

Machine learning for image-based classification of Alzheimer's disease

Katherine Rachel Gray

A dissertation submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

of

Imperial College London

October 3, 2012

Department of Computing

Imperial College London

Declaration of originality

I declare that the work presented in this thesis is my own, unless specifically acknowledged.

Katherine R. Gray

Abstract

Imaging biomarkers for Alzheimer’s disease are important for improved diagnosis and monitoring, as well as drug discovery. Automated image-based classification of individual patients could provide valuable support for clinicians. This work investigates machine learning methods aimed at the early identification of Alzheimer’s disease, and prediction of progression in mild cognitive impairment. Data are obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) and the Australian Imaging, Biomarker and Lifestyle Flagship Study of Ageing (AIBL).

Multi-region analyses of cross-sectional and longitudinal FDG-PET images from ADNI are performed. Information extracted from FDG-PET images acquired at a single timepoint is used to achieve classification results comparable with those obtained using data from research-quality MRI, or cerebrospinal fluid biomarkers. The incorporation of longitudinal information results in improved classification performance.

Changes in multiple biomarkers may provide complementary information for the diagnosis and prognosis of Alzheimer’s disease. A multi-modality classification framework based on random forest-derived similarities is applied to imaging and biological data from ADNI. Random forests provide consistent similarities for multiple modalities, facilitating the combination of different types of features. Classification based on the combination of MRI volumes, FDG-PET intensities, cerebrospinal fluid biomarkers, and genetics out-performs classification based on any individual modality.

Multi-region analysis of MRI acquired at a single timepoint is used to show volumetric differences in cognitively normal individuals differing in amyloid-based risk status for the development of Alzheimer’s disease. Reduced volumes in temporo-parietal and orbito-frontal regions in high-risk individuals from both ADNI and AIBL could be indicative of early signs of neurodegeneration. This suggests that volumetric MRI can reveal structural brain changes preceding the onset of clinical symptoms.

Taken together, these results suggest that image-based classification can support diagnosis in Alzheimer’s disease and preceding stages. Future work may lead to more finely meshed prognostic data that may be useful clinically and for research.

Acknowledgements

I would first like to express my gratitude to both my supervisors, Daniel Rueckert and Alexander Hammers, for their enthusiasm and encouragement over the past few years.

I would also like to thank the many collaborators who have helped me along the way, in particular Rolf Heckemann, for providing motivation and invaluable advice, Paul Aljabar, for his mathematical assistance, Igor Yakushev, for kindly providing his reference cluster images, and Robin Wolz, for helping me get started. Thanks also to all members of the BioMedIA group, both past and present, for providing a great working environment.

I am grateful to the whole PredictAD team for their help during the brief time we were working together, and members of the Imperial Imaging Group, especially Jo Hajnal and Shiva Keihaninejad. I must also thank the EPSRC for financial support, and both the ADNI and AIBL investigators for providing the clinical and imaging data without which this research would not have been possible.

Finally, I am most grateful to my friends, especially everyone at Burles Farm, and, above all, my family and Jad for their support and encouragement throughout my studies.

Many thanks to you all!

Contents

1	Introduction	19
1.1	Alzheimer's disease	19
1.2	Neuroanatomy	21
1.3	Neuropathology	22
1.4	Neuroimaging	24
1.4.1	Positron emission tomography	24
1.4.2	Magnetic resonance imaging	28
1.5	Biomarkers for Alzheimer's disease	32
1.5.1	Cerebrospinal fluid	33
1.5.2	Magnetic resonance imaging	34
1.5.3	Fluorodeoxyglucose positron emission tomography	36
1.5.4	Pittsburgh compound B positron emission tomography	37
1.6	Research contributions and thesis outline	39
2	Background: PET and MR image analysis	41
2.1	Introduction	41

2.2	Image registration	41
2.2.1	Transformation model	42
2.2.2	Optimisation method	50
2.2.3	Similarity metric	50
2.2.4	Regularisation method	53
2.2.5	Interpolation method	53
2.3	Anatomical segmentation	54
2.3.1	Manual segmentation	54
2.3.2	Brain atlases	55
2.3.3	Automated multi-atlas segmentation	56
2.4	Statistical parametric mapping	58
2.4.1	Image processing	58
2.4.2	Statistical analyses	59
2.4.3	Multiple comparison correction	60
2.5	Conclusion	61
3	Background: machine learning	62
3.1	Introduction	62
3.2	Classification algorithms	62
3.2.1	Linear discriminant functions	63
3.2.2	Support vector machines	65
3.2.3	Boosting	69

3.2.4	Random forests	71
3.3	Classifier performance	74
3.3.1	Performance metrics	74
3.3.2	Cross-validation	76
3.4	Application to Alzheimer’s disease	77
3.5	Conclusion	79
4	Multi-region baseline FDG-PET for classification	80
4.1	Introduction	80
4.2	Imaging data	81
4.3	Image acquisition and pre-processing	83
4.3.1	ADNI FDG-PET acquisition	84
4.3.2	FDG-PET pre-processing	85
4.3.3	ADNI MRI acquisition and pre-processing	86
4.3.4	Co-registration of FDG-PET with MRI	86
4.4	MRI anatomical segmentation	87
4.4.1	Additional image processing	88
4.4.2	Segmentation procedure	89
4.4.3	Tissue class masking	90
4.5	FDG-PET intensity normalisation	91
4.5.1	Additional image processing	91
4.5.2	Cerebral global mean intensity normalisation	93

4.5.3	Reference cluster intensity normalisation	95
4.6	Multi-region image analysis	98
4.6.1	Region-based feature extraction	99
4.6.2	Region-based group differences	99
4.7	Classification experiments	101
4.7.1	Methods	101
4.7.2	Results	103
4.7.3	Discussion	104
4.8	Conclusion	106
5	Multi-region longitudinal FDG-PET for classification	107
5.1	Introduction	107
5.2	Imaging data	109
5.3	Image acquisition and pre-processing	110
5.4	Preliminary studies	111
5.4.1	Hippocampal segmentation	112
5.4.2	Image analysis	113
5.4.3	Classification experiments	114
5.4.4	Discussion	116
5.5	Multi-region segmentation	116
5.6	Multi-region image analysis	118
5.7	Classification experiments	120

5.7.1	Methods	120
5.7.2	Results	121
5.7.3	Discussion	123
5.8	Conclusion	125
6	Random forest-based similarities for multi-modality classification	126
6.1	Introduction	126
6.2	Imaging and biological data	128
6.2.1	Region-based MRI features	128
6.2.2	Voxel-based FDG-PET features	129
6.2.3	Biological CSF and ApoE genotype features	129
6.3	Multi-modality classification framework	130
6.4	Combining FDG-PET and MR imaging data	131
6.4.1	Classification methodology	131
6.4.2	Single-modality classification results	132
6.4.3	Single-modality similarity-based classification results	133
6.4.4	Multi-modality similarity-based classification results	134
6.4.5	Discussion	135
6.5	Combining imaging and biological data	136
6.5.1	Classification methodology	136
6.5.2	Single-modality classification results	136
6.5.3	Single-modality similarity-based classification results	138

6.5.4	Multi-modality similarity-based classification results	139
6.5.5	Discussion	140
6.6	Conclusion	143
7	Early identification of Alzheimer’s disease	145
7.1	Introduction	145
7.2	Imaging and biological data	146
7.2.1	ADNI participants	147
7.2.2	AIBL participants	147
7.3	MRI acquisition and anatomical segmentation	148
7.3.1	AIBL MRI acquisition	148
7.3.2	AIBL MRI anatomical segmentation	148
7.3.3	Computation of regional MRI volumes	151
7.4	Amyloid-based risk status	152
7.4.1	ADNI participants	152
7.4.2	AIBL participants	153
7.5	Volumetric differences between risk groups	157
7.5.1	ADNI participants	158
7.5.2	AIBL 1.5 T participants	159
7.5.3	AIBL 3 T participants	160
7.5.4	Discussion	161
7.6	Conclusion	163

8 Overall conclusion	164
8.1 Contributions	164
8.2 Future work	166
9 Publications	168
Bibliography	171
A Hammers brain atlases	201
B Excluded images	204

List of Tables

3.1	Confusion matrix for a binary classifier.	74
3.2	Summary of classification results based on cross-sectional ADNI FDG-PET data.	78
3.3	Summary of classification results based on MR imaging data.	79
4.1	Clinical and demographic information for the study population.	83
4.2	MAPER steps for aligning a single atlas MRI with the target MRI.	89
4.3	FWHM of the scanner-specific Gaussian smoothing kernels.	93
4.4	Classification results for SVM and Ada-LDA classifiers.	103
5.1	Clinical and demographic information for the study population.	109
5.2	Classification results based on longitudinal hippocampal FDG-PET.	115
5.3	Classification results based on longitudinal multi-region FDG-PET.	122
5.4	Classification results based on longitudinal multi-region FDG-PET after regression.	123
6.1	Clinical and demographic information for participants with CSF data available.	128
6.2	Groupwise CSF measures and genetic information for the study population.	130
6.3	Single-modality classification results based on the original data.	132

6.4	Single-modality classification results based on the embedded data.	134
6.5	Multi-modality classification results based on the jointly embedded data.. . . .	134
6.6	Single-modality classification results based on the original data.	137
6.7	Single-modality classification results based on the embedded data.	139
6.8	Multi-modality classification results based on the jointly embedded data.	140
7.1	Clinical and demographic information for the study populations.	147
7.2	Details of the four PINCRAM iterations.	151
7.3	Regional volume differences between high- and low-risk sub-sets in ADNI.	158
7.4	Regional volume differences between high- and low-risk sub-sets in AIBL (1.5 T).	160
7.5	Regional volume differences between high- and low-risk sub-sets in AIBL (3 T).	161
8.1	Comparison of key results from Chapters 4, 5 and 6.	164
8.2	Summary of the ADNI studies.	166
A.1	Anatomically defined regions manually delineated in the Hammers atlases.	203
B.1	Participants excluded from the work described in Chapters 4 and 5.	205

List of Figures

1.1	Illustrative timeline of AD progression.	20
1.2	Sagittal views of the brain, showing its gross anatomy.	21
1.3	Sagittal views of the ventricular system.	22
1.4	Axial views of the brain, showing the sub-structures of the medial temporal lobe.	22
1.5	Stages of PET image acquisition.	26
1.6	Image reconstruction by simple back projection.	27
1.7	Hydrogen atoms in the presence of a static magnetic field.	29
1.8	The application of Fourier transforms to MRI.	30
1.9	Example MRI pulse sequence.	31
1.10	Hypothetical temporal model of biomarker dynamics during AD progression.	33
1.11	MR images of healthy individuals and AD patients.	35
1.12	FDG-PET images of healthy individuals and AD patients.	36
1.13	PiB-PET images of healthy individuals and AD patients.	38
2.1	Illustration of possible rigid transformations.	43
2.2	Illustration of possible affine transformations.	45

2.3	Illustration of a nonrigid transformation.	47
2.4	Schematic sagittal brain view showing the anterior and posterior commissures.	55
2.5	One of the manually delineated brain atlases used in this work.	56
2.6	Schematic representation of the multi-atlas segmentation procedure.	57
2.7	Illustrative example of the SPM analysis procedure.	59
3.1	2-D illustration of the decision surface of a Fisher linear discriminant function.	63
3.2	2-D illustration of the construction of a maximum-margin hyperplane.	65
3.3	A nonlinear boundary becomes a linear hyperplane in higher-dimensional space.	68
3.4	Illustration of the first two iterations in a typical boosting procedure.	70
3.5	Illustration of a random forest, showing two trees in detail.	71
3.6	Illustration of the ROC curve for a binary classifier.	76
4.1	Summary of exclusions.	83
4.2	Summary of FDG-PET and MRI acquisition and processing steps.	84
4.3	Number of baseline FDG-PET images acquired using each ADNI scanner model.	84
4.4	Illustration of the PET-MRI co-registration process.	87
4.5	Illustration of the brain masking process.	88
4.6	Illustration of the tissue classification results.	89
4.7	Typical examples of a full and masked segmentation.	90
4.8	Typical examples showing alignment of MR and PET images to the MNI template.	92
4.9	Typical example images illustrating the FDG-PET smoothing process.	93

4.10	Reduced FDG-PET intensity in AD versus HC using CGM normalisation. . . .	94
4.11	Reduced FDG-PET intensity in MCI versus HC using CGM normalisation. . . .	95
4.12	Increased FDG-PET intensity in AD and MCI versus HC.	96
4.13	Reduced FDG-PET intensity in AD versus HC using cluster normalisation. . . .	97
4.14	Reduced FDG-PET intensity in MCI versus HC using cluster normalisation. . .	97
4.15	Typical examples of the images required for regional FDG-PET feature extraction.	99
4.16	Group differences in the hippocampus and middle and inferior temporal gyri. . .	100
4.17	Regional t-values between AD patients and HC, and MCI patients and HC. . . .	100
5.1	Summary of exclusions.	109
5.2	Summary of the FDG-PET and MR image acquisition and pre-processing steps.	110
5.3	Alignment to baseline MRI space of 12-month MRI and FDG-PET images. . . .	112
5.4	Typical examples of the hippocampal segmentations.	113
5.5	Group differences in the hippocampus.	114
5.6	Propagation of the intracranial brain mask.	117
5.7	Typical examples of masked multi-region segmentations.	117
5.8	Regional t-values based on cross-sectional and longitudinal FDG-PET features. .	119
5.9	Classification accuracies based on the five feature sets studied.	121
5.10	ROC curves for the best performing combined feature.	122
6.1	Schematic overview of the proposed multi-modality classification approach. . . .	130
6.2	Feature importances for distinguishing between clinical groups.	133

6.3	Feature importances for distinguishing between clinical groups.	137
6.4	Similarity matrices for all three group pairs based on the four modalities.	138
6.5	Cobweb plots showing the distribution of parameters selected for classification.	140
7.1	Illustration of the bias correction procedure.	149
7.2	Schematic representation of a single PINCRAM iteration.	150
7.3	Illustration of the PINCRAM procedure.	151
7.4	Summary of the ADNI CSF $A\beta$ measures.	153
7.5	Illustration of PET-MRI co-registration.	154
7.6	Images required for assessment of extent of cortical β -amyloid deposition.	155
7.7	Summary of AIBL 1.5 T PiB-PET SUVR measures.	156
7.8	Summary of AIBL 3 T PiB-PET SUVR measures.	156
7.9	Illustration of a P plot.	157
7.10	Regional t-values between volumes in high- and low-risk sub-sets in ADNI.	158
7.11	Regional t-values between volumes in high- and low-risk sub-sets in AIBL (1.5 T).	159
7.12	Regional t-values between volumes in high- and low-risk sub-sets in AIBL (3 T).	160

Chapter 1

Introduction

1.1 Alzheimer's disease

Alzheimer's disease (AD), named after the German physician Alois Alzheimer, is a condition defined by progressive dementia and the abundant presence in the brain of characteristic neuropathological structures. The earliest symptom is generally memory loss, followed by further functional and cognitive decline, such that patients become gradually less able to perform even basic tasks ([de Leon, 1999](#)). AD is the most common cause of dementia in the elderly, with a worldwide prevalence that is expected to rise, as the population ages, from the 26.6 million reported in 2006 to over 100 million by 2050 ([Brookmeyer et al., 2007](#)).

There is currently no disease-modifying therapy for AD; however, symptomatic treatments can help patients to maintain mental function and manage the behavioural symptoms. Ongoing clinical trials are focused on the development of new treatments, including those aimed at lowering the risk of developing the disease or delaying its onset and progression ([Klafki et al., 2006](#)). As illustrated in [Figure 1.1](#), changes associated with AD are thought to start occurring many years before the onset of clinical symptoms. Any disease-modifying or causal therapy would therefore likely be of greatest benefit to asymptomatic individuals at high risk of developing AD, so-called pre-symptomatic patients. Amnesic mild cognitive impairment (MCI) is of interest because this can be a transitional stage between the cognitive decline associated with

normal ageing and established AD. Memory is impaired in MCI, although general cognitive function is preserved, and patients are at increased risk of developing AD. The yearly rate of conversion from MCI to AD is around 12%, in contrast to the 1-2% yearly rate of conversion reported in an age-matched general population (Petersen, 2004).

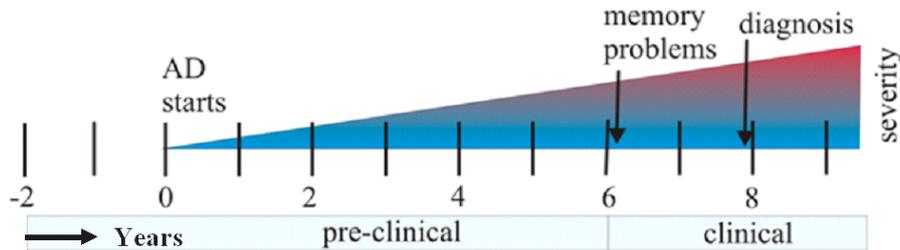


Figure 1.1: An illustrative timeline of AD progression. Produced by Jyrki Lötjönen, VTT Technical Research Centre of Finland.

A diagnosis of AD is made according to consensus criteria such as the NINCDS-ADRDA¹ Alzheimer's Criteria (McKhann et al., 1984), which provide guidelines for the classification of patients as having definite, probable, or possible AD. A diagnosis of definite AD requires that neuropathological findings be confirmed by a direct analysis of brain tissue samples, which may be obtained either at autopsy or from a brain biopsy. Since their proposal in 1984, studies have shown these criteria to have a diagnostic accuracy of up to 90% when validated against neuropathological gold standards (Ranginwala et al., 2008; Rasmusson et al., 1996). There are, however, several significant challenges to be addressed. These include pre-symptomatic diagnosis, differential diagnosis, and the assessment and prediction of progression. Research has shown biochemical and neuroimaging biomarkers to have diagnostic and prognostic value for AD, and recently published revisions to the consensus criteria aim to incorporate these advances (Albert et al., 2011; McKhann et al., 2011; Sperling et al., 2011).

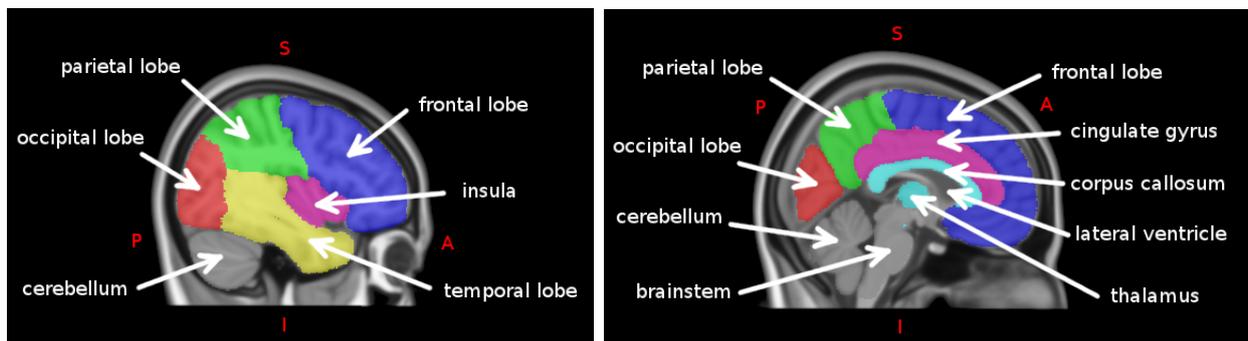
A delay of one year in both disease onset and progression would reduce the number of AD cases in 2050 by an estimated 10% (Brookmeyer et al., 2007). The early identification of pre-symptomatic patients is therefore important to allow the recruitment of appropriate participants for clinical trials. If a successful disease-modifying therapy for AD were to be developed, early identification would become even more important to allow targeting of patients for whom the

¹National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (now known as the Alzheimer's Association)

treatment may be most effective. The aim of the research presented in this thesis is to use machine learning methods with data from an array of diagnostic techniques to identify patients at highest risk of future cognitive decline.

