test	**Actual disease condition**	
	Positive	**Negative**
Positive	True positive	False positive
Negative	False negative	True negative

- $$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \quad (4.10)$$
- $$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (4.11)$$
- $$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \quad (4.12)$$

True positives (tp), false positives (fp), true negatives (tn) and false negatives (fn) can be measured by generating the respective images. These images can be generated by the *Image Calculator* routine of SPM. Image calculator is used to perform voxel-wise algebraic manipulations on a set of images such as binarizing, subtracting, taking a sum or mean of set of images, etc.

The expressions for generating tp, tn, fp and fn images are as follows:

- true positive = i4 .* (i5 > 0.2) .* (i6 > 0) *(4.13)*
- true negative = i4 .* (i5 <= 0.2) .* (i6 <= 0) *(4.14)*
- false positive = i4 .* (i5 <= 0.2) .* (i6 > 0) *(4.15)*
- false negative = i4 .* (i5 > 0.2) .* (i6 <= 0) *(4.16)*

where, i4 = white matter mask
i5 = lesion mask
i6 = hypo or hyper map

The white matter mask is used to restrict the area of brain under investigation. In the above equations, '.*' means voxel-wise multiplication of two images. *'ix > 0.2'* indicates that only the part of lesion mask image which has intensity more than *0.2* will be taken into account, the remaining lesion is ignored in masking and *' (ix ≤ 0.2)'* indicates that only the part of lesion mask image which has intensity less than *0.2* will be taken into account, the remaining lesion is ignored in masking.

Quantification Results of Optimized Framework for Database – 2 and 3

Database – 2

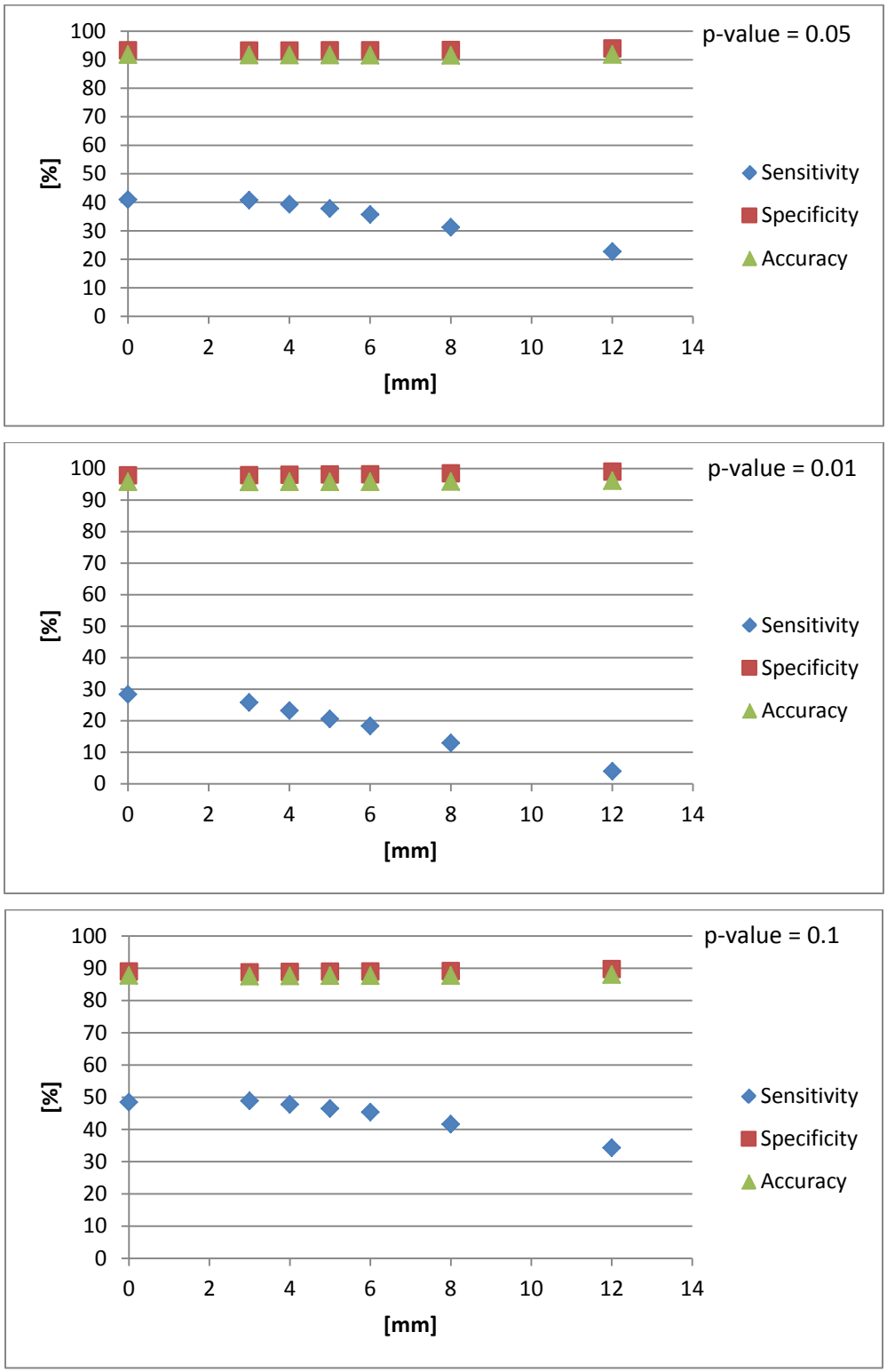


Fig. 0.5 Accuracy, sensitivity and specificity for hypo intense maps at 0.05, 0.01 and 0.1 significance levels