1.2 Neuroanatomy

The human brain, illustrated in Figure 1.2, is composed mainly of two cerebral hemispheres, each of which is divided into four lobes: frontal, temporal, parietal and occipital. Each hemisphere includes a cortex of grey matter containing the neuronal cell bodies. The cortical surface is folded into ridges (gyri) and grooves (sulci). Other cortical regions relevant to the study of AD include the cingulate gyrus and insula. The insula is folded deep within the lateral sulcus between the frontal and temporal lobes. On the lateral surface of the brain, it is covered by the operculum, which is formed from portions of the frontal, temporal and parietal lobes.



(a) Lateral view

(b) Medial view

Figure 1.2: Sagittal views of the right hemisphere of the brain, showing its gross anatomy. S: superior, I: inferior, A: anterior, P: posterior.

The cortex surrounds a core of white matter, consisting mainly of myelinated axons connecting the cell bodies. The largest white matter structure in the brain is the corpus callosum, a bundle of axons connecting the left and right cerebral hemispheres. Embedded within the cerebral white matter are deep grey matter structures, including the basal ganglia and thalamus. At the base of the brain, underneath the cerebral hemispheres, are the cerebellum and brainstem. The brainstem is continuous with the spinal cord.

The brain is separated from the skull by three layers of tissue known as meninges: the dura, the arachnoid and the pia. To protect and support the brain, cerebrospinal fluid (CSF) fills the subarachnoid space, as well as a continuous system of four cavities known as ventricles. These are illustrated in Figure 1.3. CSF also fills the central canal of the spinal cord.

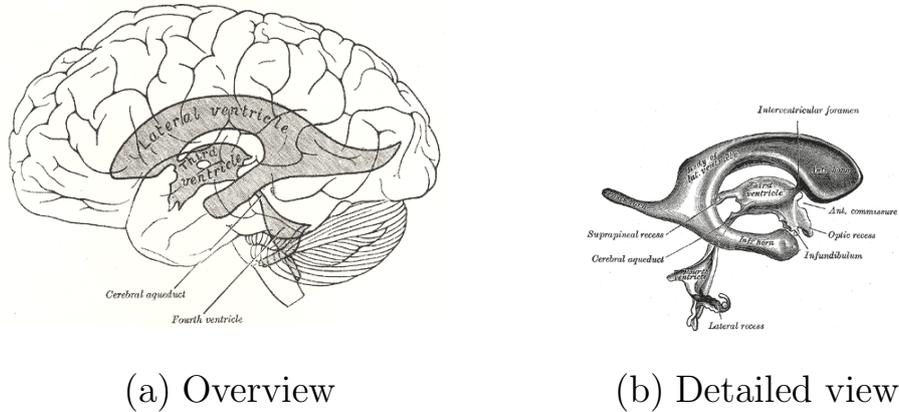


Figure 1.3: Sagittal views of the ventricular system (Gray, 1918).

Pathological changes associated with the development of AD begin in the medial temporal lobes. Several important sub-structures within this region are illustrated in Figure 1.4.

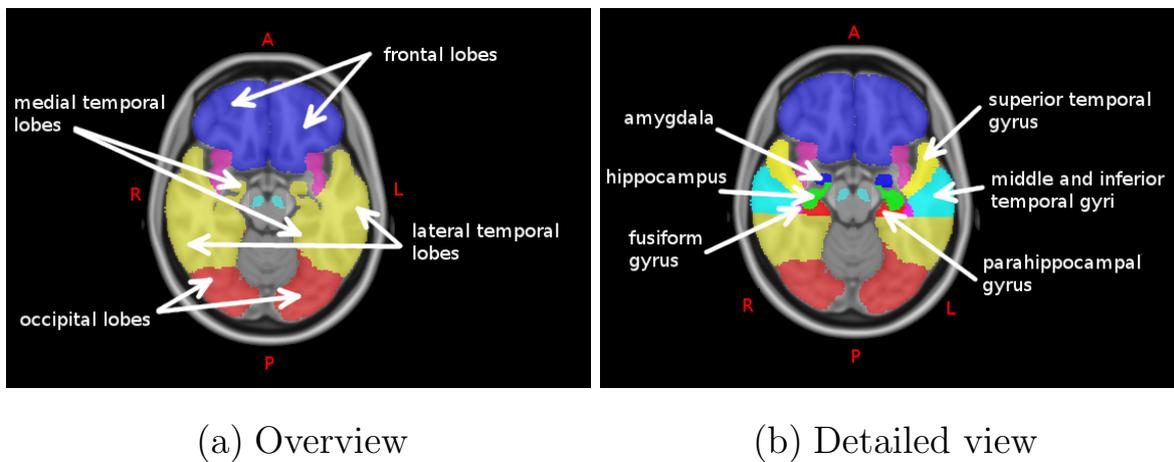


Figure 1.4: Axial views of the brain, showing the sub-structures of the medial temporal lobe. A: anterior, P: posterior, L: left, R: right.

1.3 Neuropathology

Changes occur within the brain even during the healthy ageing process. For example, the cerebral hemispheres lose volume and the ventricles become enlarged. Both such changes may

be attributed to neuronal loss (Graham and Lantos, 1997). These changes become progressively exaggerated during the development of AD, with both cerebral atrophy and neuronal loss often more pronounced in the medial temporal lobes (Dawbarn and Allen, 2007). However, a more characteristic feature of the disease is the abundant presence in the brain of neuropathological structures including extracellular amyloid plaques and intracellular neurofibrillary tangles. For a diagnosis of definite AD to be returned, the presence and distribution of these structures must be directly examined in brain tissue samples.

Amyloid plaques are dense, insoluble deposits of protein and cellular material that form around neurons. Their main protein constituent is beta-amyloid ($A\beta$), which is produced when the larger amyloid precursor protein is successively cleaved by β - and γ -secretase enzymes (Dawbarn and Allen, 2007). The dominant form of $A\beta$ found in amyloid plaques is $A\beta_{1-42}$. This is produced when cleavage by γ -secretase occurs after residue 42 of the $A\beta$ molecule, rather than the usual residue 40 (Selkoe, 2004). Once produced, $A\beta$ proteins accumulate outside the cell, forming small, soluble oligomers. These then aggregate further and combine with other proteins and cellular material, eventually forming insoluble plaques (National Institute on Aging, 2008).

Neurofibrillary tangles are insoluble, twisted fibres found inside neurons, whose main protein constituent is the microtubule-associated protein tau. With the development of AD, the balance between phosphorylation and dephosphorylation of tau is lost, and it becomes hyperphosphorylated (Dawbarn and Allen, 2007). Tau and other microtubule-associated proteins then aggregate inside the cell, forming tangles. These disrupt the stability of the microtubules that are a vital part of the neuronal communication system, ultimately leading to cell death (National Institute on Aging, 2008). The amyloid cascade hypothesis suggests that the formation of $A\beta$ is directly responsible for triggering hyperphosphorylation of tau (Selkoe, 1991).

AD can be divided into two types which share the same pathological features: late-onset AD (LOAD) which tends to manifest after age 60, and the less common familial AD (FAD) which typically has an earlier onset (Dawbarn and Allen, 2007). The work presented in this thesis relates to cases of sporadic LOAD. Age is the most significant risk factor associated with the development of LOAD (Rocca et al., 1991), although genetic, environmental, and other

factors are also relevant. The ApoE gene is the only one so far shown to be associated with the development of LOAD (Dawbarn and Allen, 2007). There are three major alleles of the ApoE gene: $\epsilon 2$, $\epsilon 3$ and $\epsilon 4$. The most common allele is $\epsilon 3$, which is present in 70-80% of most populations (Zannis et al., 1981). The $\epsilon 4$ allele is associated with an increased risk of developing LOAD, while the $\epsilon 2$ allele has a neuroprotective effect (Corder et al., 1993).

1.4 Neuroimaging

Neuroimaging techniques provide a way for clinicians to examine the structural and functional changes in the brain associated with the development of AD *in vivo*. Commonly used modalities include magnetic resonance imaging (MRI), X-ray computed tomography (CT), positron emission tomography (PET), single-photon emission computed tomography (SPECT), and diffusion tensor imaging (DTI). The work presented in this thesis will focus on PET and MRI, both of which are described in the following subsections.

1.4.1 Positron emission tomography

The basic procedure for a PET scan involves injecting the patient with a tracer, labelled with a positron-emitting radionuclide, and then scanning them. A positron emitted inside the body can travel only a short distance through tissue, losing kinetic energy by Coulomb scattering from atomic electrons as it does so, until it is almost at rest. When this low energy positron interacts with an atomic electron, the particles can annihilate to produce two gamma ray photons that are detectable outside the body. To conserve energy and momentum, the photons must be emitted in opposite directions and each with an energy of 511 keV. Since the elements of the PET detector form closed rings around the patient, the two photons are detected simultaneously in opposite detector elements. This process, known as coincidence detection, allows spatial localisation of the tracer in the body and the production of an image showing its distribution.

Radiotracers

Positron-emitting nuclei are unstable, and they stabilise by the decay of a proton into a neutron, positron and electron neutrino. The time taken for half the radioactive nuclei in a sample to decay is known as its half-life. Two nuclei commonly used in PET imaging are ^{11}C and ^{18}F , which have half-lives of 20 and 110 minutes, respectively (Rudin, 2005). Tracer molecules for PET imaging are selected to target a particular physiological process, and then radiolabelled with a suitable positron-emitting nuclide. Isotopic labelling, such as the replacement of ^{12}C by ^{11}C , is preferable because the resulting tracer has identical behaviour to the unlabelled molecule. However, labelling with ^{18}F is attractive because its longer half-life means that synthesis of the tracer does not have to occur on-site. ^{18}F is used as a pseudo-isotopic substitute for hydrogen in a variety of PET tracers because this exchange generally has only a small effect on the behaviour of the molecule *in vivo*.

Image acquisition

A PET scanner consists of a series of coaxial rings around the patient, each containing a number of detector elements. In the most commonly used detection systems, these elements are made up of an array of scintillating crystals which are optically coupled to location-sensitive photomultiplier tubes (Rudin, 2005). When a gamma ray photon interacts with the scintillating crystal, electrons in the lattice are excited from the valence band up into the conduction band. These electrons return to the valence band at impurities in the crystal and, in doing so, dissipate energy in the form of light. This is converted into a weak electronic pulse which is then amplified into a measurable signal in the photomultiplier tubes.

To describe how data are acquired, a single ring of the detector is considered in isolation. Each element in the ring is connected in a coincidence circuit with every other element, and an event is registered if photons are detected simultaneously in two elements. The detection of two photons must occur within a short coincidence window to be considered simultaneous. This is between 10 and 12 nanoseconds in modern clinical PET scanners (Rudin, 2005). Registration

of an event determines a path across the detector, known as the line of response, along which the two photons were emitted, as shown in Figure 1.5 (a). Parallel lines of response are grouped together to form projections for every possible orientation of the ring, as illustrated in Figure 1.5 (b). The number of events recorded along each line of response in a single projection forms one row of a data matrix called a sinogram. The complete sinogram therefore contains information recorded from all projections in a single ring, as shown in Figure 1.5 (c).

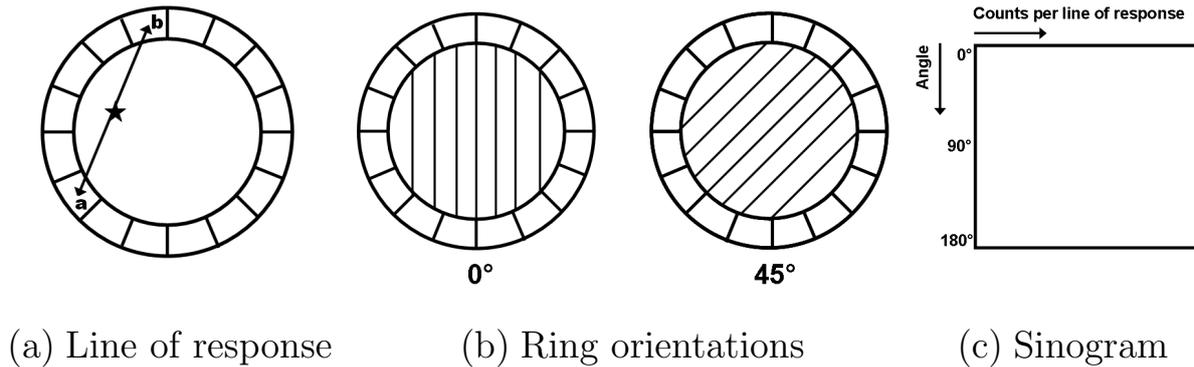


Figure 1.5: Stages of PET image acquisition, showing (a) an annihilation event and the corresponding line of response, (b) the grouping of parallel lines of response to form projections, and (c) the construction of a sinogram.

Photon attenuation in tissue

At energies around 511 keV, the dominant interaction of photons with tissue is by Compton scattering from outer-shell electrons. This results in both a loss of energy and deflection from the original path. Data must be corrected for errors occurring due to this attenuation, as well as other effects, before an image can be reconstructed. The probability that a photon undergoes no interactions as it travels through tissue along a line l is known as its survival probability. The survival probabilities of the pair of photons produced as shown in Figure 1.5 (a) are independent. The combined probability that neither photon interacts may therefore be expressed as

$$P_C = \exp\left(-\int_a^b \mu(x)dx\right)$$

for tissue with the linear attenuation coefficient $\mu(x)$ (Ollinger and Fessler, 1997). The attenuation factor $(1 - P_C)$ is thus independent of the position of the annihilation event along the line l . It can be calculated for every line of response, and the resulting values used to correct the PET image for attenuation. The attenuation map may be obtained either from a transmission scan acquired using an external radiation source prior to injection of the radiotracer, or from the CT image for combined PET/CT scanners. Not all attenuated photons are deflected out of the field of view, and an incorrect line of response may be registered if such scattered photons are detected. However, since energy loss is correlated with the angle of scatter, the registration of scattered photons may be suppressed by only considering those with sufficiently high energy (Rudin, 2005).

Image reconstruction

The aim of PET image reconstruction is to obtain a quantitative map of the spatial distribution of radiotracer in the body. A commonly used method is filtered back projection (FBP), which is described here by consideration of a single slice through an object, illustrated in Figure 1.6 (a). The projection at each angle is first extracted from the sinogram as an intensity profile, shown in Figure 1.6 (b). Since the angle at which each projection was acquired is known, the intensities can be back-projected to reconstruct the image, as shown in Figure 1.6 (c). The resulting star artefact can be suppressed by the application of a ramp filter (Jain, 1989).

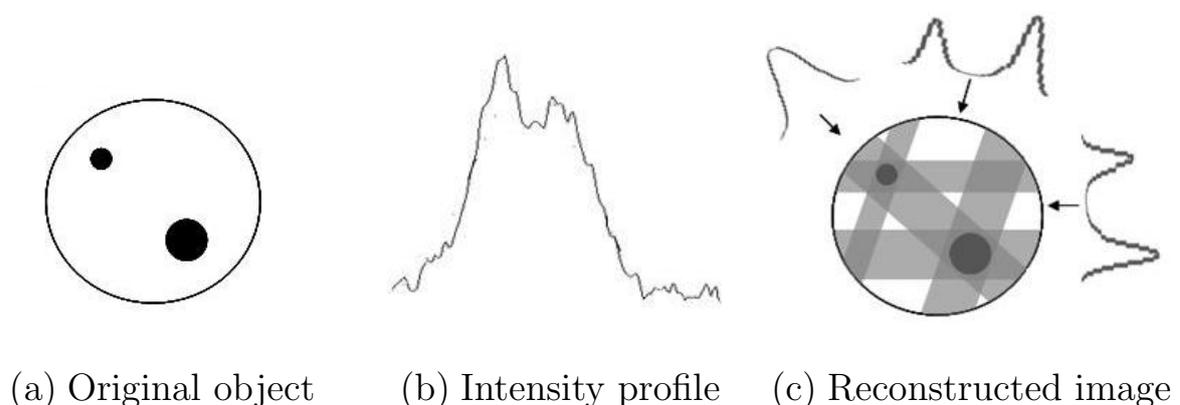


Figure 1.6: Image reconstruction by simple back projection, showing (a) a slice through an object, (b) the intensity profile extracted from a single projection, and (c) the image obtained following reconstruction.

Iterative approaches may alternatively be used, such as the maximum likelihood expectation maximisation (MLEM) algorithm. This aims to find the image most likely to result in the observed projections, given some modelling of the data, noise and detection procedure. The algorithm begins with an estimate of the image, often that obtained using FBP, which it then modifies based on a comparison of the observed projections with those obtained from the image estimate (Qi and Leahy, 2006). In theory, this procedure is repeated until convergence, but in practice it can be very slow and a maximum number of iterations is often specified.

Image analysis

PET images may be acquired in static, dynamic, or gated modes. In static mode, images of several planes through the body provide visual information showing the radiotracer distribution. Although visual analysis can be a useful diagnostic tool, it lacks objectivity. Semi-quantitative objective measures may be obtained from static images. For example, the standardised uptake value (SUV) is the ratio of radioactivity in a region to a subject-specific scale factor which is determined from the injected dose and body weight of the patient (Rudin, 2005). In some cases, there is a region in which the radiotracer accumulates to the same extent in both patients and healthy individuals. The SUV ratio (SUVR) between the region of interest and this reference region provides an alternative measure of regional radiotracer accumulation. In dynamic mode, a time-series of PET images is acquired, from which curves showing the regional tracer kinetics can be extracted. The temporal behaviour of the tracer can then be modelled, and pharmacokinetic parameters derived. In gated mode, the image acquisition is synchronised with a physiological function, such as the cardiac cycle.

1.4.2 Magnetic resonance imaging

MRI exploits the phenomenon of nuclear magnetic resonance (NMR) to produce high quality structural images of the internal organs and other tissues. When undergoing a structural MRI scan, the patient is placed in a powerful static magnetic field, with which the spins of hydrogen

atoms in their body align. This alignment can be perturbed by the application of a radio frequency (RF) electromagnetic pulse, resulting in the resonance emission of a measurable RF signal. The strength of the static magnetic field determines the achievable image quality, and current clinical systems generally employ field strengths of 1.5 T or 3 T. Spatial localisation within the body is achieved by the application of magnetic field gradients, such that the static field varies in strength across the body. The frequency of the resonance signal detected therefore becomes dependent on the location from which it was emitted. Different tissues can be distinguished by the characteristic properties of their emitted RF signals.