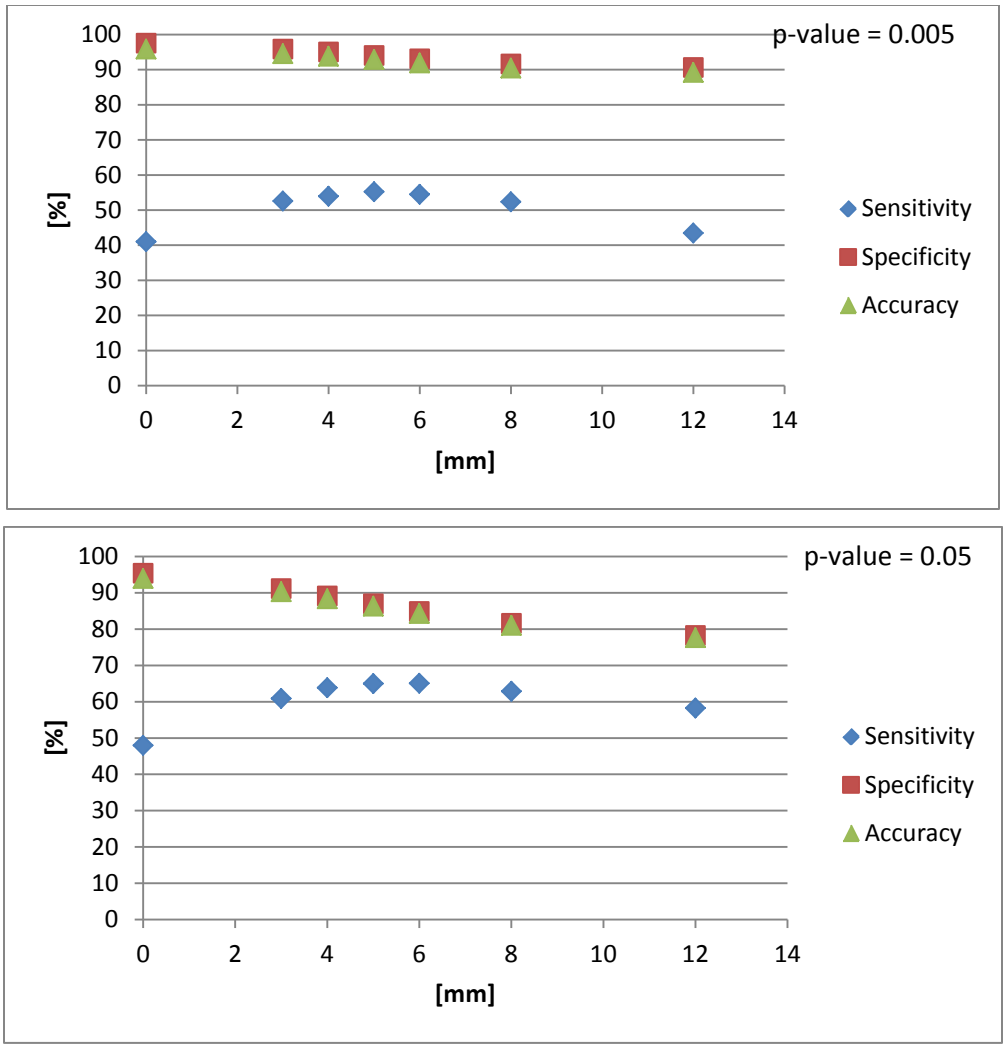


Fig. 0.6 Accuracy, sensitivity and specificity for hyper intense maps at 0.005 and 0.05 significance levels

Database – 3

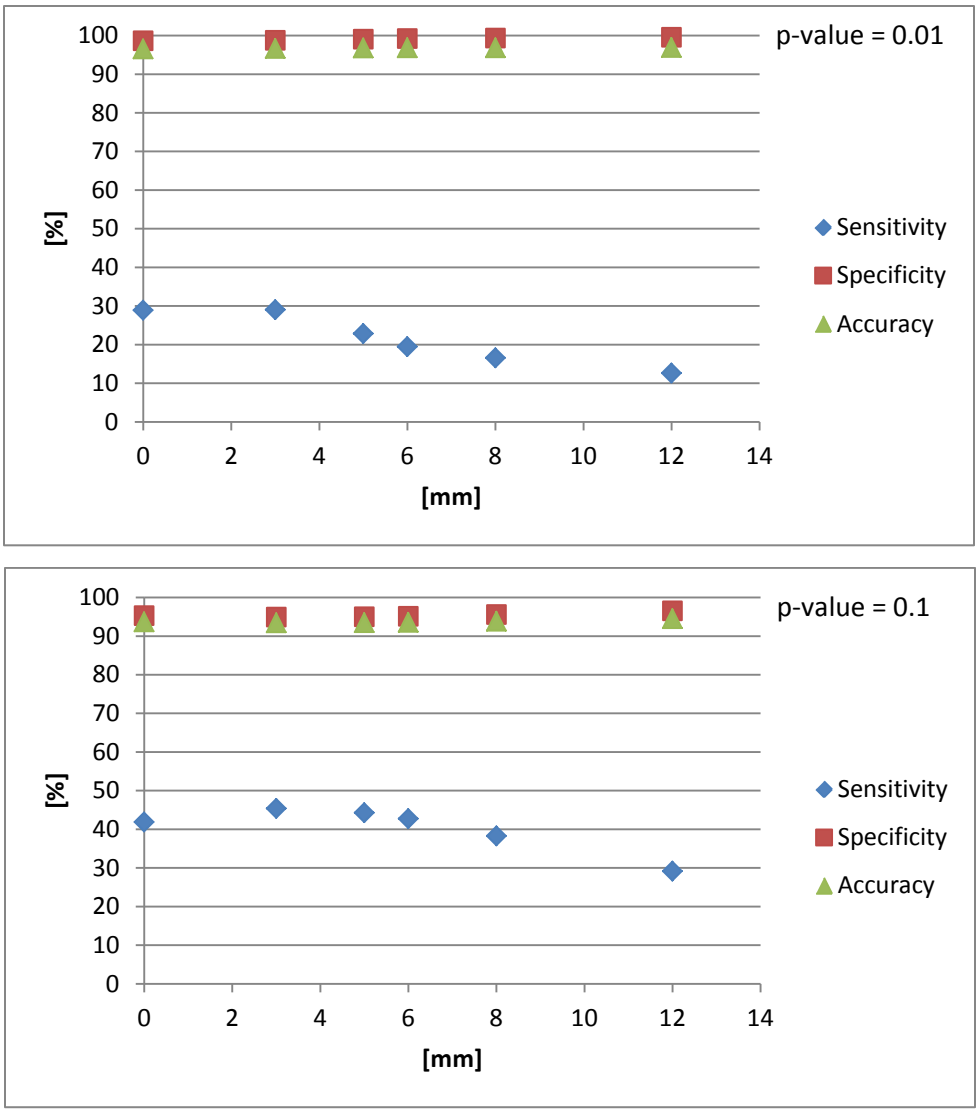


Fig. 0.7 Accuracy, sensitivity and specificity for hypo intense maps at 0.05 and 0.01 significance levels