Nuclear magnetic resonance

Although nuclei behave according to the laws of quantum mechanics, the principles of NMR can be described with sufficient accuracy using a classical vector model in which nuclear spin is viewed as a physical gyroscopic rotation. In the presence of a static magnetic field B_0 , the spins of hydrogen atoms in the body align either parallel or anti-parallel to the field, as illustrated in Figure 1.7. By convention, the coordinate system is defined such that B_0 is oriented along the z -axis. A net longitudinal magnetisation M_z results from the small excess of spins which align in the lower energy parallel configuration. The spins precess about the static field at a frequency which is dependent on its strength. This is known as the Larmor frequency $\omega_L = \gamma B_0$, where the gyromagnetic ratio γ is characteristic of the nuclei under consideration. The Larmor frequency for hydrogen atoms in the presence of a 1 T static field is 42.6 MHz (Becker, 2000).

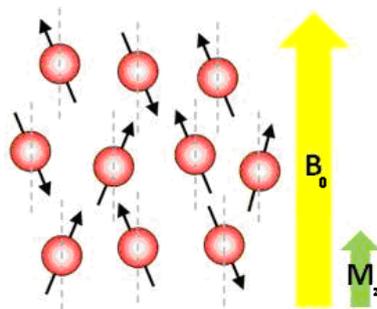


Figure 1.7: Hydrogen atoms in the presence of a static magnetic field B_0 , which induces a net longitudinal magnetisation M_z .

Since the spins do not precess in phase about the z -axis, there is no net magnetisation in the transverse plane. The application of a RF pulse oscillating at the Larmor frequency can establish phase coherence amongst the randomly precessing spins. The application of such a pulse perpendicular to the z -axis results in the rotation of the net magnetisation into the transverse plane. When the RF signal is then switched off, the spins precess in phase about the static magnetic field, thus inducing a measurable voltage in a receiver coil. The amplitude of this signal is maximal immediately following the RF pulse, but then decays with time as the precession loses phase coherence, and the system returns to equilibrium.

Fourier transforms

In the presence of a static magnetic field B_0 oriented along the z -axis, the spins of hydrogen atoms in the body precess in the xy -plane, as illustrated in Figure 1.8 (a). This precession can be described by oscillating components in both the x - and y -directions, as shown in Figure 1.8 (b). A mathematical technique known as a Fourier transform can be used to convert these temporal signals into a frequency distribution. The temporal signals illustrated in Figure 1.8 (b) correspond to a single peak at the Larmor frequency, as illustrated in Figure 1.8 (c).

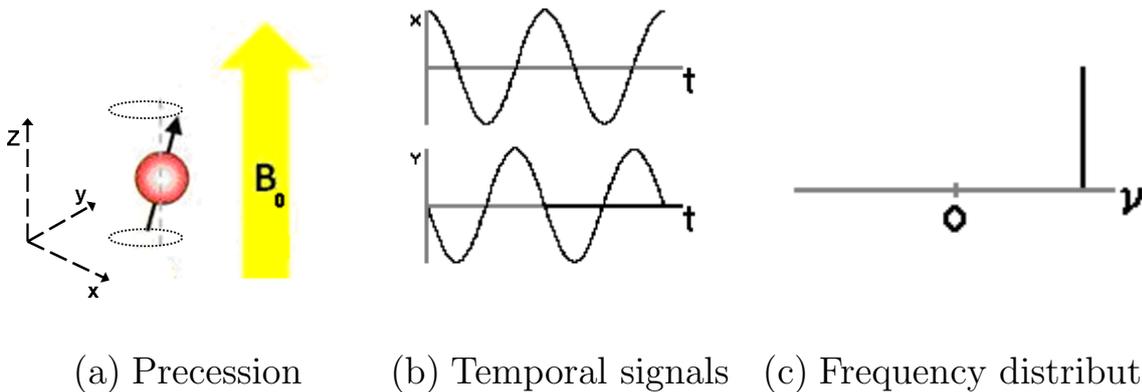


Figure 1.8: The application of Fourier transforms to MRI illustrated by consideration of (a) the precession of spins in the presence of a static magnetic field. A Fourier transform converts (b) the associated temporal signals into (c) the corresponding frequency distribution.

The RF signal detected following resonance emission from nuclei within the body is a temporal signal consisting of many frequency components. A Fourier transform can convert this temporal

signal $f(t)$ into a multi-spectral frequency distribution $f(\omega)$, according to the expression

$$f(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt = \int_{-\infty}^{\infty} f(t)[\cos(\omega t) - i \sin(\omega t)] dt.$$

More detailed information about Fourier transforms and their properties may be found in [Jennison \(1961\)](#).

Spatial localisation

Spatial localisation is achieved by using magnetic field gradients to modify the static field so that it varies in strength across the body. Since the Larmor frequency is proportional to the applied magnetic field, the location of the source signal can then be inferred from the frequency of the resonance signal detected. Figure 1.9 depicts an example MRI pulse sequence illustrating the additional gradients required for spatial localisation in 3-D ([Hornak, 2010](#)).

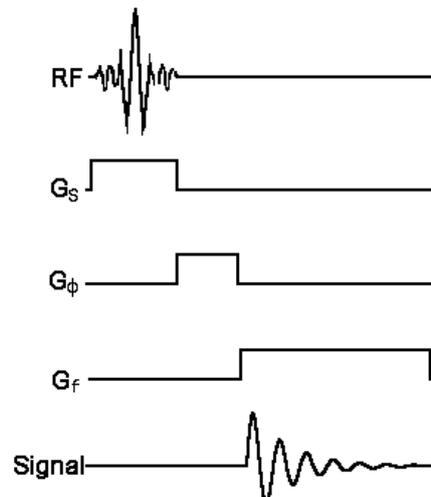


Figure 1.9: An example MRI pulse sequence, showing the RF pulse, slice-selection gradient G_S , phase encode gradient G_ϕ , frequency encode gradient G_f , and the detected RF signal.

The application of a slice-selection gradient G_S along the z -axis results in a Larmor frequency which varies linearly with z . A RF pulse with a narrow band of frequencies therefore excites resonance within a single transverse section of the body. To achieve spatial localisation within this transverse section, additional gradients are applied along the x - and y -axes. A phase

encode gradient G_ϕ is first applied along one axis, such that the precession frequencies of the nuclei within the transverse section become dependent on their position. When this gradient is switched off, the precession frequencies of the nuclei are once again identical; however, they precess out of phase. A frequency encode gradient G_f is applied along the remaining axis during the signal detection. The locations of nuclei within a transverse section of the body can then be unequivocally identified from the frequency distribution of the detected RF signal.

Tissue contrast

The decay of the RF signal as the nuclear spins return to equilibrium is associated with time constants describing its longitudinal and transverse components. Recovery of the longitudinal magnetisation as the spins realign with the static magnetic field is known as spin-lattice, or T_1 , relaxation. Decay of the transverse magnetisation as the spins dephase is known as spin-spin, or T_2 , relaxation. Magnetic field inhomogeneities cause the signal to decay faster in the transverse plane than can be explained by T_2 relaxation alone. This effect is known as T_2^* relaxation. The environment of the nuclei under consideration influences these time constants, and therefore the decay properties of the RF signal. MRI contrast is dependent on the differing T_1 and T_2 relaxation properties of various biological tissues.

1.5 Biomarkers for Alzheimer's disease

Since the publication of the NINCDS-ADRDA Alzheimer's Criteria in 1984, significant progress has been made in identifying the structural and molecular changes in the brain that are associated with AD. Much of the recent research has been based on data from the Alzheimer's Disease Neuroimaging Initiative (ADNI; <http://adni.loni.ucla.edu>), which aims to compare neuroimaging, biological, and clinical assessment of the cognitive and behavioural changes associated with normal ageing, MCI and AD. Participants undergo regular cognitive and functional assessments, and some also opted to undergo lumbar punctures for the collection of CSF biomarkers such as $A\beta$ and tau. All ADNI participants had structural MRI scans, and approx-

imately 50% also underwent PET imaging with the tracer [^{18}F]-fluorodeoxyglucose (FDG). Some participants additionally underwent PET imaging with the tracer [^{11}C]-Pittsburgh compound B (PiB). FDG-PET images depict brain function in terms of the rate of cerebral glucose metabolism, and PiB-PET images show the distribution of amyloid deposition in the brain.

Figure 1.10 shows a hypothetical model of the temporal profiles of various biomarkers throughout the development of AD. Amyloid accumulation occurs earliest in the disease process, preceding both cognitive and functional decline by years, and changing only gradually after symptoms develop. Compared to measures of amyloid deposition, CSF tau levels, MRI volumes, and FDG-PET intensities are more dynamic biomarkers of AD progression. At present, a clinical diagnosis of AD is made based on assessments of cognition and behaviour, which start to decline fairly late in the disease process. Other biomarkers may therefore be better suited for the early detection and prediction of AD, and for monitoring progression. These are briefly reviewed in the following subsections.

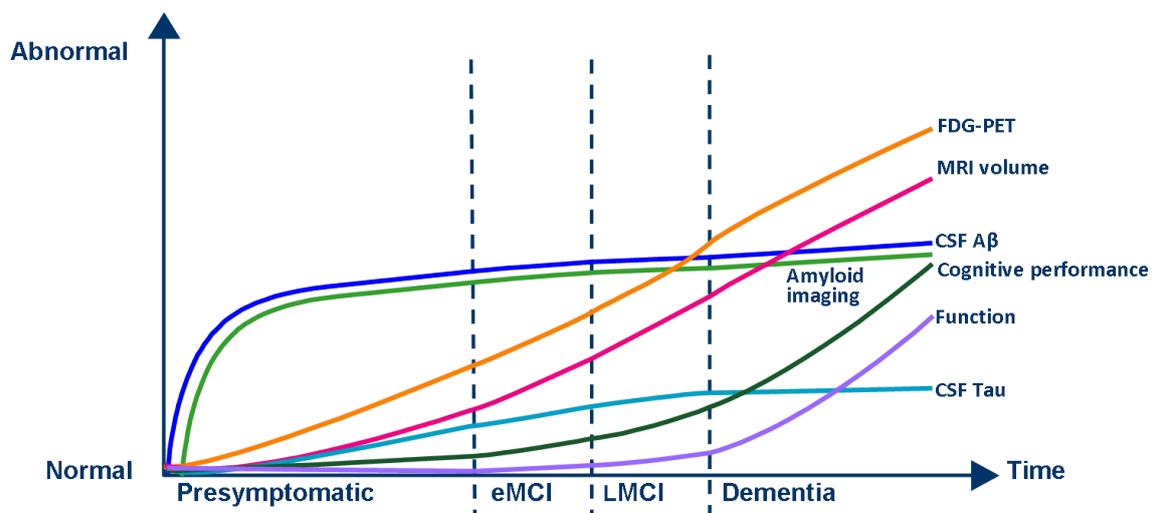


Figure 1.10: Hypothetical temporal model of biomarker dynamics during AD progression. Biomarker measures vary from normal to maximally abnormal as a function of the disease stage. eMCI: early MCI, LMCI: late MCI. Adapted from (Aisen et al., 2010).

1.5.1 Cerebrospinal fluid

Recent consensus reports have identified CSF levels of A β and tau as among the most promising potential AD biomarkers (Frank et al., 2003; The Ronald and Nancy Reagan Research Institute

of the Alzheimer's Association and the National Institute on Aging Working Group, 1998). CSF levels of $A\beta$ are approximately 100 times greater than those found in blood plasma, and this biomarker is best measured in the CSF (Scheuner et al., 1996). The same is true of tau, which is thought to be released from damaged neurons as they undergo neurofibrillary degeneration (Kahle et al., 2000). CSF is extracted by lumbar puncture, in which a needle is inserted between the lumbar vertebrae into the subarachnoid space of the spinal canal

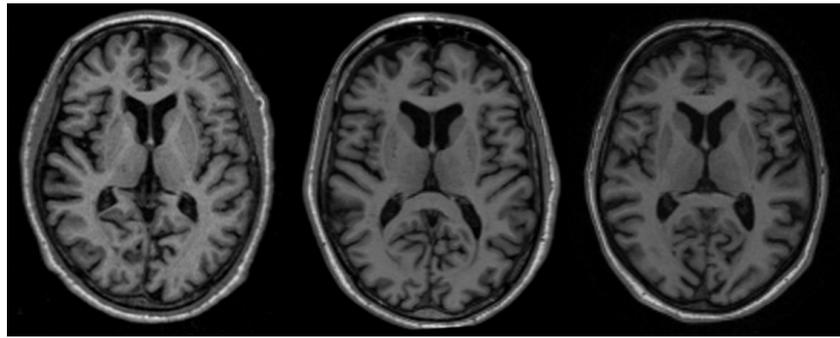
Various studies have shown AD patients to have reduced CSF $A\beta$ and elevated CSF tau compared with cognitively normal individuals (Ishiguro et al., 1999; Motter et al., 1995; Vandermeeren et al., 1993). When considered in combination, these two biomarkers can effectively distinguish AD patients from healthy individuals (Sunderland et al., 2003), as well as from patients with other types of dementia (Clark et al., 2003). AD patients with at least one ApoE $\epsilon 4$ allele have lower CSF $A\beta$ and higher CSF tau than those without (Tapiola et al., 2000). This finding aligns with the observation that more extensive AD pathology is generally found in AD carriers of the ApoE $\epsilon 4$ allele than in non-carriers (Roses and Saunders, 1997).

MCI patients tend to have CSF $A\beta$ and tau levels that lie between those expected of AD patients and healthy individuals. Preliminary data suggest that MCI patients with AD-like biomarker levels have a greater likelihood of converting to AD than those with biomarker levels more typical of cognitively normal individuals (Hansson et al., 2006). The first study of baseline CSF biomarker data from ADNI largely confirmed previous findings (Shaw et al., 2009). CSF $A\beta$ was found to be the most sensitive single CSF biomarker, and the overall best group discrimination was achieved by combining CSF $A\beta$ and tau, along with the number of ApoE $\epsilon 4$ alleles. The majority of MCI patients who converted to AD over the course of one year had baseline CSF $A\beta$ and tau levels that were more typical of AD patients than of healthy controls.

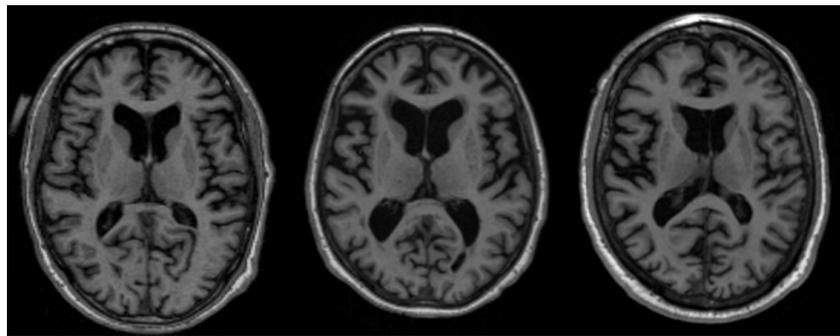
1.5.2 Magnetic resonance imaging

The structural changes in the brain associated with AD can be non-invasively assessed using MRI. As shown in Figure 1.11, AD patients typically have evidence of cortical atrophy, and

enlarged ventricles in comparison with healthy individuals.



(a) Healthy individuals



(b) AD patients

Figure 1.11: Transverse sections from MR images of (a) healthy individuals, and (b) AD patients. These images demonstrate that AD patients typically show evidence of cortical atrophy, and enlarged ventricles in comparison with healthy individuals.

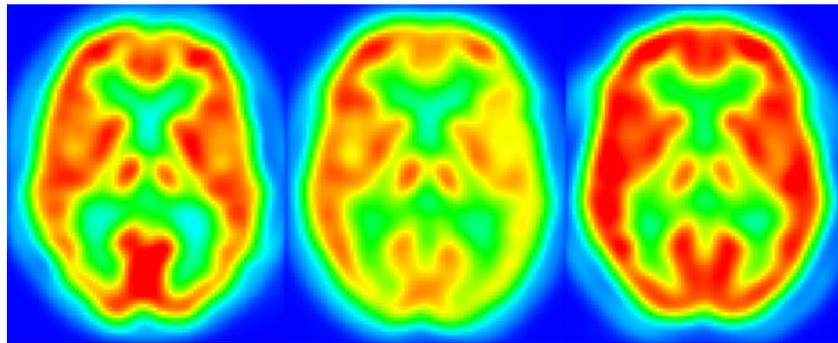
Temporal lobe atrophy is closely associated with AD, and histological studies show that the hippocampus, amygdala and entorhinal cortex are particularly vulnerable to AD pathology (Braak and Braak, 1998). Correlation has been found between the rate of temporal lobe atrophy and both current cognitive performance and future decline, even among healthy individuals (Hua et al., 2008). Increased rates of hippocampal atrophy compared with cognitively normal individuals have been measured using MRI in both AD and MCI patients (Schuff et al., 2009; van de Pol et al., 2007). Longitudinal studies have additionally shown that the rate of hippocampal atrophy accelerates over time in both AD and MCI patients (Jack Jr. et al., 2008c; Ridha et al., 2006). However, hippocampal atrophy alone is not sufficient to predict conversion from MCI to AD, and other structures may prove more sensitive (Dickerson et al., 2001).

A recent analysis of the ADNI MRI data found that an increased rate of hippocampal volume

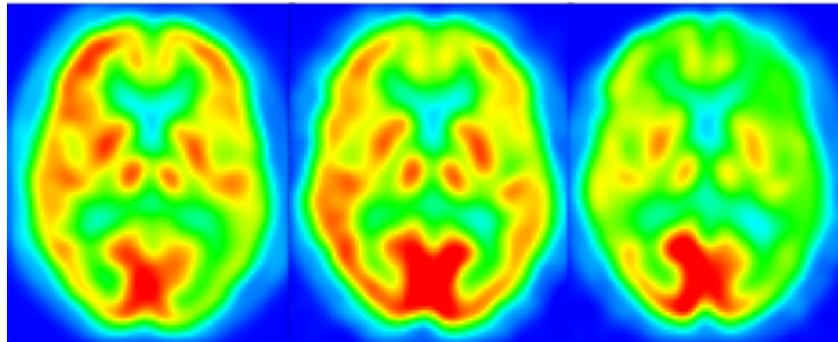
loss was associated with presence of the ApoE $\epsilon 4$ allele in AD patients, and with reduced levels of CSF A β in MCI patients (Schuff et al., 2009). Another analysis showed that the rate of temporal lobe atrophy in AD is correlated with reduced CSF A β and elevated CSF tau, and that it is significantly faster in MCI subjects that later convert to AD than in non-converters (Leow et al., 2009).

1.5.3 Fluorodeoxyglucose positron emission tomography

FDG is a ^{18}F labelled glucose analogue, whose distribution in the brain gives an indication of the cerebral metabolic rate of glucose (CMRgl). As shown in Figure 1.12, AD patients typically have reduced glucose metabolism in temporo-parietal regions of the brain in comparison with healthy individuals.



(a) Healthy individuals



(b) AD patients

Figure 1.12: Transverse sections from FDG-PET images of (a) healthy individuals, and (b) AD patients. These images demonstrate that AD patients typically have reduced glucose metabolism in temporo-parietal regions of the brain in comparison with healthy individuals.

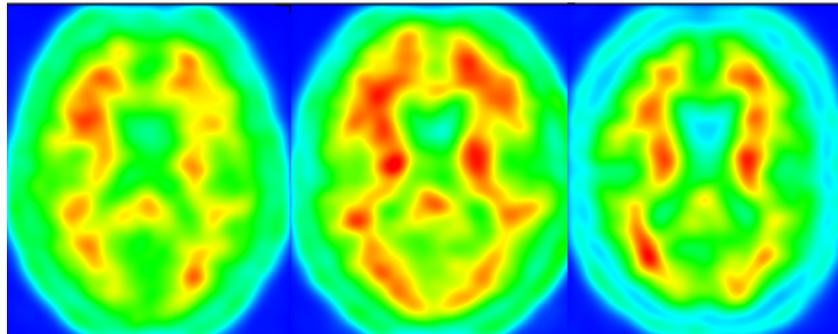
Numerous FDG-PET studies have shown that both MCI and AD are associated with significant reductions in the CMRgl in brain regions preferentially affected by the disease (de Leon et al., 2001, 1983; Herholz et al., 2002; Langbaum et al., 2009; Mosconi et al., 2008, 2007; Mosconi, 2005; Nestor et al., 2003). AD patients display reductions of greater magnitude and spatial extent than MCI patients. Reductions in the CMRgl in AD patients can predict both their cognitive decline and histopathological diagnosis (Hoffman et al., 2000; Minoshima et al., 2001; Silverman et al., 2001), while those in MCI patients can predict their conversion to AD (Anchisi et al., 2005; Mosconi et al., 2004). Longitudinal studies have shown these changes to be progressive (Alexander et al., 2002; Mosconi et al., 2005). Based on comparisons of AD and MCI patients, it has been suggested that posterior regions are preferentially affected in the earlier stages of AD, with anterior regions such as the frontal cortex becoming involved only in the later stages of the disease (Alexander et al., 2002; Langbaum et al., 2009).

Cognitively normal individuals with one or two ApoE ϵ 4 alleles already have reduced CMRgl in some of the regions affected by AD (Langbaum et al., 2009; Reiman et al., 2005). This finding suggests that FDG-PET can provide an early indicator for the disease. A single study of a small group of MCI patients (Drzezga et al., 2005) has demonstrated complete separation of those that rapidly converted to AD and those remaining stable, using a combination of reduced CMRgl in AD-typical regions and ApoE ϵ 4 status.

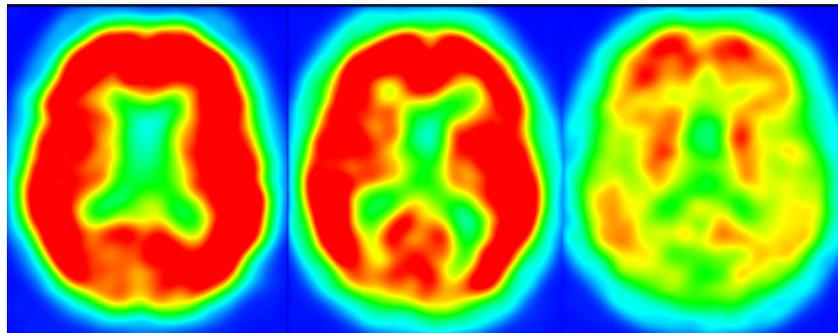
1.5.4 Pittsburgh compound B positron emission tomography

PiB is a ^{11}C labelled thioflavin-T derivative that binds to amyloid plaques *in vivo*. It can thus be used to assess one of the characteristic neuropathological features of AD. As shown in Figure 1.13, AD patients typically have increased PiB retention in areas known to accumulate significant amyloid deposits in comparison with healthy individuals. A number of PiB-PET studies have reported cortical PiB retention in AD patients, and mostly non-specific retention in the white matter in healthy individuals (Forsberg et al., 2008; Jack Jr. et al., 2009; Klunk et al., 2004; Villemagne et al., 2008). Cortical PiB retention is also observed in MCI patients, but to a lesser extent than in AD. An inverse correlation has been found between cortical PiB

retention and levels of CSF $A\beta$ (Fagan et al., 2006). Patients are often classified as PiB positive or negative, where a global cortical to cerebellar ratio is defined to separate the two groups. Independent studies have consistently found that approximately 30% of cognitively normal elderly individuals would be classified as PiB positive according to such criteria (Jack Jr. et al., 2008b; Mintun et al., 2006). This suggests that PiB alone is not a sufficient marker for AD, although it may indicate individuals who will subsequently develop the disease. Longitudinal follow-up of cognitively normal PiB positive individuals will be required to verify this suggestion.



(a) Healthy individuals



(b) AD patients

Figure 1.13: Transverse sections from PiB-PET images of (a) healthy individuals, and (b) AD patients. These images demonstrate that AD patients typically show cortical PiB retention, while healthy individuals typically show non-specific retention in the white matter.

Significant amyloid plaque deposition occurs before the onset of clinical symptoms (Mintun et al., 2006), continuing at a slower rate as AD progresses. Progression may therefore be better assessed by considering measures of neurodegeneration. A study of ADNI MRI and PiB-PET found the rate of ventricular expansion greater in MCI patients that were PiB positive at baseline than in those that were PiB negative (Jack Jr. et al., 2009). This supports other studies suggesting that PiB negative MCI patients may not have early AD (Archer et al., 2006;

Forsberg et al., 2008). Additional follow-up data will again be required for verification.

1.6 Research contributions and thesis outline

The research presented in this thesis contributes to the growing body of literature surrounding the image-based classification of MCI and AD. In particular, a framework for multi-modality classification, based on the combination of similarity measures derived from random forest classifiers, is presented. In addition, early signs of neurodegeneration are identified in cognitively normal individuals at high risk of developing AD, based on multi-region analysis of MR images from two independent cohorts.

Relevant concepts from the fields of image analysis and machine learning are first described in Chapters 2 and 3, respectively. Chapter 2 describes image analysis techniques including registration, anatomical segmentation, and statistical parametric mapping. Chapter 3 then provides an overview of machine learning concepts relevant to image-based classification, including classification algorithms, and methods with which to assess their performance. A review of the current state-of-the-art for image-based classification of AD is additionally presented.

In terms of the classification of AD and MCI, there are many more studies based on structural MR imaging data than on FDG-PET imaging data. This is because anatomical imaging with either MRI or CT is routinely used in clinical practice for dementia patients. Investigations of the potential utility of multi-region FDG-PET features for image-based classification of AD and MCI are described in Chapters 4 and 5. In particular, attempts are made to distinguish between MCI patients who subsequently progress to AD and those who remain stable. Chapter 4 demonstrates that regional information extracted from FDG-PET images acquired at a single timepoint can be used to achieve classification results in line with those obtained using data from MRI, or biomarkers obtained invasively from the CSF. Chapter 5 then demonstrates the additional benefit of incorporating longitudinal FDG-PET information for classification. By combining cross-sectional and longitudinal multi-region FDG-PET features, classification results in line with the current state-of-the-art are achieved. The findings described in these

chapters support the use of FDG-PET for the early diagnosis of AD and for monitoring its progression.

Changes in multiple neuroimaging and biological measures may provide complementary information for the diagnosis and prognosis of AD. Chapter 6 presents a multi-modality classification framework in which manifolds are constructed based on pairwise similarity measures derived from random forest classifiers. Similarities from multiple modalities are combined to generate an embedding that simultaneously encodes information about all the available features. Multi-modality classification is then performed using coordinates from this joint embedding. Random forests provide consistent pairwise similarity measures for multiple modalities, thus facilitating the combination of different types of feature data. Classification results based on the combination of regional MRI volumes, voxel-based FDG-PET signal intensities, CSF biomarker measures, and ApoE allele status are comparable with those obtained in other studies using multi-kernel learning. Since random forest classifiers extend naturally to multi-class problems, the framework described here could be used for other applications in the future, such as the differential diagnosis of AD.

Novel findings of early signs of neurodegeneration in cognitively normal individuals at high risk of developing AD are presented in Chapter 7. Multi-region analysis of MR images acquired at a single timepoint is used to show volumetric differences in cognitively normal individuals differing in amyloid-based risk status for the development of AD. Reduced volumes in temporo-parietal and orbito-frontal regions in high-risk individuals from two independent cohorts could be indicative of very early changes associated with AD. These findings suggest that volumetric MRI can reveal structural brain changes that precede the onset of clinical symptoms. It may therefore be useful in identifying early signs of neurodegeneration in healthy elderly individuals, potentially providing a useful early screening tool, or outcome measure for clinical trials.

Chapter 2

Background: PET and MR image analysis

2.1 Introduction

This chapter describes state-of-the-art techniques relevant to the PET and MR image analysis presented in this work. Image registration is first described in Section 2.2. This allows the alignment of different images so that they share a common coordinate system. Anatomical segmentation techniques are then reviewed in Section 2.3. Much of the work presented in this thesis involves the use of multi-region imaging features obtained by using segmentation to label anatomically defined structures in the brain. The initial focus of this research was FDG-PET image analysis, and an overview of statistical parametric mapping is given in Section 2.4. This provides a voxel-based analysis method for studying group differences amongst PET images.

2.2 Image registration

The goal of image registration is to estimate a spatial correspondence between two images. Approaches can be broadly divided into those based on image intensity values, and those

which instead rely on image features such as lines or contours. The focus for this work is on intensity-based approaches. These typically comprise several related components: a transformation model, optimisation method, similarity metric, regularisation method, and interpolation method. A transformation model defines the way in which one image (the source) should be deformed into the coordinate system of another (the target). Having selected a transformation model, the spatial correspondence between the two images is estimated by applying an optimisation method to find the transformation which maximises the image similarity. To ensure that the transformation is plausible, regularisation may be incorporated into the registration process. Finally, intensities may need to be interpolated to compensate for any mismatch between the deformed voxel grid of the source image and the target grid.

Several applications of image registration will be described in this thesis: alignment of images of a subject acquired at a single timepoint using different modalities (MRI and PET), alignment of images of a subject acquired serially using a single modality, and alignment to a standard template space of images of a group of subjects acquired using a single modality. Registration also forms part of an image segmentation procedure in which the labels from a set of manually segmented images are propagated to the target.

The components of the voxel-based registration method used throughout this research are described in detail in the following subsections. A more comprehensive review of image registration techniques may be found in [Hajnal et al. \(2001\)](#).

2.2.1 Transformation model

A transformation \mathbf{T} defines a parametric representation mapping a voxel in the target image to a location in the source image, $\mathbf{T} : (x, y, z) \rightarrow (x', y', z')$. Transformations may be broadly divided into linear and nonlinear models. Linear models include both rigid transformations, which preserve distances between points, and affine transformations, which preserve colinearity of points. Linear transformations are global in nature, and cannot model local geometric differences between images. Nonlinear (or nonrigid) transformations, however, can represent varying

local deformations, thus allowing the source image to be locally warped into the coordinate system of the target.

The choice of transformation model is dependent on the application of interest. For example, a rigid transformation may be sufficient for the registration of serially acquired brain MR images of a healthy adult, since there should be very little change in the shape of the cranium. An affine transformation may be more appropriate for the intra-subject registration of brain MR and PET images acquired at a single timepoint, where some global scaling may be required. A nonrigid transformation may be applied following a global transformation to reduce any residual differences remaining between images. For example, local differences are likely to remain following global registration of serially acquired brain MR images of an AD patient.

Rigid transformations

A rigid transformation can be represented by the application of translations and rotations, as illustrated in Figure 2.1.

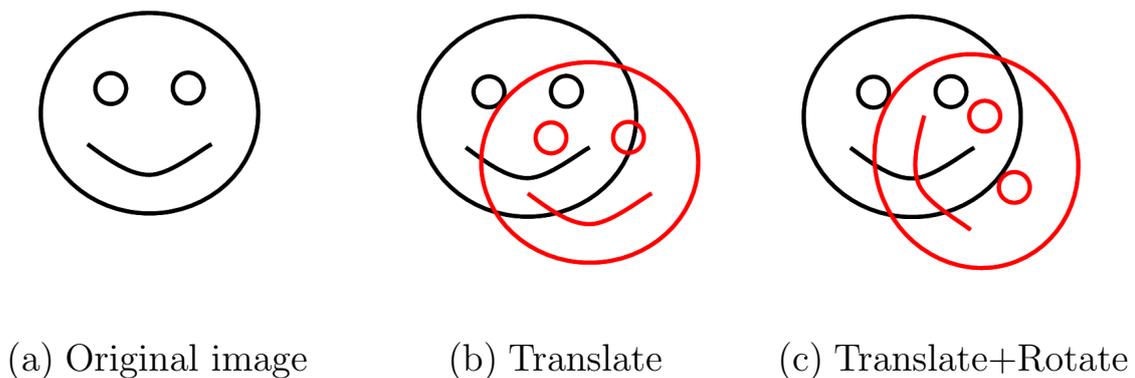


Figure 2.1: Illustration of possible rigid transformations applied to (a) the original image, showing the effect of (b) translation, and (c) both translation and rotation.

In 2-D, a translation in the xy -plane can be represented by a vector \mathbf{t} , and rotations may be made about an axis perpendicular to the plane. An anti-clockwise rotation about the z -axis by an angle θ can be expressed as

$$\mathbf{R}_z(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

The effect of applying a rigid transformation comprising a rotation matrix \mathbf{R} and translation vector \mathbf{t} to a point \mathbf{x} can be written as $\mathbf{T}_{\text{rigid}}(\mathbf{x}) = \mathbf{R}\mathbf{x} + \mathbf{t}$. This may alternatively be represented by the following single transformation matrix using homogenous coordinates:

$$\mathbf{T}_{\text{rigid}}(\mathbf{x}) = \begin{pmatrix} \mathbf{R}\mathbf{x} + \mathbf{t} \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix},$$

where $\mathbf{T}_{\text{rigid}}$ can be decomposed into the following block form rotation and translation matrices:

$$\mathbf{T}_{\text{rigid}} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{t} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R} & 0 \\ 0 & 1 \end{pmatrix}.$$

In 3-D, a general rotation can be decomposed into rotations about each of the coordinate axes.

In a right-handed frame, these rotations may be expressed as

$$\mathbf{R}_x(\theta_1) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta_1 & -\sin \theta_1 & 0 \\ 0 & \sin \theta_1 & \cos \theta_1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \mathbf{R}_y(\theta_2) = \begin{pmatrix} \cos \theta_2 & 0 & \sin \theta_2 & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \theta_2 & 0 & \cos \theta_2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{R}_z(\theta_3) = \begin{pmatrix} \cos \theta_3 & -\sin \theta_3 & 0 & 0 \\ \sin \theta_3 & \cos \theta_3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

A general rotation comprising sequential rotations about the x -, y - and z -axes can be expressed as $\mathbf{R} = \mathbf{R}_z(\theta_3)\mathbf{R}_y(\theta_2)\mathbf{R}_x(\theta_1)$. The single matrix representation for the rigid transformation

$\mathbf{T}_{\text{rigid}}$ is therefore the same as for the 2-D case. A general 3-D rigid transformation can therefore be represented using six parameters, three describing translation, and three rotation.

Affine transformations

An affine transformation can be represented by the application of translations and rotations, as well as scales and shears, which are illustrated in Figure 2.2.

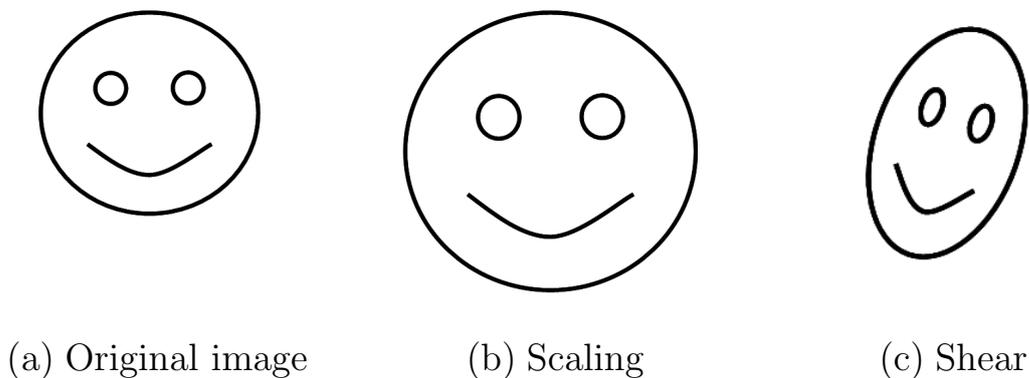


Figure 2.2: Illustration of possible affine transformations applied to (a) the original image, showing the effect of (b) scaling, and (c) shear.

In 3-D, scale factors s_x , s_y and s_z can be applied independently along each of the coordinate axes, such that a general scaling transformation may be expressed as

$$\mathbf{T}_{\text{scale}} = \begin{pmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

In 2-D, shear along the x -direction describes the translation of a point $\mathbf{x} = (x_x, x_y)$ parallel to the x -axis by an amount $sh_x x_y$, where sh_x is a scalar shear coefficient. Shear in the y -direction can be similarly described by a scalar shear coefficient sh_y . The general shear matrix can then be expressed as

$$\mathbf{T}_{\text{shear}} = \begin{pmatrix} 1 & sh_y & 0 \\ sh_x & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In 3-D, shears can be characterised as either beam shears or slice shears, although these two representations are equivalent. A beam shear is defined as the translation of a point \mathbf{x} parallel to one axis by an amount equal to a linear combination of the other two coordinate values. A slice shear involves translation of the point \mathbf{x} along a pair of axes by an amount proportional to the coordinate value of the third. For example, shear along the x - and y -axes due to z is described by scalar shear coefficients sh_{zx} and sh_{zy} respectively. A general shear can therefore be expressed as

$$\mathbf{T}_{\text{shear}} = \begin{pmatrix} 1 & sh_{xy} & sh_{xz} & 0 \\ sh_{yx} & 1 & sh_{yz} & 0 \\ sh_{zx} & sh_{zy} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

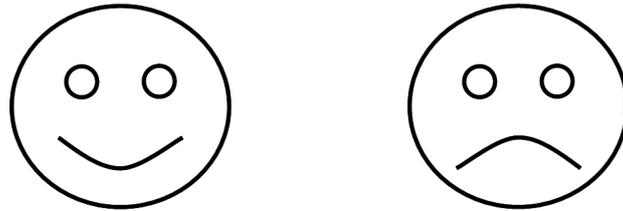
A general affine transformation may be expressed as $\mathbf{T}_{\text{affine}} = \mathbf{T}_{\text{shear}}\mathbf{T}_{\text{scale}}\mathbf{T}_{\text{rigid}}$. In 3-D, translation, rotation and scale can each be represented using three parameters, and shear using six. It therefore appears that 15 parameters are required to specify an affine transformation. However, the parameters are not independent, and a general 3-D affine transformation may be expressed as

$$\mathbf{T}_{\text{affine}} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Only 12 independent parameters are required to represent a 3-D affine transformation.

Nonrigid transformations

A nonrigid transformation represents local deformations which can vary across the image, as illustrated in Figure 2.3. There are various possible ways to characterise nonrigid transformations because they require many more parameters than global transformations and therefore cannot be simply represented using a single matrix.



(a) Original image (b) Deformed image

Figure 2.3: Illustration of a nonrigid transformation applied to (a) the original image, showing the effect of (b) a locally varying deformation.