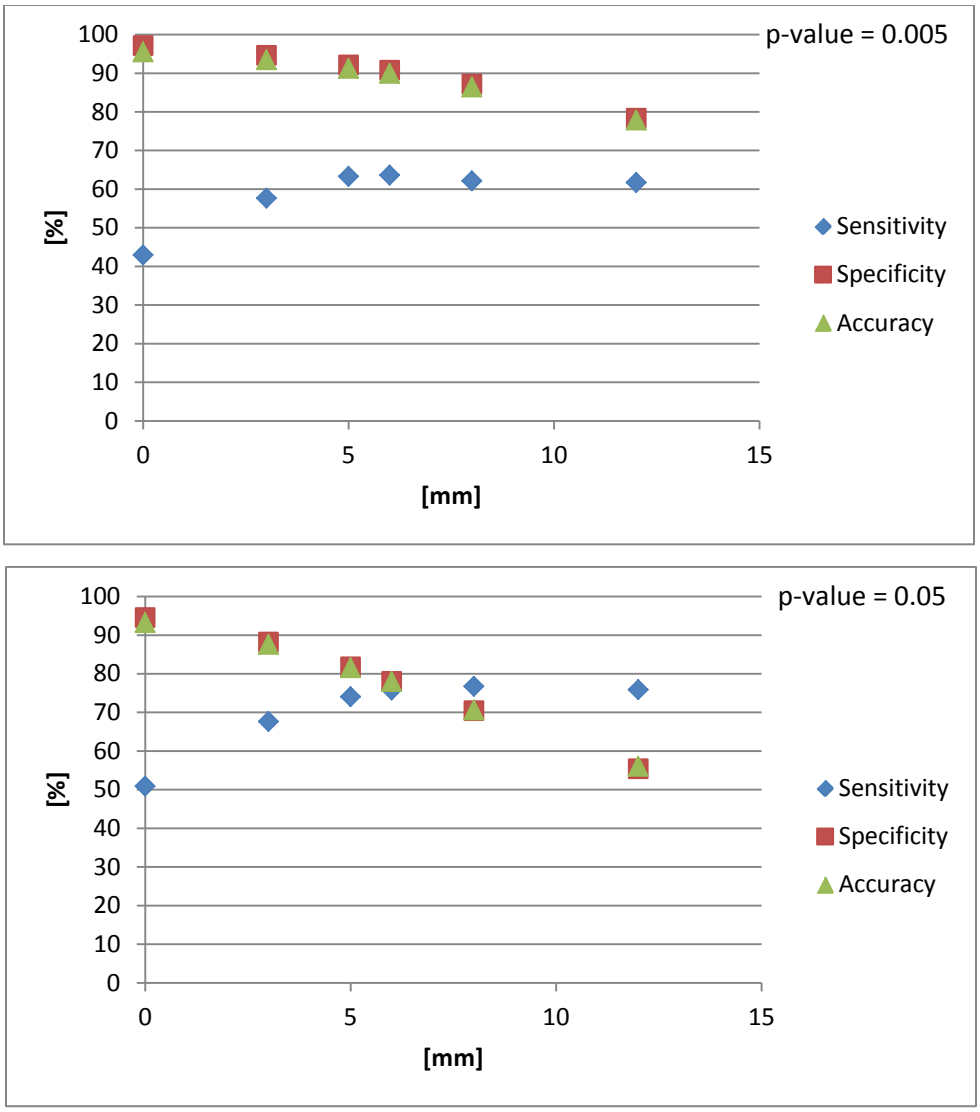


Fig. 0.8 Accuracy, sensitivity and specificity for hyper intense maps at 0.005 and 0.05 significance levels

VBM Results for Clinical Data

Subject – 2

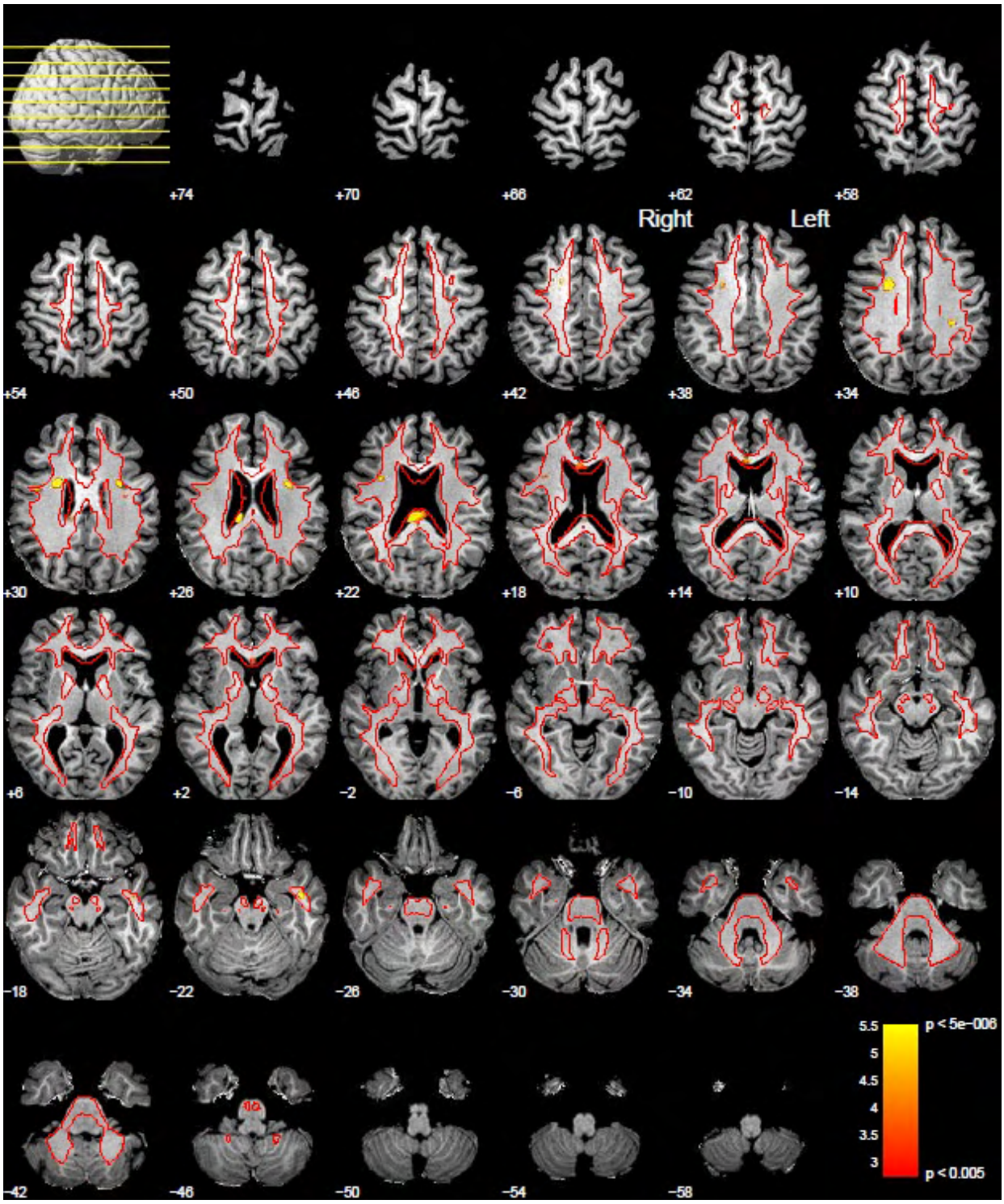


Fig. 0.9 t-map – subject 2 (hpyermap, significance level 0.005 and smoothing filter 4mm)

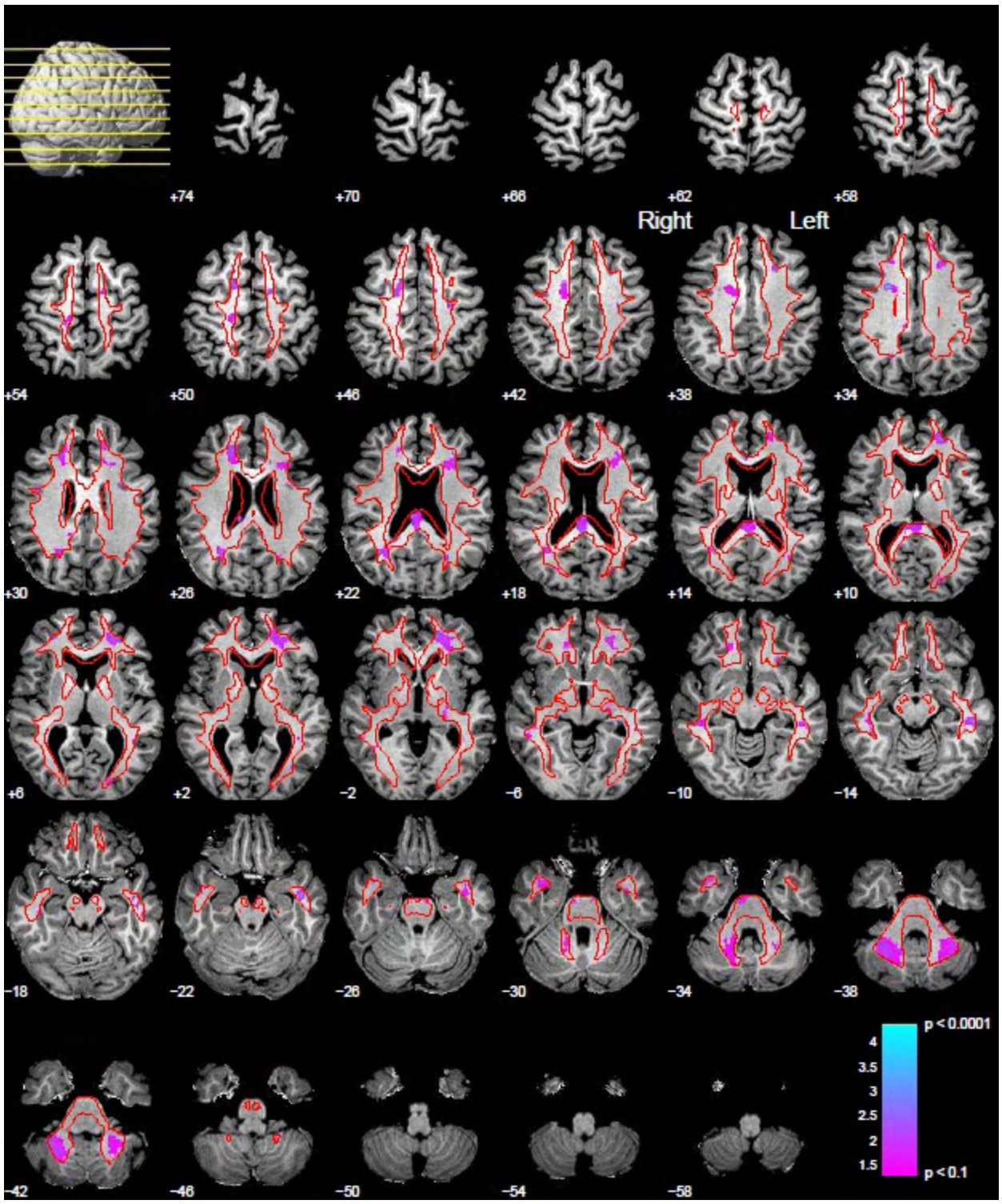


Fig. 0.10 t-map – subject 2 (hpyomap, significance level 0.1 and smoothing filter 3mm)