A brief description of the relevant mathematical terminology is now presented. A function f on a domain Ω is described as continuous if an infinitesimal change in the input results in an infinitesimal change in the output. The class of all such continuous functions, whose n th derivatives are also continuous, is denoted $C^n(\Omega)$. If all derivatives of a continuous function f are also continuous, the function is described as smooth, and it belongs to the class $C^\infty(\Omega)$. In practice, when modelling transformations, functions belonging to classes $C^2(\Omega)$ and $C^1(\Omega)$ may be considered sufficiently smooth. A function f which maps points from a set X to a set Y is described as a homeomorphism if it is a bijection, and both f and its inverse f^{-1} are continuous. A bijection describes an exact pairing of the sets X and Y , such that each element in X is paired with exactly one element in Y , and *vice-versa*. A homeomorphic transformation preserves topology, and should therefore be used if the underlying topology between two images is assumed to be identical. Enforcing the additional restriction that both f and its inverse f^{-1} must belong to the class $C^n(\Omega)$ defines f as a C^n -diffeomorphism. The term diffeomorphism is typically used to refer to the case of C^∞ -diffeomorphism. If the anatomical structures represented within a pair of images are assumed to be smooth, the transformation between them

must be a diffeomorphism of the appropriate order. Diffeomorphic transformations are often used as a theoretical basis for the nonrigid registration of medical images.

A nonrigid transformation can be represented by a smooth displacement field, which requires the smooth assignment of vectors to every location in the image. In 3-D, this requires that a displacement vector is specified for every voxel, and the number of parameters is therefore three times the number of voxels. It is possible to reduce the number of parameters required by either using a model, or exploiting a property of the transformation. For example, a smooth transformation may be globally defined based on displacement vectors assigned to each of a set of control points defined within the image. For the free-form deformation (FFD) model used in this work (Rueckert et al., 1999), the control points are arranged on a regular $n_x \times n_y \times n_z$ axis-aligned lattice with spacings $\delta_x, \delta_y, \delta_z$ along each coordinate axis. A FFD can then be parametrised by a set of vectors $\{\Phi\}$, each of which is associated with one of the control points. Displacement vector $\phi_{i,j,k}$, for example, represents the control point located at position $\mathbf{x} = (i, j, k)$. The nonrigid transformation is thus parametrised by the displacement vectors at the locations of the control points, but must also be defined at general locations within the image. This is achieved by convolving $\{\Phi\}$ with a suitable basis function, resulting in a smoothly varying displacement field across the entire image. In this work, the control points are convolved using 1-D cubic B-splines, which are expressed as

$$\begin{aligned} B_0(u) &= \frac{(1-u)^3}{6} & B_2(u) &= \frac{(-3u^3 + 3u^2 + 3u + 1)}{6} \\ B_1(u) &= \frac{(3u^3 - 6u^2 + 4)}{6} & B_3(u) &= \frac{u^3}{6}. \end{aligned}$$

The local displacement at the general location $\mathbf{x} = (x, y, z)$ can then be written as the 3-D tensor product over the control point vectors, as

$$\mathbf{T}_{\text{local}}(x, y, z) = \sum_{l=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 B_l(u) B_m(v) B_n(w) \phi_{i+l, j+m, k+n},$$

where $i = \lfloor x/\delta x \rfloor - 1$, $j = \lfloor y/\delta y \rfloor - 1$, $k = \lfloor z/\delta z \rfloor - 1$, $u = x/\delta x - \lfloor x/\delta x \rfloor$, $v = y/\delta y - \lfloor y/\delta y \rfloor$, $w = z/\delta z - \lfloor z/\delta z \rfloor$, and $\lfloor x \rfloor = \text{floor}(x)$, which gives the largest integer not greater than x .

B-splines are computationally efficient because the displacement at a particular control point affects the transformation only in its local neighbourhood. Similarly, its displacement depends only on control points within the neighbourhood.

When performing nonrigid registration using FFDs, it is possible to employ a hierarchical, coarse-to-fine strategy (Schnabel et al., 2001) to reduce the likelihood of convergence to a local optimum. Using this approach, the FFD parameters are first optimised based on a relatively sparse control point lattice with a large spacing. This results in a transformation that captures large-scale local differences between the two images being registered. The control point lattice is then sub-divided to generate a lattice with half the spacing of the original (Forsey and Bartels, 1988). The FFD parameters are re-optimised based on this new, denser lattice of control points. This process of lattice sub-division and parameter optimisation may be repeated as required, with smaller-scale local differences between the images being captured at each iteration. It is important to ensure that the scale of the image features is appropriate for the selected control point spacing. The images should be blurred and resampled at each step so that information relating to structures smaller than a certain size is neglected. Images may be blurred by convolution with a Gaussian kernel of width σ . For capturing large-scale differences, a relatively wide Gaussian kernel is appropriate. Successively smaller values of σ may then be selected as smaller-scale differences are captured by the FFD.

Combining global and local transformations

In the work presented in this thesis, nonrigid image registration is performed as a multi-stage process. Global transformation parameters are estimated, and used as the starting point for the nonrigid registration step. The global transformation itself is performed in two steps, with rigid transformation parameters estimated first, and used as the starting point for an affine registration. Rigid transformation parameters make up a sub-set of the affine parameters, and their combination is therefore straightforward. The affine transformation is simply initialised with the translation and rotation parameters from the rigid registration, and the scales and shears set to one and zero respectively. The global transformation is thus obtained following

the affine registration step, and the subsequent nonrigid registration aims to estimate the local residual displacements required to accurately align the images. This means that the global and local components of the final transformation are combined by addition, such that $\mathbf{T}(\mathbf{x}) = \mathbf{T}_{\text{global}}(\mathbf{x}) + \mathbf{T}_{\text{local}}(\mathbf{x})$.

2.2.2 Optimisation method

The goal of an optimisation method, as applied to the registration of medical images, is to select transformation parameters which maximise the similarity between the two images being registered. In the work presented in this thesis, global transformations are optimised using downhill descent, and nonrigid transformations using steepest gradient descent. Both methods consider the optimisation as a minimisation problem, and therefore aim to minimise the difference between the two images. This is equivalent to maximising their similarity.

In downhill descent, the transformation parameters are initialised, and the similarity between the two images is determined using the chosen metric. Each parameter is then perturbed by the chosen step size, the image similarity re-evaluated, and the parameter providing the greatest increase in similarity is modified. This process is repeated until no further increase in similarity is achieved following perturbation. A large step size is selected for the earliest iterations, and successively reduced to obtain an accurate global alignment.

In steepest gradient descent, the set of transformation parameters Φ are updated by moving along the direction of maximum increase in similarity. For a similarity metric C and step size δ , this may be expressed as $\Phi \rightarrow \Phi + \delta \nabla_{\Phi} C$. As for the downhill descent method, successively smaller step sizes may be applied at each iteration. The gradient term $\nabla_{\Phi} C$ can be evaluated analytically for certain similarity metrics, or estimated using a finite difference method.

2.2.3 Similarity metric

A similarity metric which measures the correspondence between a pair of images must be defined in order to optimise the transformation parameters of a registration. For a transformation \mathbf{T} ,

which maps locations $\{\mathbf{x}\}$ in the target image I_t to locations $\{\mathbf{T}(\mathbf{x})\}$ in the source image I_s , the set of intensity pairs $\{(\mathbf{I}_t(\mathbf{x}), \mathbf{I}_s(\mathbf{T}(\mathbf{x})))\}$ is used to calculate the similarity metric. In the following section, this set of intensity pairs will be more compactly represented by $\{(\mathbf{t}, \mathbf{s})\} = \{(t_1, s_1), \dots, (t_n, s_n)\}$.

The simplest similarity metric, which penalises differences in intensity between corresponding voxels in two images, is the sum of squared differences, defined as

$$SSD = \frac{1}{n} \sum_{i=1}^n (t_i - s_i)^2.$$

This actually provides a measure of the distance between images, and $-SSD$ is therefore used to measure the image similarity. This metric should be used only when the correct alignment of two images would result in them having identical intensities, bar Gaussian noise. If this is not the case, cross-correlation may be an appropriate similarity metric if it can be assumed that the intensities at corresponding voxels in the two images would have a linear relationship when correctly aligned, again allowing for residual Gaussian noise. Cross-correlation is a general measure of statistical agreement, and is defined as

$$CC = \frac{\sum_n (t_i - \bar{t})(s_i - \bar{s})}{\sqrt{\sum_n (t_i - \bar{t})^2} \sqrt{\sum_n (s_i - \bar{s})^2}},$$

where \bar{t} and \bar{s} represent the mean intensities of I_t and I_s respectively.

For cases such as multi-modality image registration, where a linear relationship between the intensities in corresponding voxels cannot be assumed, an appropriate similarity metric can be defined using measures from the field of information theory. By allocating the intensities in images I_t and I_s to a series of bins, the joint histogram of the image pair is constructed. This histogram consists of a matrix of values $h(t, s)$, each of which represents the number of times a binned intensity pair (t, s) has co-occurred at the same voxel in the two images. The joint probability $p(t, s)$ of co-occurrence of the intensity pair (t, s) may be estimated as

$$p(t, s) = \frac{h(t, s)}{N},$$

where N is the total number of samples in the joint histogram. The marginal probabilities of the occurrence of intensity t in image I_t , $p(t)$, and of the occurrence of intensity s in image I_s , $p(s)$, may then be estimated as $p(t) = \sum_s p(t, s)$ and $p(s) = \sum_t p(t, s)$, respectively.

Using these probability estimates, the Shannon entropy (Shannon, 1949) of the target image I_t can be expressed as

$$H(I_t) = - \sum_t p(t) \log(p(t)),$$

with a corresponding expression for the source image I_s . The Shannon entropy describes the information content of an image. It is maximal if all possible intensities have equal probability, and minimal (zero) if $p(t) = 1$ for a single intensity value, such that the image is completely uniform. The joint entropy of the image pair can be expressed in a similar way as

$$H(I_t, I_s) = - \sum_t \sum_s p(t, s) \log(p(t, s)).$$

A similarity metric can then be defined such that it maximises the information content of each image, while minimising their shared information content. Mutual information (Collignon et al., 1995; Viola and Wells III, 1995), defined by $MI = H(I_t) + H(I_s) - H(I_t, I_s)$, thus describes the extent to which one image can be explained by another. The two images are aligned when their mutual information is maximal. Normalised mutual information, defined by

$$NMI = \frac{H(I_t) + H(I_s)}{H(I_t, I_s)},$$

is a more robust metric with respect to changes in the overlap of the two images during the registration process (Studholme et al., 1999).

2.2.4 Regularisation method

Nonrigid image registration requires the optimisation of a large number of parameters. It is therefore an ill-posed problem to which multiple solutions may exist. Restricting the possible solutions to those considered plausible transformations can be beneficial. For example, if a pair of images represent the same topology, a plausible transformation between them should not alter this topology. The smoothness of a transformation is often an appropriate measure of its plausibility (Rueckert et al., 1999). The term regularisation refers to the process of providing smooth, and therefore plausible, transformations. Regularisation methods include the application of a processing step to obtained deformations (Karaçali and Davatzikos, 2004), incorporation of models describing the mechanical properties of tissues into the registration step (Soza et al., 2004), and modification of the optimisation step such that it maximises both the image similarity and transformation smoothness. The final approach is the method used throughout this research.

For a similarity metric C_{sim} and regularisation term C_{reg} , the overall objective function can be expressed as $C = C_{sim} + \lambda C_{reg}$, where λ determines the relative contribution of each term. There is a trade-off between maximising the image similarity and maximising the transformation plausibility, and it is important that an appropriate value is selected for λ . A value that is too low could result in an implausible transformation, whereas a value that is too high could prevent the registration from achieving accurate alignment. For the nonrigid registrations presented in this work, high curvature solutions were penalised by using the bending energy associated with the transformation as a regularisation term (Rueckert et al., 1999).

2.2.5 Interpolation method

The computation of the similarity metric for a transformation \mathbf{T} is based on the correspondence between intensities at voxels \mathbf{x} in the target image I_t , and those at voxels $\mathbf{T}(\mathbf{x})$ in the source image I_s . It is unlikely that locations in the target image should exactly coincide with voxels in the deformed source image. Intensity values in the source image must therefore be interpolated

to coincide with locations in the target image before calculating the image similarity.

There are several choices of interpolation method. The simplest is nearest neighbour interpolation, in which the source image intensity at a location in the target image is assigned the value of its nearest neighbour in the source image. This method is appropriate for image data with intensity values on a nominal or ordinal scale, such as label maps for anatomical structures. For images with a continuous distribution of intensities, tri-linear interpolation may be more appropriate, in which the source image intensity at a location in the target image is determined from a linear combination of its neighbours in the source image.

2.3 Anatomical segmentation

Anatomical image segmentation describes the process of assigning labels to voxels in an image according to the anatomically defined structures they represent. In the work presented in this thesis, segmentations are defined on structural MR images and propagated to the corresponding PET images. Regional imaging features may then be extracted, such as volumes and average PET signal intensities. The following subsections provide an overview of relevant image segmentation concepts, including manual segmentation, brain atlas generation, and multi-atlas segmentation.

2.3.1 Manual segmentation

An experienced human rater can produce accurate manual segmentations of structural brain images. This requires a detailed segmentation protocol which clearly defines the way in which each structure should be delineated. Manual segmentation is the gold-standard method, but it is a challenging and time-consuming process. Its use is often not practical, particularly for a large set of images.

An automated, or semi-automated, approach could present a valuable alternative, provided that the resulting segmentations are of sufficiently high quality. Automated approaches may

also be more robust to sources of error occurring between segmentations of different images. For manual segmentation, these can include human factors such as inter- and intra-rater variation, and differences in the interpretation of the segmentation protocol. Both manual and automated segmentation methods may be affected by image acquisition effects such as image contrast characteristics, motion artefacts, and scanner calibration issues.

2.3.2 Brain atlases

The term brain atlas refers to the pairing of a brain image, or series of histological sections of a post-mortem brain, with a corresponding set of anatomical labels. The Talairach and Tournoux atlas of 1988 (Talairach and Tournoux, 1988) is additionally associated with a corresponding coordinate system, allowing the alignment of other brains with the atlas. This coordinate system is defined by requiring that the anterior and posterior commissures lie on a straight horizontal line, known as the AC-PC line, as illustrated in Figure 2.4. The anterior and posterior commissures lie in the mid-sagittal plane, and the coordinate system is therefore fully defined by requiring this plane to be vertical.

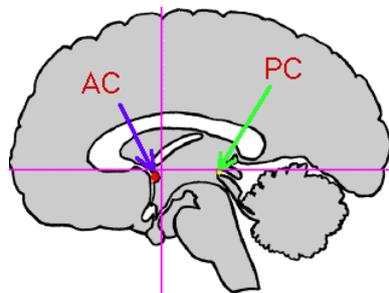


Figure 2.4: Schematic sagittal brain view showing the anterior commissure (AC) and posterior commissure (PC), and construction of the AC-PC line (Rorden, 2002).

The Talairach and Tournoux atlas is based on post-mortem sections of one hemisphere of a single brain. The Montreal Neurological Institute (MNI) aimed to define a more representative template based on a large set of MR images of normal controls. The International Consortium for Brain Mapping (ICBM) adopted one such template as their standard in 2001 (Mazziotta et al., 2001), and it remains the most commonly used today. This template is known as the ICBM152, since it was generated by averaging 152 normal MR images that had been affinely

aligned to the Talairach and Tournoux atlas.

In the work presented in this thesis, a set of 30 manually delineated brain atlases is used. Each atlas consists of the 83 anatomically defined structures described in Appendix A. The MR images used for atlas creation were acquired from 30 young, healthy adults (age range 20–54, median age 30.5 years), as described in [Hammers et al. \(2003\)](#). Protocols for the manual delineation are described in [Hammers et al. \(2003\)](#) and [Gousias et al. \(2008\)](#). An example of one of the manual segmentations overlaid onto the corresponding MRI is shown in Figure 2.5.

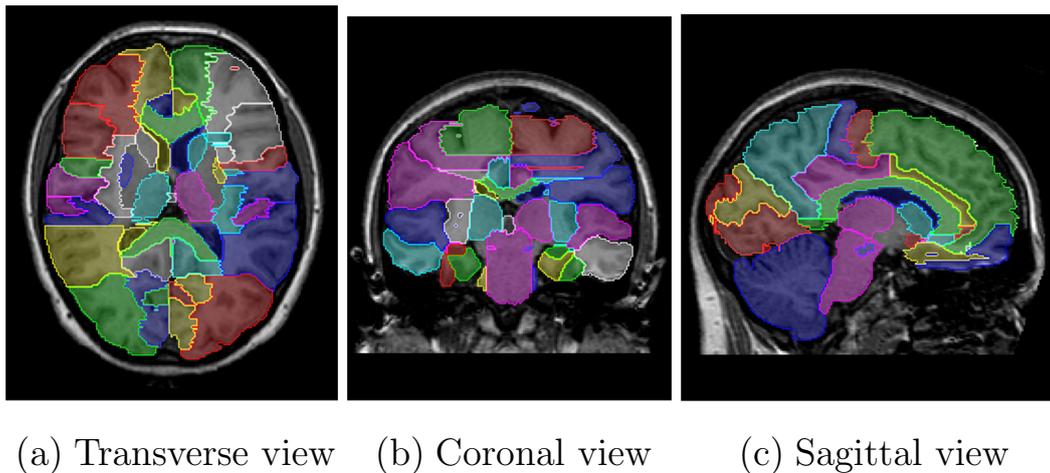


Figure 2.5: One of the manually delineated brain atlases used in this work shown overlaid onto the corresponding MR image.

2.3.3 Automated multi-atlas segmentation

There are numerous ways to approach automated image segmentation. The focus for this work is on atlas-based methods, in which expert manual segmentations are propagated to new (unseen) images. For example, the MR image of a single brain atlas may be registered to the MR image of a new subject, and the resulting transformation used to propagate the atlas label volume into the space of the unseen image. An atlas-based segmentation of the unseen image is thus automatically generated. If a set of atlases is available, multiple label volumes can be propagated to the unseen image. These transformed label volumes may then be viewed as classifiers, whose outputs can be combined to generate a consensus segmentation for the new subject. This procedure is illustrated in Figure 2.6.

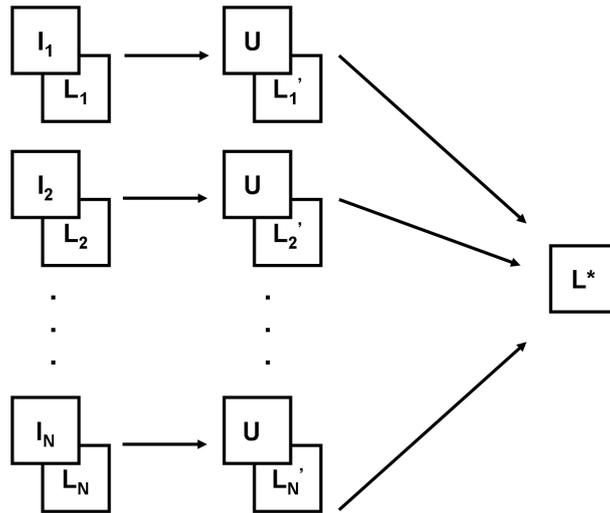


Figure 2.6: Schematic representation of the multi-atlas segmentation procedure. A set of atlas anatomical images I_i is registered to the unseen image U , and the corresponding atlas label volumes L_i propagated to the unseen image. The transformed label volumes L'_i are then fused to generate the consensus segmentation L^* for the unseen image.