Subject – 3

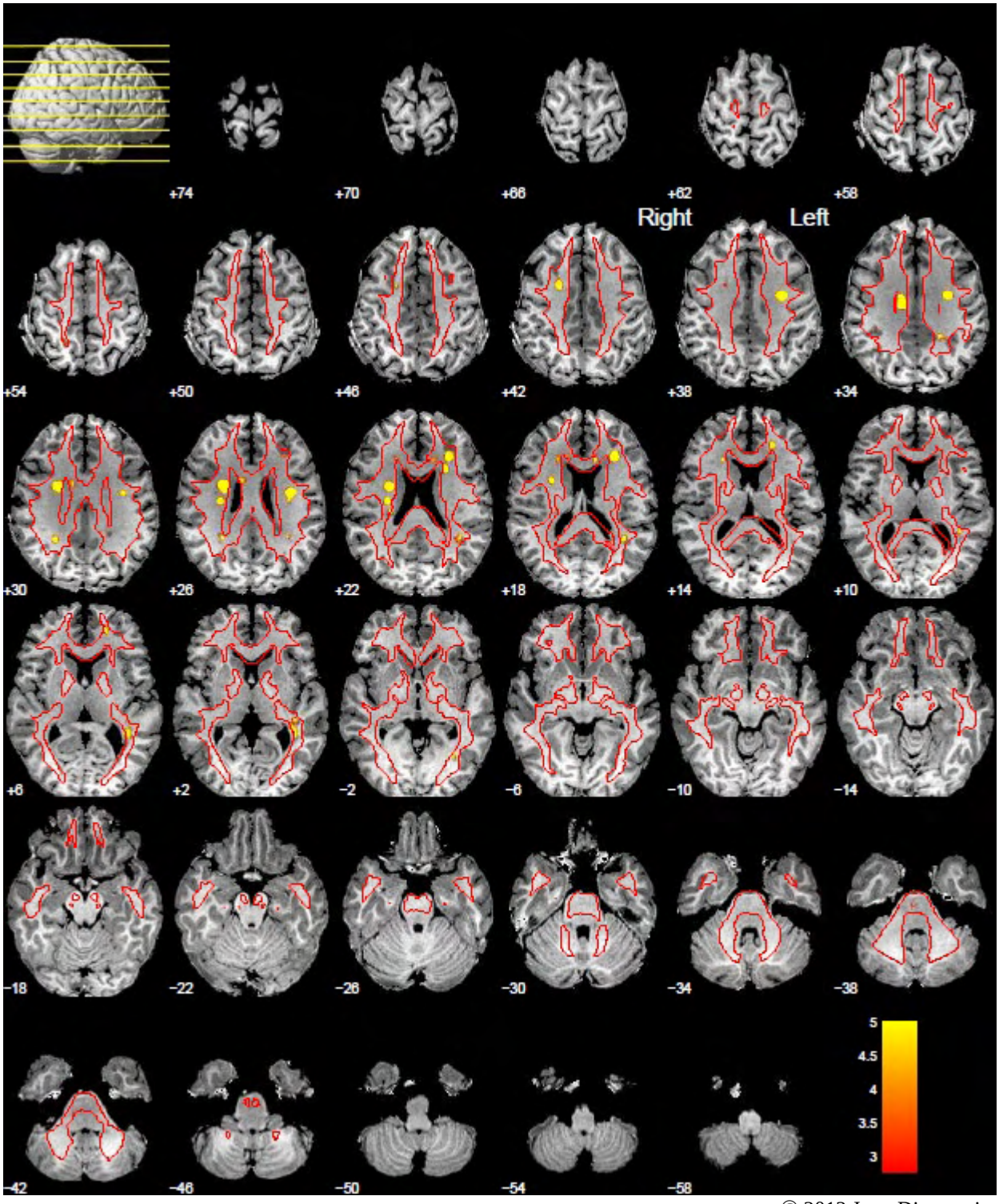


Fig. 0.11 t-map – subject 3 (hpyermap, significance level 0.005 and smoothing filter 4mm)