An overview of classifier fusion methods is provided in [Kittler et al. \(1998\)](#). A simple approach is to consider the multiple labels assigned to a voxel as votes. A majority voting rule may then be applied, such that the consensus label for a voxel is defined as the mode of the distribution of its individual labels. Classifier fusion according to such a majority voting rule performs well in comparison with other fusion approaches in the general pattern recognition context ([Kittler et al., 1998](#)). It has also been shown to be both robust and accurate for medical image segmentation, using a database of bee brain confocal microscopy images ([Rohlfing et al., 2004](#)), as well as human brain MR images ([Heckemann et al., 2006](#)). The key benefit of fusing the propagated labels from multiple atlases is that the effect of non-systematic errors associated with the propagation of any single atlas may be reduced. For brain MR images, the fused consensus segmentation has greater accuracy than that of any individual segmentation based on the propagation of a single atlas label volume ([Heckemann et al., 2006](#)).

If a sufficiently large set of atlases is available, it can be beneficial to apply a strategy in which multi-atlas segmentation of a new subject is based on only the most suitable atlases. The STAPLE algorithm ([Warfield et al., 2004](#)) generates a probabilistic final segmentation by computing a weighted combination of the propagated label volumes, in which the weighting factor for an individual label volume is based on its estimated performance. Alternatively, an

intensity-based similarity metric may be used to rank each atlas according to its suitability for segmenting a new subject. Multi-atlas segmentation may then be performed using only the top-ranked atlases (Aljabar et al., 2009).

2.4 Statistical parametric mapping

Statistical parametric mapping (SPM) is a technique for testing hypotheses about functional neuroimaging data, such as PET. The software was first made available to the research community in 1991 (Friston et al., 1991, 1990), and there have since been many theoretical and technical advances. The latest versions of the SPM software are available from the Wellcome Trust Centre for Neuroimaging at University College London (<http://www.fil.ion.ucl.ac.uk/spm>), along with detailed documentation, background information, and related publications.

The current SPM8 software incorporates various image processing tools as well as the statistical interface. In the work presented in this thesis, SPM is used to investigate voxel-wise differences in PET signal intensities between two groups. The aspects of the software that are relevant to this particular type of analysis are described in the following subsections. The images must first undergo several processing steps, including alignment to a standard template space, smoothing, and intensity normalisation (if required). A general linear model is applied at each voxel in the processed images, and the resulting parameters are used to compute t-statistics (Student, 1908). Voxels for which the PET signal intensities are significantly different between the two groups can then be identified. Finally, since statistical tests are performed independently at each voxel, a correction for multiple comparisons is required.

2.4.1 Image processing

Before voxel-wise statistical analyses may be performed, spatial correspondences between voxels must be established across the set of images. This is achieved by alignment to a standard template space. The images are then smoothed, both to reduce residual anatomical variation,

and to improve the signal-to-noise ratio. Intensity normalisation is additionally performed to account for inter-subject variability in overall radioactivity. This can arise from physiological differences in metabolic rate, as well as variations in, for example, the injected dose or time allowed for tracer uptake before scanning. Intensity normalisation is often performed using proportional scaling relative to the cerebral global mean.

2.4.2 Statistical analyses

A general linear model is applied at each voxel in the processed images. This model describes the signal intensities at a voxel \mathbf{Y} in terms of the parameters \mathbf{B} and residual variability \mathbf{E} . For investigating differences in the PET signal intensities between two groups, this can be expressed as the multiple linear regression $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}$. The design matrix $\mathbf{X} = (\mathbf{X}_P, \mathbf{X}_C)$ contains variables indicating to which group each image belongs, as illustrated in Figure 2.7 (a).

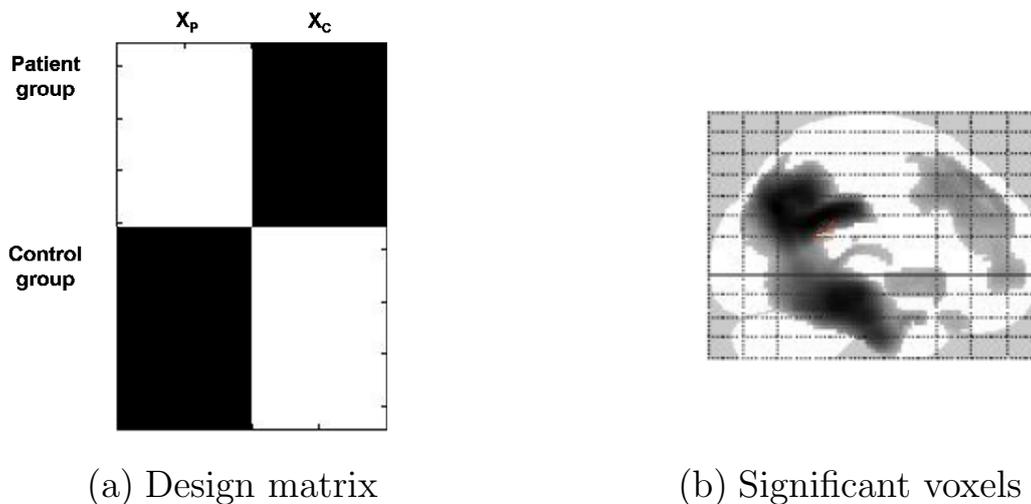


Figure 2.7: Illustrative example of the SPM analysis procedure, showing (a) the design matrix for investigating differences between two groups, and (b) an example maximum intensity projection image depicting voxels whose adjusted signal intensities are significantly different between two groups. The vectors \mathbf{X}_P and \mathbf{X}_C indicate the group to which each image belongs.

The parameters $\mathbf{B} = (B_P, B_C)$ are determined for each voxel from the analysis of variance, and then hypothesis testing is performed using classical statistical inference. To identify voxels with significantly higher adjusted PET signal intensities in a patient group than a group of healthy individuals, voxel-wise t-statistics are computed as

$$t = \frac{(B_P - B_C)}{SE},$$

where SE is the standard error, which may be determined from \mathbf{E} . In SPM, a contrast vector $\mathbf{C} = (1, -1)$ is defined, such that the t-statistic may be expressed as $t = \mathbf{CB}/SE$. A threshold α for the t-statistic can be set to define the significance level at which the null hypothesis, that the adjusted PET signal intensity at a voxel is no higher in the patient group than in the healthy group, can be rejected. An image depicting the significant voxels may then be produced, as illustrated in Figure 2.7 (b).

2.4.3 Multiple comparison correction

Multiple comparison correction is an integral part of SPM. If the large number of voxels which make up medical imaging data were treated independently, it would be likely that a number of individual voxels would reach statistical significance by chance, even if the null hypothesis were true. The significance level can be corrected to account for the number of voxels, thus controlling the family-wise error (FWE). This is the likelihood of the chance observation of a family of statistically significant voxels across the entire brain.

One standard method for controlling the FWE is the Bonferroni correction (Bonferroni, 1936, 1935), in which the significance level is adjusted to $p_{FWE} = \alpha/n$, where n is the number of independent statistical tests. However, the Bonferroni correction generally sets too conservative a threshold for functional neuroimaging data, since the spatial correlations in the images mean that the statistical tests at each voxel are not independent.

Random field theory provides a way to define a corrected threshold which takes these spatial correlations into account, based on the Euler characteristic of the t-statistic image. This can be considered as the number of “blobs” present in the image after thresholding. At high thresholds, the expected Euler characteristic is a good approximation for the FWE, and a corrected voxel-level threshold which permits a known false positive rate across the entire brain may thus be determined. The Euler characteristic depends only on the number of resolution elements

(resels) in an image (Worsley et al., 1992), which itself depends on the image smoothness, and the number of voxels. A more detailed description of the application of random field theory to functional neuroimaging data is provided in Worsley et al. (1996). Voxel-level tests identify the individual voxels which reach statistical significance based on the corrected threshold. In addition, cluster-level tests identify regions comprising a number of significant voxels, while set-level tests identify the number of significant clusters.

2.5 Conclusion

This chapter has provided an overview of methods relevant to the analyses of PET and MR imaging data which are presented in this thesis. Using these methods, multi-region and voxel-based imaging features will be extracted from MR and PET images. Image registration and anatomical segmentation techniques are important processing steps used throughout this work. Statistical parametric mapping is applied for FDG-PET image intensity normalisation, which is described in detail in Chapter 4. The following chapter presents additional background information about the use of machine learning techniques for image-based classification.

Chapter 3

Background: machine learning

3.1 Introduction

Machine learning is a branch of artificial intelligence involving the design of algorithms whose performance automatically improves through experience. Such algorithms can learn to make intelligent decisions based on their recognition of complex patterns, with applications including handwriting recognition, stock market analysis, and medical diagnosis. The focus of this research is medical diagnosis. This is a classification problem in which the aim could be, for example, to use neuroimaging data to determine whether a newly presenting patient has AD. This chapter provides an overview of machine learning concepts that are relevant to the image-based classification analyses presented in this thesis. Details of various classification algorithms are first presented in Section 3.2, followed by a description of methods with which to assess their performance in Section 3.3. Finally, a review of literature relating to the use of machine learning for image-based classification of AD is provided in Section 3.4.

3.2 Classification algorithms

An image consisting of D voxels may be represented by the D -dimensional feature vector $\mathbf{x} = (x_1, x_2, \dots, x_D)$. The overall goal of a classification algorithm is to assign this feature

vector to one of K discrete classes C_k . For the purposes of this work, only the case of $K = 2$ is considered, and the classes C_1 and C_2 are taken to be disjoint, such that each feature vector belongs to one, and only one, of the two classes. A classification algorithm could therefore take the form of a function $y(\mathbf{x})$ which returns a value that indicates the class to which the feature vector \mathbf{x} should be assigned. The parameters of the function $y(\mathbf{x})$ are optimised during a training phase in which the algorithm is presented with a set of N training examples for which the correct diagnoses are known, $\{(\mathbf{x}_i, t_i) \mid \mathbf{x}_i \in \mathbb{R}^D, t_i \in \{-1, 1\}\}_{i=1}^N$. After training, the classification performance may be assessed using new data. There are a large number of classification algorithms, and details of those relevant to this work are now provided.

3.2.1 Linear discriminant functions

The two-class Fisher linear discriminant function classifier (Fisher, 1936) aims to determine the linear combination of features that results in the maximum separation between the class means relative to the sum of the within-class variances, and therefore with the minimum possible class overlap. This linear combination may be expressed as $y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} - b$, where the decision surface $y(\mathbf{x}) = 0$ is a $(D - 1)$ -dimensional hyperplane, as illustrated in Figure 3.1. A feature vector \mathbf{x} belongs to class C_1 if $y(\mathbf{x}) \geq 0$, and to class C_2 if $y(\mathbf{x}) < 0$. The value of $y(\mathbf{x})$ gives a signed measure of the distance from the decision surface, which may be interpreted as a measure of the certainty with which the feature vector \mathbf{x} is assigned to a class.

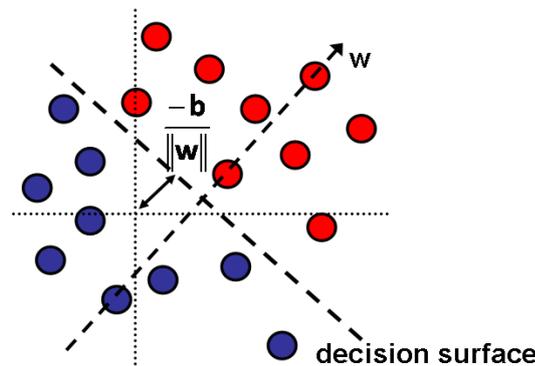


Figure 3.1: A 2-D illustration of the decision surface obtained using a Fisher linear discriminant function. Feature vectors belonging to class C_1 are shown in red, and those belonging to class C_2 are shown in blue. The decision surface is defined by its orthogonality to the feature weight vector \mathbf{w} and its distance from the origin, which depends on the threshold b .

The maximisation criterion described above may be written as

$$\max_{\mathbf{w}} J(\mathbf{w}) = \max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}},$$

where \mathbf{S}_B is the between-class scatter matrix, and \mathbf{S}_W the within-class scatter matrix. For two classes C_1 and C_2 , the scatter matrices are defined as

$$\begin{aligned} \mathbf{S}_B &= \sum_{i=1}^2 N_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T = \frac{N_1 N_2}{N} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \\ \mathbf{S}_W &= \sum_{i=1}^2 \sum_{j \in C_i} (\mathbf{x}_j - \boldsymbol{\mu}_i)(\mathbf{x}_j - \boldsymbol{\mu}_i)^T, \end{aligned}$$

where N_i is the number of feature vectors in class C_i , N the total number of feature vectors, $\boldsymbol{\mu}_i$ the mean of class C_i , and $\boldsymbol{\mu}$ the overall mean. Since only the direction of \mathbf{w} is important, not its magnitude, the maximisation may equivalently be expressed as the constrained optimisation

$$\begin{aligned} \max_{\mathbf{w}} \quad & \mathbf{w}^T \mathbf{S}_B \mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^T \mathbf{S}_W \mathbf{w} = 1. \end{aligned}$$

By using Lagrange multipliers (Fletcher, 1987), this may be re-expressed as the unconstrained optimisation of the Lagrangian $L(\mathbf{w}, \alpha) = \mathbf{w}^T \mathbf{S}_B \mathbf{w} - \alpha(\mathbf{w}^T \mathbf{S}_W \mathbf{w} - 1)$. By finding the stationary point, the feature weight vector may be expressed as $\mathbf{w} \propto \mathbf{S}_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, and the threshold as $b = \mathbf{w}^T \boldsymbol{\mu}$.

3.2.2 Support vector machines

A two-class support vector machine (SVM) classifier aims to construct a hyperplane that maximises the margin, which is the distance between the closest points on either side of the boundary. These points are known as the support vectors, and their role in the construction of a maximum-margin hyperplane is illustrated in Figure 3.2. The original SVM algorithm was a linear classifier (Vapnik and Lerner, 1963), but there have since been modifications to deal with data that are not linearly separable. A soft-margin formulation, which allows for mislabelled data, has been proposed (Cortes and Vapnik, 1995), as well as a way to use the kernel trick (Aizerman et al., 1964) to create nonlinear classifiers (Boser et al., 1992). These three formulations are described in further detail in the following subsections.

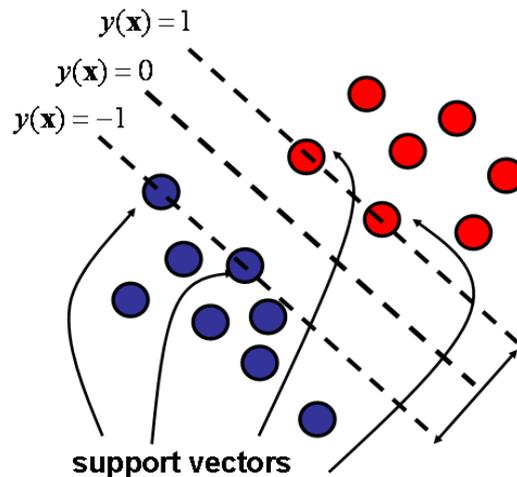


Figure 3.2: 2-D illustration of the construction of a maximum-margin hyperplane. This decision surface maximises the distance between the support vectors, indicated by the arrow.

Before the application of a SVM classifier, it is important that both the training and test data are scaled so that features with high variance do not dominate those with lower variance (Hsu et al., 2010; Juszczak et al., 2002). For the work presented in this thesis, all features are linearly scaled to the range $[-1, +1]$.

Linear SVM

The decision surface of a linear SVM classifier is described by $y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} - b = 0$, as for the Fisher linear discriminant function classifier. The feature weight vector \mathbf{w} and threshold b

are then chosen such that the margin, or distance between the support vectors, is maximised. As illustrated in Figure 3.2, the support vectors lie on two parallel hyperplanes described by $y(\mathbf{x}) = 1$ and $y(\mathbf{x}) = -1$, such that the distance between them is $2/\|\mathbf{w}\|$. The maximisation of the margin can therefore be expressed as the constrained optimisation

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} \\ \text{subject to} \quad & t_i(\mathbf{w}^T \mathbf{x}_i - b) \geq 1, \end{aligned}$$

where the constraint ensures that no feature vectors fall within the margin. By using Lagrange multipliers, this may be re-expressed as the unconstrained optimisation

$$\begin{aligned} \min_{\mathbf{w}, b} \max_{\alpha} \quad & \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^N \alpha_i [t_i(\mathbf{w}^T \mathbf{x}_i - b) - 1] \right\} \\ \text{subject to} \quad & \alpha_i \geq 0, \end{aligned}$$

from which an expression for the feature weight vector \mathbf{w} can be derived in terms of a linear combination of the feature vectors,

$$\mathbf{w} = \sum_{i=1}^N \alpha_i t_i \mathbf{x}_i.$$

The decision surface is thus expressed in terms of the support vectors, since only their corresponding α_i are non-zero. A robust solution for the threshold b may then be found by averaging over the N_{sv} support vectors,

$$b = \frac{1}{N_{sv}} \sum_{i=1}^{N_{sv}} (\mathbf{w}^T \mathbf{x}_i - t_i).$$

The primal form of the Lagrangian $L(\mathbf{w}, b, \boldsymbol{\alpha})$ may be equivalently written in dual form by substituting the above expression for \mathbf{w} . The dual form,

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \tilde{L}(\boldsymbol{\alpha}) = & \max_{\boldsymbol{\alpha}} \left\{ \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j t_i t_j \mathbf{x}_i^T \mathbf{x}_j \right\} \\ \text{subject to } & \alpha_i \geq 0 \quad \text{and} \quad \sum_{i=1}^N \alpha_i t_i = 0, \end{aligned}$$

expresses the optimisation criterion in terms of inner products of the feature vectors. This is an important property for the creation of nonlinear SVM classifiers.

Soft-margin SVM

The soft-margin SVM formulation may be applied in cases where no linear hyperplane exists which can separate the data. Slack variables $\boldsymbol{\xi}$ are introduced, which measure the degree of misclassification of the feature vectors. The optimisation becomes a trade-off between maximising the margin and minimising the degree of misclassification. This trade-off is controlled by the penalty parameter C , such that the constrained optimisation may be expressed as

$$\begin{aligned} \min_{\mathbf{w}, \boldsymbol{\xi}, b} & \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i \right\} \\ \text{subject to } & t_i (\mathbf{w}^T \mathbf{x}_i - b) \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0. \end{aligned}$$

By using Lagrange multipliers, the problem may be re-expressed as the unconstrained optimisation

$$\min_{\mathbf{w}, \xi, b} \max_{\alpha, \beta} \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i [t_i (\mathbf{w}^T \mathbf{x}_i - b) - 1 + \xi_i] - \sum_{i=1}^N \beta_i \xi_i \right\}$$

subject to $\alpha_i, \beta_i \geq 0$,

which may be written in dual form as

$$\max_{\alpha} \tilde{L}(\alpha) = \max_{\alpha} \left\{ \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j t_i t_j \mathbf{x}_i^T \mathbf{x}_j \right\}$$

subject to $0 \leq \alpha_i \leq C$ and $\sum_{i=1}^N \alpha_i t_i = 0$.

The only change from the linear SVM optimisation is the upper bound on the α_i .

Nonlinear SVM

In cases where the data are not linearly separable in the input feature space, a nonlinear function $\phi(\mathbf{x})$ may be used to map each feature vector into a higher-dimensional space. As illustrated in Figure 3.3, the data are separated by a linear hyperplane in this new space.