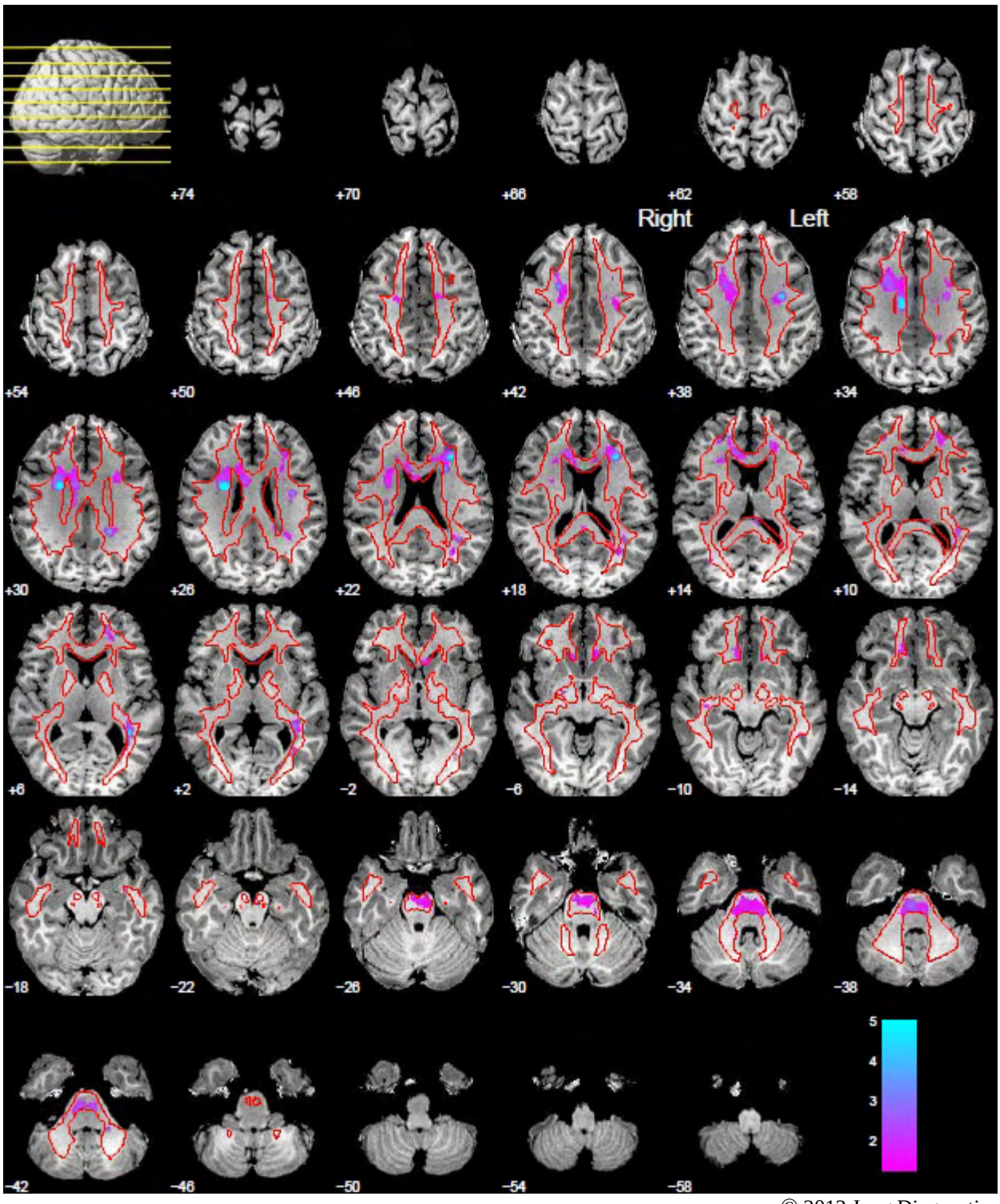


Fig. 0.12 t-map – subject 3 (hpyomap, significance level 0.1 and smoothing filter 3mm)

Subject – 5

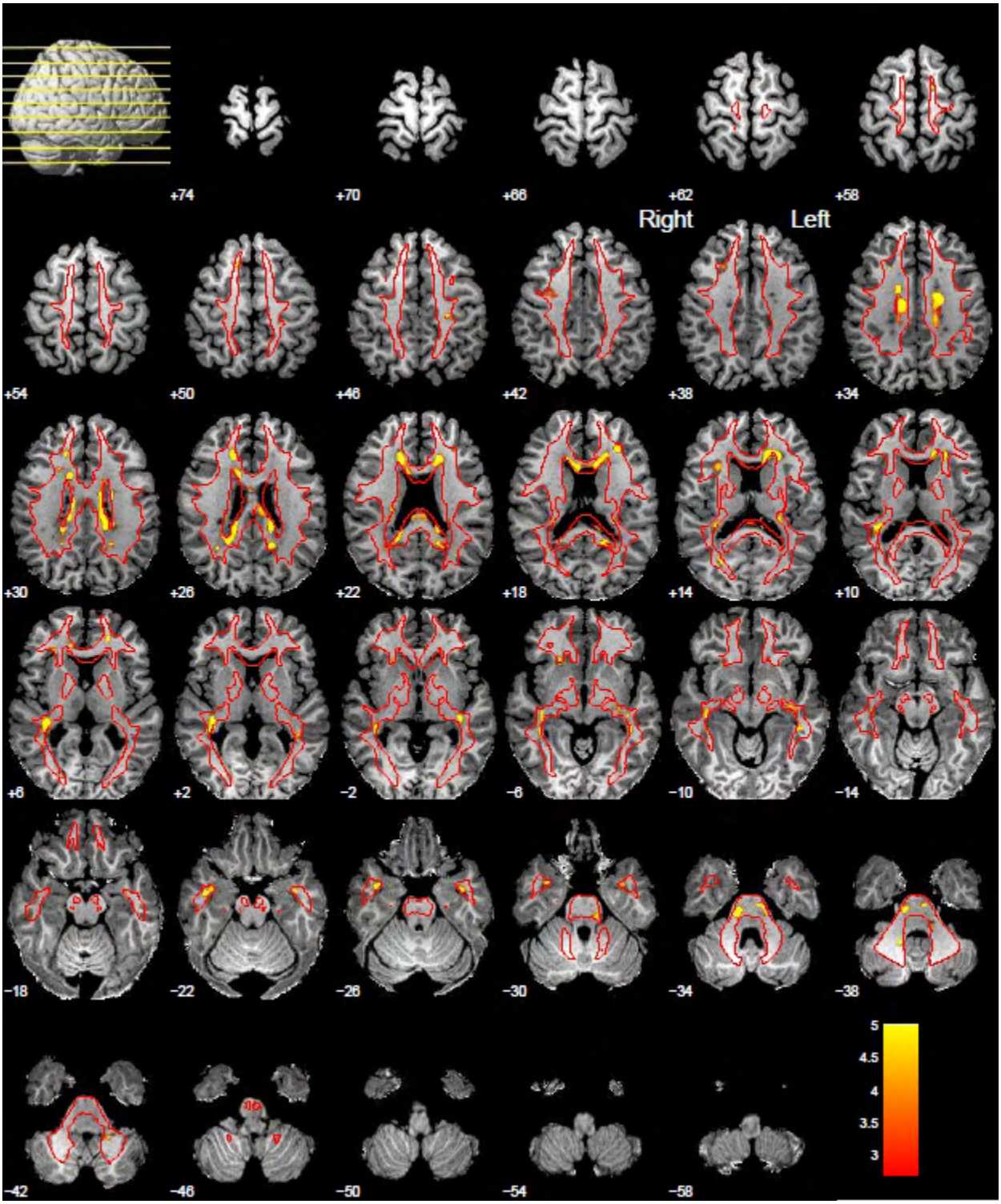


Fig. 0.13 t-map – subject 5 (hpyermap, significance level 0.005 and smoothing filter 4mm)