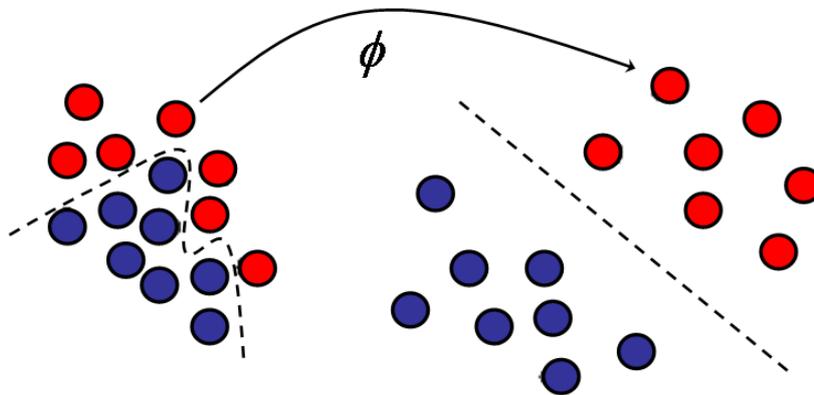


Figure 3.3: A nonlinear boundary in the input feature space becomes a linear hyperplane in a higher-dimensional space to which feature vectors are mapped using the nonlinear function ϕ .

The linear SVM algorithm may then be solved in the transformed feature space by optimising the dual form Lagrangian

$$\tilde{L}(\boldsymbol{\alpha}) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j t_i t_j \boldsymbol{\phi}(\mathbf{x}_i)^T \boldsymbol{\phi}(\mathbf{x}_j).$$

The optimisation criterion is thus expressed in terms of inner products of the transformed feature vectors. By choosing the nonlinear mapping $\boldsymbol{\phi}$ such that these inner products can be expressed in terms of a kernel function $\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) \equiv \boldsymbol{\phi}(\mathbf{x}_i)^T \boldsymbol{\phi}(\mathbf{x}_j)$, it is not necessary to explicitly perform the mapping. The optimisation problem may therefore be solved even in very high-dimensional spaces. The most commonly used kernel is the Gaussian radial basis function, given by $\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$, where $\gamma > 0$ describes the width.

3.2.3 Boosting

Boosting is an ensemble method which aims to generate a single strong classifier by combining a number of weak classifiers. A weak classifier is defined as having an accuracy little better than 50%, whereas a strong classifier may achieve an arbitrarily good accuracy. In this work, weak classifiers are simple thresholds on each of the D features.

The majority of boosting algorithms work by iteratively selecting weak classifiers, and then combining them to form a single strong classifier using a weighted summation, in which each weak classifier is weighted according to its performance. All feature vectors are initially assigned equal weightings. They are then re-weighted at every iteration, after the inclusion of a new weak classifier, such that the weights for incorrectly classified examples are increased, while those for correctly classified examples are decreased. Later weak classifiers therefore focus more on those cases which were most difficult to classify during the earlier iterations. Some boosting algorithms, however, decrease the weightings of repeatedly mis-classified examples in order to reduce the influence of outliers. An overview of the boosting process is illustrated in Figure

3.4.

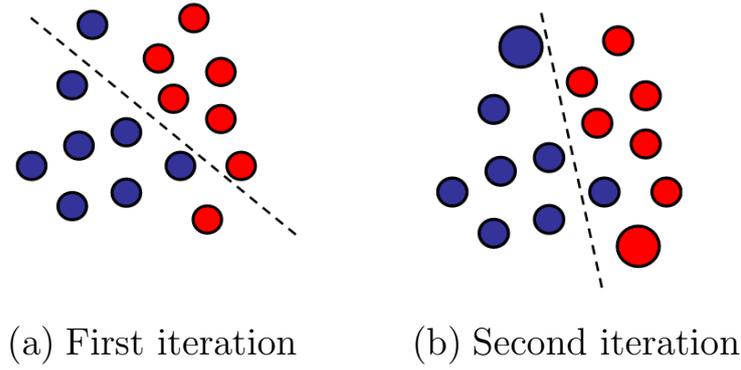


Figure 3.4: Illustration of the first two iterations in a typical boosting procedure. At the first iteration (a), all feature vectors are weighted equally and a weak classifier selected. At the second iteration (b), the weights of incorrectly classified examples are increased and a second weak classifier selected.

There are a number of boosting algorithms available, of which adaptive boosting (AdaBoost) (Freund and Schapire, 1997) is one of the most popular. The AdaBoost algorithm begins by assigning equal weightings to all N feature vectors, such that $D_1(i) = 1/N$. At each iteration ($j = 1, 2, \dots, J$), the weak classifier h_j resulting in the minimum classification error is selected from the set of weak classifiers \mathcal{H} , such that

$$h_j = \operatorname{argmax}_{h_j \in \mathcal{H}} |0.5 - \epsilon_j|,$$

where the error ϵ_j is the sum of the weights of the incorrectly classified examples. The selected weak classifier h_j is assigned a weighting $\alpha_j = 1/2 \ln((1 - \epsilon_j)/\epsilon_j)$ according to its performance, and the weightings are then updated such that

$$D_{j+1}(i) = \frac{D_j(i) \exp(-\alpha_j t_i h_j(\mathbf{x}_i))}{Z_j}$$

where $Z_j = \sum_{i=1}^N D_j(i) \exp(-\alpha_j t_i h_j(\mathbf{x}_i))$

$$= 2\sqrt{\epsilon_j(1 - \epsilon_j)}.$$

The final strong classifier is constructed from the selected weak classifiers as $y(\mathbf{x}) = \text{sign} \sum_{j=1}^J \alpha_j h_j(\mathbf{x})$.

3.2.4 Random forests

A random forest is an ensemble classifier consisting of many decision trees, where the final predicted class for a test example is the mode of the predictions of all individual trees, as illustrated in Figure 3.5. Random forests, developed by Leo Breiman and Adele Cutler (Breiman, 2001), combine bootstrap aggregation (bagging) (Breiman, 1996) and random feature selection (Amit and Geman, 1997; Ho, 1998) in order to construct a collection of decision trees exhibiting controlled variation. The training set for each individual tree in a random forest is constructed by sampling N examples at random with replacement from the N available examples in the dataset. This is known as bootstrap sampling, and bagging describes the aggregation of predictions from the resulting collection of trees. As a result of the bootstrap sampling procedure, approximately one third of the available N examples are not present in the training set of each tree. These are referred to as the out-of-bag data of the tree, for which internal test predictions can be made. By aggregating the predictions of the out-of-bag data across all trees, an internal estimate of the generalisation error of the random forest can be determined.

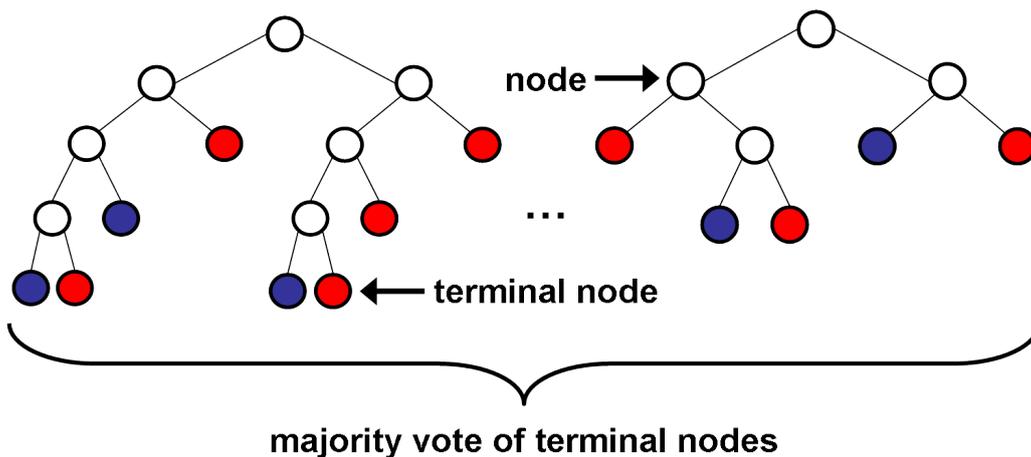


Figure 3.5: Illustration of a random forest, showing two trees in detail. Each node is partitioned based on a single feature, and each branch ends in a terminal node which provides a prediction for the class of a test example based on the path taken through the tree. The colour of a terminal node indicates its class prediction. The final predicted class for a test example is the mode of the predictions of all individual trees.

At each node in a tree, $d \ll D$ features are randomly selected, and the node is partitioned using the best possible binary split. It has been shown (Breiman, 2001) that the random forest error rate depends on both the correlation between trees $\bar{\rho}$ and the strength of the collection of trees s , such that an upper bound for the generalisation error is given by $\bar{\rho}(1 - s^2)/s^2$. A reduction in the selected value of d results in a corresponding reduction in both $\bar{\rho}$ and s . An optimal range for d therefore exists, which is generally quite wide. The recommended default value for most applications is $d = \sqrt{D}$ (Liaw and Wiener, 2002).

A parent node n_p is partitioned into child nodes n_l and n_r according to an impurity criterion which aims to maximise the homogeneity of the child nodes with respect to the parent node. Impurity is assessed using the Gini index I_G , which measures the likelihood that an example would be incorrectly labelled if it were randomly classified according to the distribution of labels within the node. For two classes C_1 and C_2 , the Gini index of a node n may be expressed as

$$I_G(n) = 1 - \sum_{K=1}^2 p_K^2,$$

where p_K is the relative proportion of examples belonging to class K present in the node n . The Gini index therefore ranges from a minimum of zero to a maximum value of $(1 - 1/K)$. A value of zero indicates that the node contains only examples belonging to a single class, and the maximum value indicates that the node contains examples belonging to both classes in equal proportions. The best possible binary split is the one which maximises the improvement in the Gini index $\Delta I_G(n_p) = I_G(n_p) - p_l I_G(n_l) - p_r I_G(n_r)$, where p_l and p_r are the proportions of examples in node n_p that are assigned to child nodes n_l and n_r , respectively. The Gini index can also be used to assess the relative importances of the various features for classification. A measure of the importance of an individual feature may be computed by summing the decreases in the Gini index ΔI_G occurring at all nodes in the forest which are partitioned based on that feature.

Random forests can also provide pairwise measures of the similarities between examples in the dataset. All N feature vectors are passed down each tree in the forest, and if feature vectors

\mathbf{x}_i and \mathbf{x}_j finish in the same terminal node in a tree, their similarity s_{ij} is increased by one. The final similarity measures are normalised by the total number of trees in the forest. The similarities form an $N \times N$ matrix with elements s_{ij} , and corresponding distance matrix elements $d_{ij} = 1 - s_{ij}$ (Cox and Cox, 2001). The matrix is symmetric, positive-definite, bounded from above by one, and has diagonal elements equal to one.

Manifold learning techniques may be applied to the pairwise similarity measures to find an appropriate coordinate embedding for the feature vectors, such that the similarity relationships between them are preserved. A review of the most popular manifold learning techniques, as applied to medical imaging, is provided in Aljabar et al. (2012). Classical multidimensional scaling (MDS) (Torgerson, 1952) is used for the work presented in this thesis, and is now described. Using MDS, the matrix of coordinates \mathbf{X} is derived by performing an eigenvalue decomposition on the matrix of scalar products

$$\mathbf{B} = \mathbf{X}^T \mathbf{X} = -\frac{1}{2} \left\{ d_{ij}^2 - \frac{1}{N} \sum_{i=1}^N d_{ij}^2 - \frac{1}{N} \sum_{j=1}^N d_{ij}^2 + \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N d_{ij}^2 \right\}.$$

Retaining only the eigenvectors corresponding to the k largest-valued eigenvalues leads to a lower-dimensional embedding for the data, which can be useful for visualising its structure. A goodness-of-fit parameter G , describing the extent to which the selected k eigenvectors represent the full $N \times N$ similarity matrix, can be useful in selecting an appropriate dimensionality for the lower-dimensional embedding. One possible measure of goodness-of-fit is given by

$$G = \frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^N \max(\lambda_j, 0)},$$

where the eigenvalues λ_j are sorted in decreasing order (Mardia et al., 1979).

3.3 Classifier performance

Measures of classifier performance are required to assess the applicability of a trained algorithm using independent test data, as well as for the optimisation of parameters during training. The simplest performance metric is accuracy, the proportion of examples that are correctly labelled by the classifier. However, this does not always provide an appropriate measure of performance, and other relevant metrics are described in Section 3.3.1. The technique of cross-validation may be used to assess the generalisation performance of a classifier, as there is often no appropriate independent dataset available for testing. A single round of cross-validation involves partitioning the data into two subsets, so that it may be trained using one and tested using the other. Results are generally reported as the average over multiple rounds in which different partitions of the data are used. Details of the most common cross-validation methods are provided in Section 3.3.2.

3.3.1 Performance metrics

The performance of a binary classifier can be visualised using a confusion matrix, as shown in Table 3.1. The number of examples correctly labelled by the classifier are located on the diagonal. These may be divided into true positives TP , representing correctly identified patients, and true negatives TN , representing correctly identified controls. The number of examples incorrectly labelled by the classifier may be divided into false positives FP , representing controls incorrectly classified as patients, and false negatives FN , representing patients incorrectly classified as controls.

True class	Predicted class	
	C_1 (patients)	C_2 (controls)
C_1 (patients)	TP	FN
C_2 (controls)	FP	TN

Table 3.1: Confusion matrix for a binary classifier which aims to distinguish between classes C_1 (patients) and C_2 (controls).

The accuracy measures the proportion of examples that are correctly labelled by a classifier,

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}.$$

This may not be a good performance metric if the class distribution of the dataset is unbalanced. For example, if class C_1 is much larger than C_2 , a high accuracy value could be obtained by a classifier which labels all examples as belonging to class C_1 . The sensitivity and specificity,

$$\text{sensitivity} = \frac{TP}{TP + FN} \quad \text{and} \quad \text{specificity} = \frac{TN}{TN + FP},$$

may provide a better assessment of the overall performance of a classifier. Sensitivity measures the proportion of correctly identified patients, and specificity measures the proportion of correctly identified controls. The balanced accuracy, which treats both classes with equal importance, may then be expressed as

$$\text{balanced accuracy} = \frac{\text{sensitivity} + \text{specificity}}{2}.$$

An ideal classifier would achieve 100% sensitivity and specificity, but in general there is a trade-off between these two measures. This can be investigated using a receiver operating characteristic (ROC) curve. As shown in Figure 3.6, a ROC curve shows the relationship between the true positive rate (sensitivity) and false positive rate ($1 - \text{specificity}$) as the discrimination threshold of the binary classifier is varied. This curve could be used to select the optimal threshold for a particular application. For example, to identify patients in the earliest stages of disease, it may be desirable to select a threshold which results in high sensitivity, at the expense of reduced specificity. The area under a ROC curve (AUC) may be interpreted as an aggregated measure of classifier performance (Flach et al., 2011).

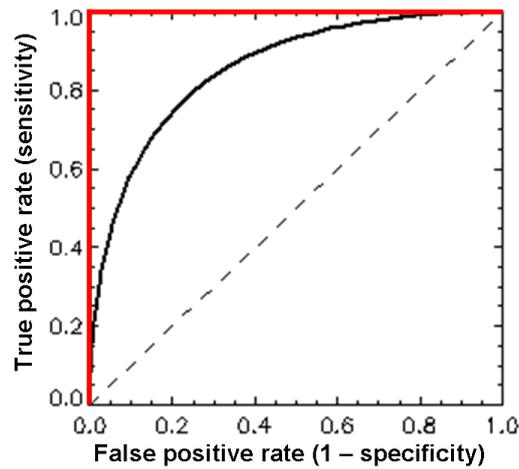


Figure 3.6: Illustration of the ROC curve for a binary classifier. The solid line shows the relationship between the sensitivity and specificity as the discrimination threshold of the classifier is varied. This may be compared with the dashed line of no-discrimination, and the red line depicting an ideal classifier.

3.3.2 Cross-validation

The parameters of a classifier are optimised based on the training data. An independent test set is therefore required for making a reliable assessment of the applicability of the classifier to new data. Cross-validation provides a way to measure this generalisation performance when no such test data are available. One commonly used method is k -fold cross-validation, in which the data are randomly partitioned into k subsets. A single cross-validation fold involves using $(k - 1)$ subsets for training the classifier, and the remaining data for testing. This process is repeated k times, such that each of the subsets is used once for testing, and the results are averaged over the folds. An alternative method is repeated random sampling, in which the dataset is randomly partitioned into training and test sets of fixed sizes. For example, a single round may involve randomly selecting 75% of the data for training, with the remaining 25% used for testing. This process can then be repeated, and the results averaged over the repetitions. Repeated random sampling has the advantage that the proportions of the training and test sets are not dependent on the number of repetitions. However, there may be some overlap between test sets, and the method also exhibits Monte Carlo variation. This means that the results will vary if the analysis is repeated using different partitions of the data. If two classes C_1 and C_2 are not of equal sizes, the training and test sets should be selected such that they contain examples from the two classes in approximately equal proportions to the full

dataset. This is known as stratified cross-validation, and has been shown to produce results with a lower variance than regular cross-validation ([Kohavi, 1995](#)).

Both the k -fold and repeated random sampling cross-validation methods generate a distribution of performance values which may be averaged across the folds or repetitions. The statistical significance of differences between the results of two classifiers may be assessed by performing unpaired t-tests between these distributions. In addition, permutation testing may be applied to assess whether the results of a classifier are significantly different from chance. Permutation testing involves performing cross-validation on data for which the diagnostic labels have been randomly permuted. This results in a distribution of classification results under the null hypothesis that the classifier cannot accurately predict the clinical labels from the data. Unpaired t-tests between the distribution of observed results and that obtained from permutation testing indicate whether the observed results are significantly different from chance.

3.4 Application to Alzheimer's disease

Automated image-based classification of individual patients could provide valuable diagnostic support for clinicians, when considered alongside cognitive assessment scores and traditional visual image analysis. The ADNI study provides an ideal dataset for classification experiments, since it approximates a clinical population due to its large size and diversity. The gender ratio of ADNI participants is, however, not that expected of a general clinical population. Approximately 43% of ADNI participants are females, compared with the 57% which would be expected based on the gender ratio observed amongst adults aged over 65 in the United States ([Central Intelligence Agency, 2012](#)).

Three recent studies which have performed image-based AD classification using cross-sectional ADNI FDG-PET data are summarised in Table 3.2. [Salas-Gonzalez et al. \(2010\)](#) additionally report an accuracy of 83% for discriminating between MCI patients and healthy controls (HC). Machine learning techniques using FDG-PET images from other datasets have also been successful in distinguishing AD patients from HC, as well as from patients with frontotemporal

dementia (Kippenhan et al., 1994; Xia et al., 2008).

Study	Results (%)		
	Acc.	Sens.	Spec.
Haense et al. (2009) sum of abnormal t-values in predefined areas (Herholz et al., 2002)	-	83	78
Hinrichs et al. (2009) boosting applied to voxel-wise features	84	-	-
Salas-Gonzalez et al. (2010) linear SVM applied to selected voxel-wise features	87	-	-

Table 3.2: Summary of classification results based on cross-sectional ADNI FDG-PET data. Studies report either the accuracy (acc.), or sensitivity (sens.) and specificity (spec.) for distinguishing AD patients from healthy controls.