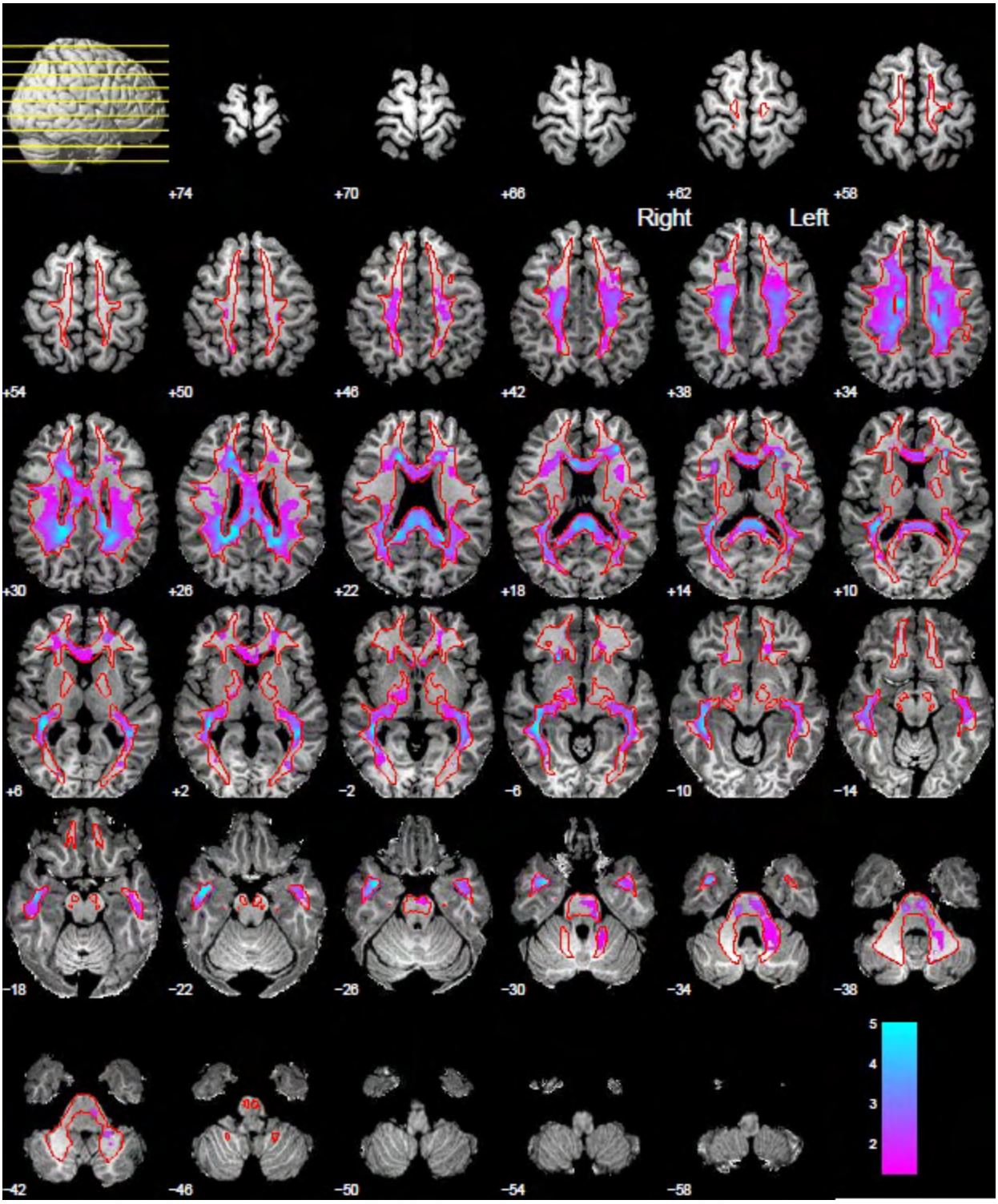


Fig. 0.14 t-map – subject 5 (hpyomap, significance level 0.1 and smoothing filter 3mm)

Subject – 6

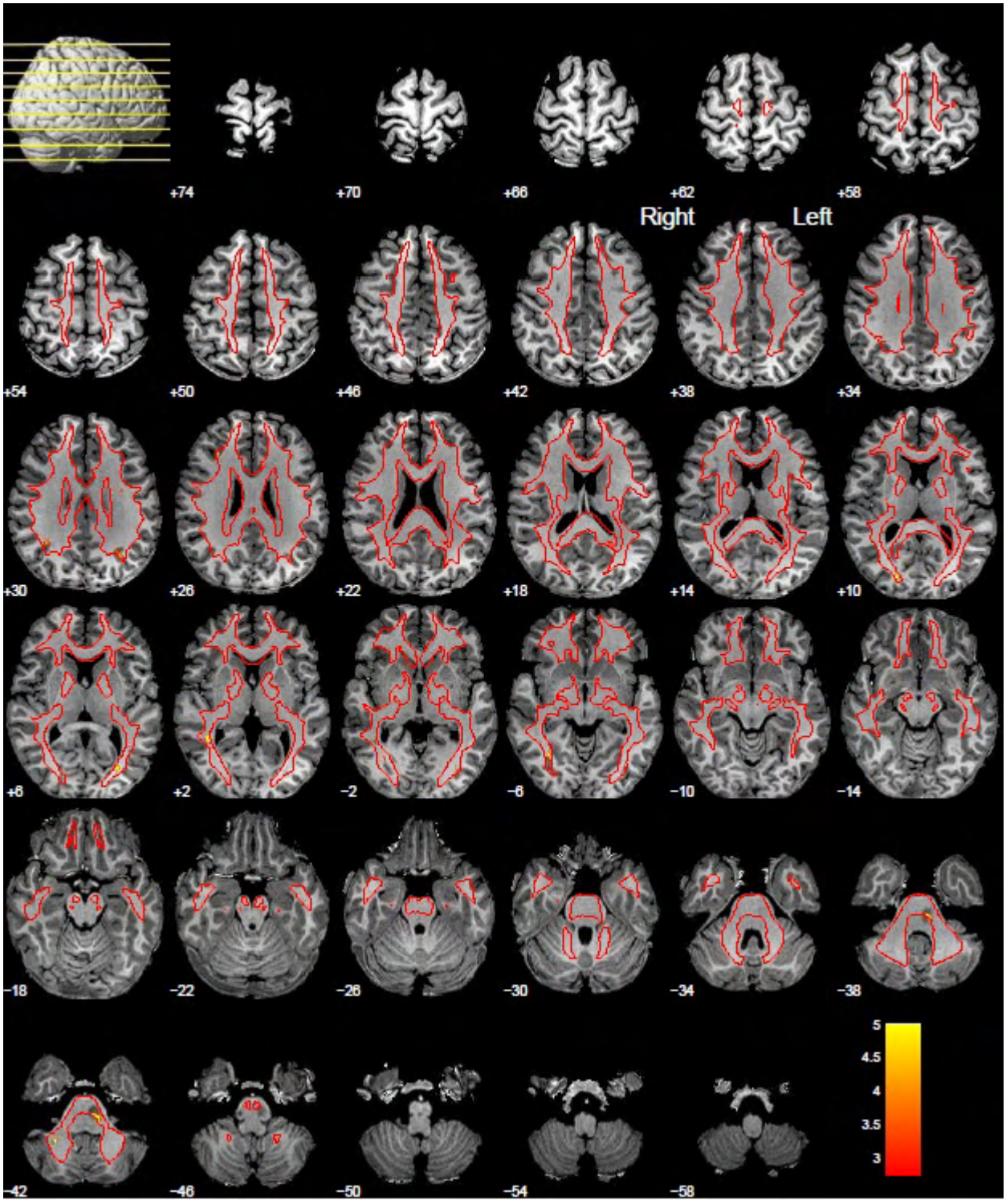


Fig. 0.15 t-map – subject 6 (hpyermap, significance level 0.005 and smoothing filter 4mm)

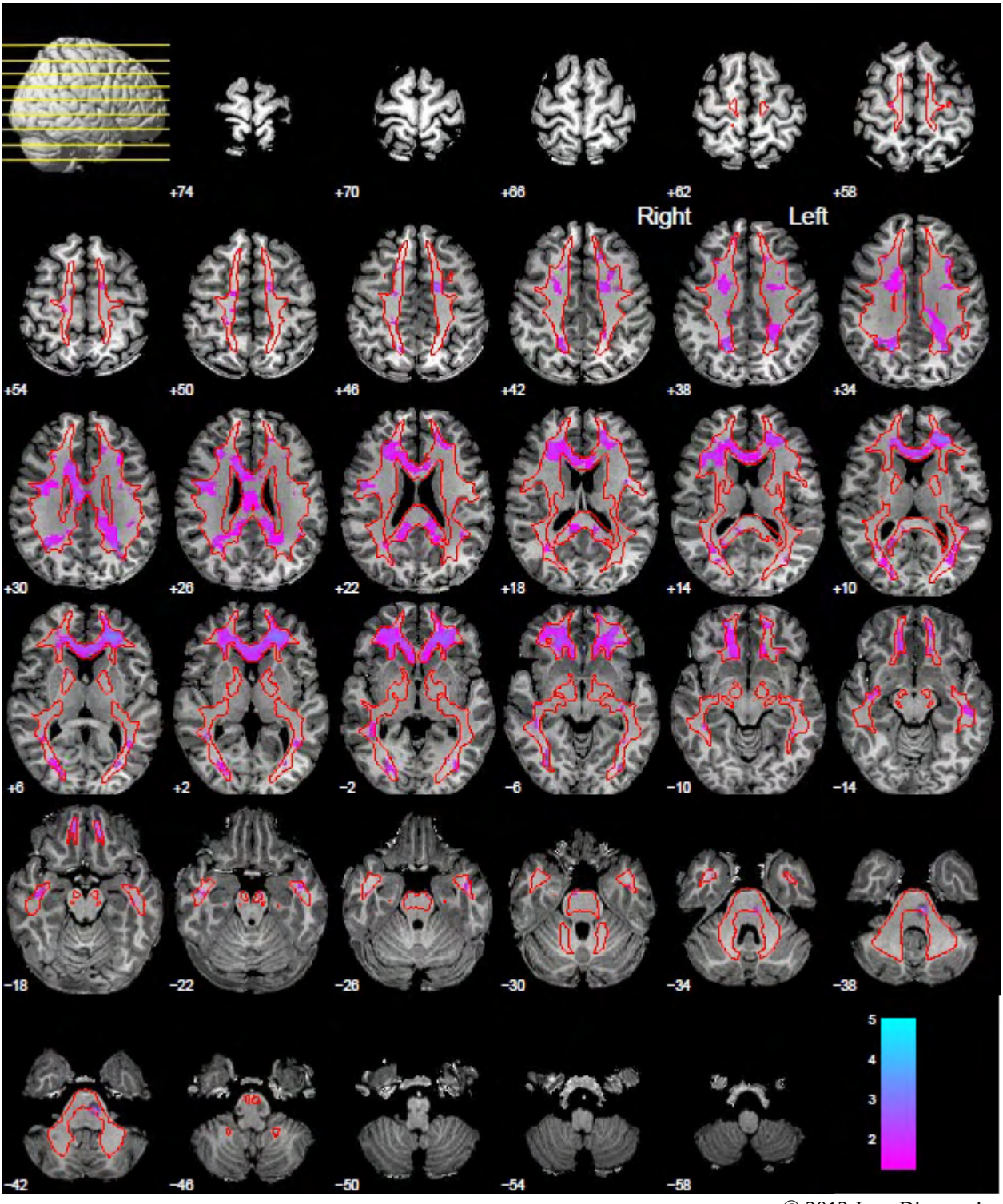


Fig. 0.16 t-map – subject 6 (hpyomap, significance level 0.1 and smoothing filter 3mm)