There are, however, many more classification studies based on structural MR imaging data, since anatomical imaging with either MRI or CT is routinely used in clinical practice for dementia patients. For example, a recent study (Cuingnet et al., 2011) compares ten high-dimensional classification methods applied to 509 baseline ADNI 1.5 T MR images. Two methods use only the hippocampal shape or volume, while the remainder are whole-brain approaches. These use either cortical thickness measures, or voxel-wise tissue class probabilities for grey matter, white matter and CSF. High accuracies in distinguishing AD patients from HC (up to 81% sensitivity and 95% specificity) are reported for whole-brain approaches. Four of the ten methods were able to distinguish MCI patients who later progressed to AD (pMCI) from those who remained stable over 18 months (sMCI) slightly more accurately than a random classifier. However, the results were not significantly different from chance ($p > 0.05$).

SVM classifiers have been successfully applied by several groups for discriminating between AD patients and HC based on structural MRI. Reported accuracies generally fall between 80% and 95% (Duchesne et al., 2008; Klöppel et al., 2008; Magnin et al., 2009). Vemuri et al. (2008) achieve an accuracy of 89% based on a large cohort of 190 AD patients and 190 age- and gender-matched HC. Three studies performed using other machine learning techniques are summarised in Table 3.3. Davatzikos et al. (2008a) have additionally applied their high-dimensional pattern classification approach in a small study comparing 15 HC with 15 MCI patients, at least 10 of whom later converted to AD. They report an accuracy of 90% for discriminating between MCI patients and HC. Colliot et al. (2008) perform classification based on automatically segmented

hippocampal volumes, achieving accuracies of 84% between AD patients and HC, 73% between MCI patients and HC, and 69% between AD and MCI patients.

Study	No. AD patients	No. HC	Accuracy (%)
Davatzikos et al. (2008b) high-dimensional pattern classification	37	37	100
Plant et al. (2010) data mining	32	18	92
Aljabar et al. (2008) spectral clustering	8	18	84

Table 3.3: Summary of classification results based on MR imaging data.

There is increasing interest in using multi-modality imaging and non-imaging data for classification. For example, [Zhang et al. \(2011\)](#) apply a multi-kernel learning approach to cross-sectional FDG-PET and MR imaging data and CSF biomarker measures. They report classification accuracies of 93% between AD patients and HC, and 76% between MCI patients and HC, when using all three modalities in combination. These results are superior to those obtained when using any one modality independently. [Hinrichs et al. \(2011\)](#) have also investigated the application of kernel combination methods. They apply these to both cross-sectional and longitudinal FDG-PET and MR imaging data, as well as CSF biomarker measures, neuropsychological status examination scores, and ApoE genotype information. They also report that the use of multi-modality data leads to superior classification performance compared with that based on any individual modality.

3.5 Conclusion

This chapter has provided details of the machine learning algorithms applied as part of the work presented in this thesis, as well as descriptions of the methods with which their performance may be assessed. In addition, an overview of recent research related to the image-based classification of AD and MCI has been presented. The later chapters incorporate more focussed and detailed reviews of the most recent related research. The following four chapters present the main contributions of the thesis.

Chapter 4

Multi-region baseline FDG-PET for classification

Work in this chapter has, in part, been presented in:

K. R. Gray, R. Wolz, S. Keihaninejad, R. A. Heckemann, P. Aljabar, A. Hammers, and D. Rueckert. Regional analysis of FDG-PET for use in the classification of Alzheimer's disease. (In Proceedings) IEEE 8th International Symposium on Biomedical Imaging (ISBI'11), 1082–1085, 2011.

4.1 Introduction

This chapter presents a regional analysis of baseline FDG-PET imaging data from the ADNI study. Whole-brain segmentations into 83 anatomically defined regions were automatically generated in the native MRI space of each subject, and used to extract regional signal intensities from the corresponding FDG-PET images. The resulting regional FDG-PET data were used to investigate group differences between AD and MCI patients and HC, as well as in binary classification experiments. Part of this work involved the investigation of image intensity normalisation methods with which to account for non-disease related inter-subject variations in the CMRgl.

The ADNI participants whose imaging data are used in this work are first described in Section 4.2, followed by the image acquisition and pre-processing steps in Section 4.3, and the anatomical segmentation procedure in Section 4.4. Investigations into image intensity normalisation methods are then presented in Section 4.5. FDG-PET intensity normalisation is often performed relative to the cerebral global mean (CGM). Recent research, based on a study of 36 AD patients, 11 MCI patients and 15 HC, suggests that using regions of the brain which are relatively preserved in AD may provide improved group discrimination (Yakushev et al., 2009). These two intensity normalisation approaches were compared using ADNI FDG-PET imaging data from 71 AD patients, 147 MCI patients and 69 HC. The superior group discrimination observed in this large cohort using the method proposed by Yakushev et al. (2009) supports the previously reported findings.

Multi-region analyses of group differences are described in Section 4.6. The findings are consistent with previously reported voxel-based group differences amongst the ADNI subjects (Langbaum et al., 2009). Binary classification experiments are presented in Section 4.7. In particular, the most challenging, but clinically significant, comparison between pMCI and sMCI patients was investigated. Classification was performed using a SVM classifier, as well as an algorithm which combines feature extraction using AdaBoost with a linear discriminant function classifier. The value for image-based classification of regional information extracted from FDG-PET images acquired at a single timepoint was thus investigated using two classifiers exhibiting different behaviours. Classification results are comparable with those obtained using data from MRI, or biomarkers obtained invasively from the CSF. These findings support the use of FDG-PET for the early diagnosis of AD and for monitoring its progression.

4.2 Imaging data

Clinical and imaging data used in this chapter were obtained from the ADNI database (<http://adni.loni.ucla.edu>), which was briefly introduced in Section 1.5. The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging

and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organisations, as a \$60 million, five-year public-private partnership. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, M.D., VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations ([ADNI Data Sharing and Publications Committee, 2012](#)).

ADNI participants were recruited from over 50 sites across the United States and Canada, to make up a total of 819 adults aged between 55 and 90. These include 229 HC to be followed for three years, 380 MCI patients to be followed for three years, and 210 mild AD patients to be followed for two years. The key eligibility criteria for enrolment into one of the three groups were as follows. HC had a mini mental state examination (MMSE) score ([Folstein et al., 1975](#)) of between 24 and 30 (inclusive), a clinical dementia rating (CDR) ([Berg, 1988](#)) of 0, and no diagnosis of depression, MCI, or dementia. MCI patients had a MMSE score of between 24 and 30 (inclusive), a subjective memory complaint, objective memory loss measured by education-adjusted scores on the Wechsler memory scale logical memory II ([Wechsler, 1987](#)), a CDR of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia. Mild AD patients had a MMSE score of between 20 and 26 (inclusive), a CDR of 0.5 or 1.0, and met the NINCDS-ADRDA criteria for probable AD ([McKhann et al., 1984](#)). Further information is available on the ADNI information website (<http://www.adni-info.org>).

Approximately 50% of ADNI participants had FDG-PET scans, and 404 baseline images were available to download in total (as of December 2010). A number of images were excluded from the analyses presented in this chapter, due to scanner quantification issues, image processing problems, or because the associated diagnoses did not clearly fall into one of the four clinical categories (AD, pMCI, sMCI, HC). These exclusions are summarised in Figure 4.1, and will be

explained in further detail later in the chapter.

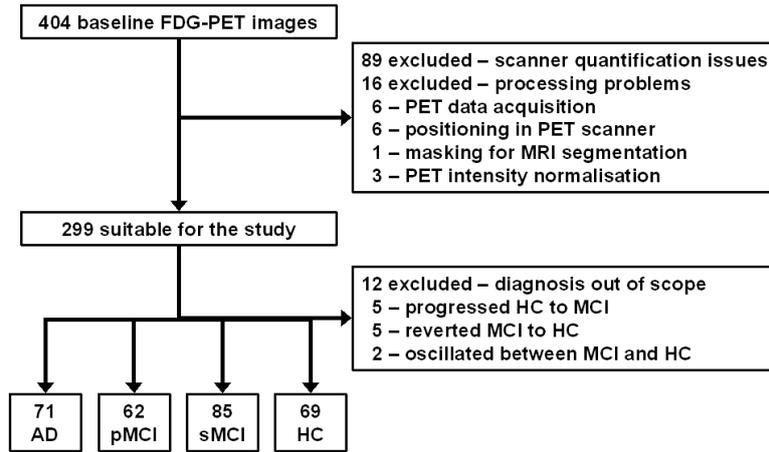


Figure 4.1: Summary of exclusions, showing that of the 404 baseline FDG-PET images available to download, 287 were suitable for the analyses presented in this chapter.

The MCI patients were divided into pMCI and sMCI based on changes in clinical status occurring over 24 ± 11 (range 0 – 36) months. ADNI subject identifiers for all excluded images are listed in Appendix B, along with the reasons for their exclusion. Groupwise characteristics are provided in Table 4.1 for the 287 subjects whose imaging data were used in this work. The mean age does not vary significantly ($p > 0.01$) on t-test between the clinical groups.

	N(F)	Age (mean \pm std. dev.)	MMSE Score (mean \pm std. dev.)	CDR (%)		
				0	0.5	1
AD	71(29)	76.2 \pm 7.0	23.3 \pm 2.2	0	38	62
pMCI	62(22)	75.2 \pm 6.9	26.8 \pm 1.7	0	100	0
sMCI	85(23)	76.0 \pm 6.9	27.5 \pm 1.7	0	100	0
HC	69(27)	75.6 \pm 5.0	29.0 \pm 1.1	100	0	0

Table 4.1: Clinical and demographic information for the study population ($N = 287$). For each clinical group, the total number of subjects (N) and number of females (F) are shown, along with the average age, average MMSE score, and CDR distribution.

4.3 Image acquisition and pre-processing

The anatomical segmentations required for regional sampling were automatically generated in the native MRI space of each subject (see Section 4.4). It was therefore necessary to co-register each FDG-PET image with its corresponding MRI. The following subsections describe the

acquisition of both the FDG-PET and MR images, the necessary pre-processing steps, and the inter-modality image registration. These steps are summarised in Figure 4.2.

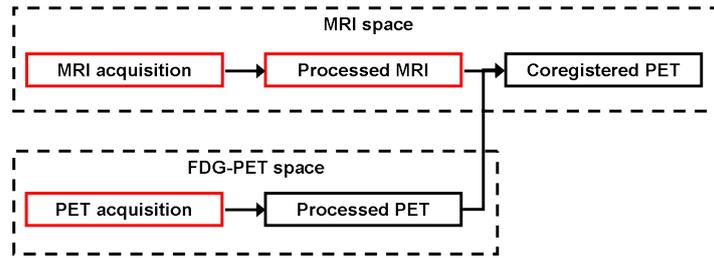


Figure 4.2: Summary of the FDG-PET and MR image acquisition, pre-processing and co-registration steps. Those outlined in red were performed by ADNI investigators, while those outlined in black were performed as part of this work.

4.3.1 ADNI FDG-PET acquisition

Baseline FDG-PET images from 404 ADNI participants, acquired using Siemens, General Electric (GE) and Philips PET scanners, were available. For this work, all those acquired using either the Siemens HRRT or BioGraph HiRez scanners ($n = 89$) were excluded, due to differences in the pattern of FDG metabolism that were discovered during the ADNI quality control process. Further information is available on the ADNI PET Core website (<http://www.loni.ucla.edu/twiki/bin/view/ADNI/ADNIPETCore>). The remaining imaging data consisted of baseline FDG-PET scans from 315 subjects, acquired using 12 scanner models, as shown in Figure 4.3.

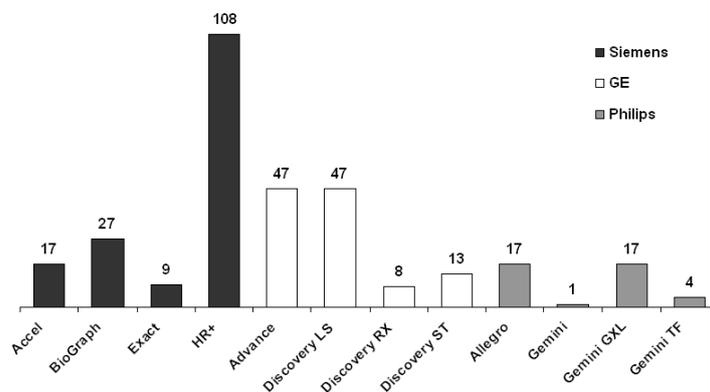


Figure 4.3: Number of baseline FDG-PET images acquired using each of the 12 scanner models.

FDG-PET images were collected from 40 sites, and acquired according to one of three standard

protocols. In the majority of cases ($n = 263$), a 30-minute dynamic scan consisting of six 5-minute frames was acquired, beginning approximately 30 minutes after the intravenous injection of 185 ± 19 MBq of FDG. Subjects were instructed to fast for at least four hours prior to scanning, and then rest in a dimly lit room with their eyes open during the tracer uptake period. Since dynamic scanning was not available on the Siemens BioGraph, scans performed using this model consisted of a single 30-minute frame ($n = 27$). Static acquisitions were also performed for seven subjects using other models. A limited number of subjects ($n = 18$) had 60-minute quantitative dynamic scans, including venous sampling, which began simultaneously with the injection of FDG, and whose final 30 minutes consisted of six 5-minute frames.

Data were corrected for both scatter and measured attenuation, which was determined using either the CT scan for PET/CT scanners, or a transmission scan with ^{68}Ge or ^{137}Cs rotating rod sources for PET-only scanners. Images were reconstructed using scanner-specific algorithms, and then sent to the University of Michigan, where they were reviewed for artefacts, de-identified, and transmitted to the Laboratory of NeuroImaging (LONI) for storage. Further details are available in the ADNI PET technical procedures manual ([ADNI PET Core, 2005](#)). The 315 baseline FDG-PET scans were downloaded from the LONI image data archive in their original DICOM or ECAT format. Only 309 images were suitable for further processing, since three lasted for less than 30 minutes, and three had missing timeframe information in their ECAT headers (see Appendix B).

4.3.2 FDG-PET pre-processing

The 309 remaining baseline FDG-PET images were converted to NIfTI format using (X)MedCon (<http://xmedcon.sourceforge.net>). Each image was examined for major artefacts, and its orientation adjusted if necessary. In six images, the positioning of the subject in the scanner was such that the cerebellum, or frontal cortex in one case, was partially excluded from the field of view (see Appendix B). These images were excluded from further processing. Since FDG-PET acquisition was performed according to one of three protocols (30-minute static, 30-minute dynamic, 60-minute dynamic), standardisation was necessary before the images could be com-

pared. Dynamic scans acquired over 30-60 minutes were corrected for patient motion by using tools from the Image Registration Toolkit (IRTK; <http://www.doc.ic.ac.uk/~dr/software>) to register each of the subsequent frames rigidly to the image's first frame. Normalised mutual information was used as the similarity criterion for the registration, and transformations were performed using linear interpolation. The co-registered frames were then averaged to produce a single 30-minute static image. For the 60-minute quantitative scans, the final six 5-minute frames were extracted, and co-registered and averaged into a static image in the same way.

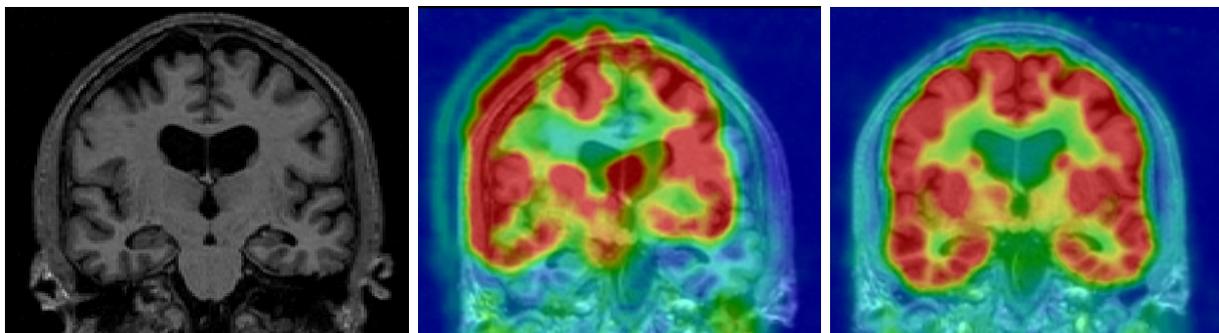
4.3.3 ADNI MRI acquisition and pre-processing

Baseline T1-weighted 1.5 T MRI scans for all 303 subjects with corresponding pre-processed baseline FDG-PET images were available. MRI and FDG-PET scans had been acquired within 30 ± 16 days of each other. The baseline MR images were downloaded from the LONI image data archive in NIfTI format. The scans had been acquired according to a standard protocol (Jack Jr. et al., 2008a) involving two scans per subject acquired using Siemens, GE and Philips MR scanners. Further details are available in the ADNI MRI technical procedures manual (ADNI MRI Core, 2005). Of the two images acquired per subject, the ADNI quality assurance team selected the best image for pre-processing based on the presence and severity of common image artefacts, as well as general image quality. Pre-processing involved the application of a scanner-specific correction for gradient non-linearity distortion (Gradwarp; Jovicich et al. (2006)), a correction for image intensity non-uniformity (B1; Jack Jr. et al. (2008a)), and a histogram peak sharpening algorithm for bias field correction (N3; Sled et al. (1998)). Only the N3 step was necessary for images acquired on Philips scanners, since B1 correction was already implemented and their gradient systems tended to be linear (Jack Jr. et al., 2008a).

4.3.4 Co-registration of FDG-PET with MRI

Each of the 303 pre-processed FDG-PET images was co-registered with its corresponding pre-processed native space MRI, again using tools from IRTK. Rigid registration parameters were

estimated, and used as a starting point for the estimation of a twelve-parameter affine registration, using normalised mutual information as the similarity criterion. The resulting affine registration parameters were applied to transform the FDG-PET image into the higher resolution space of its corresponding MRI using linear interpolation. An affine transformation was preferred over a rigid one because it can account for any scaling or voxel size errors which may remain after phantom correction of the MRI (Clarkson et al., 2009). There is currently no established method with which to assess the accuracy of PET-MRI co-registration. The transformed images were therefore interactively visually reviewed before further processing. The image registration process is illustrated in Figure 4.4.



(a) Native space MRI (b) Native space PET (c) MRI-space PET

Figure 4.4: Illustration of the PET-MRI co-registration process, showing (a) the native space MR image overlaid with (b) the native space FDG-PET image, and (c) the MRI-space FDG-PET image.

4.4 MRI anatomical segmentation

Automatic whole-brain segmentations of all available baseline ADNI MR images had been prepared in native MRI space as part of the work described in Heckemann et al. (2011), in which full details of both the segmentation procedure and subsequent morphometric analyses are presented. Each MRI was segmented into 83 anatomically defined regions using multi-atlas propagation with enhanced registration (MAPER) (Heckemann et al., 2010). This is a refined version of a previously validated segmentation approach (Heckemann et al., 2006), whose accuracy was shown to be only slightly inferior to that of expert manual segmentation. MAPER

was the first automatic whole-brain multi-region segmentation method shown to produce robust results in subjects with anatomical variations often seen in neurodegenerative disease, such as ventricular enlargement (Heckemann et al., 2010). MAPER is an atlas-based method, as described in Section 2.3.3, and the required atlas data consisted of the 30 manually segmented MRI described in Section 2.3.2. The additional image processing steps, segmentation procedure, and tissue class masking are described in the following subsections.

4.4.1 Additional image processing

Additional image processing was applied for brain extraction and tissue classification. For brain extraction, binary masks covering intracranial white and grey matter were available as the starting point. These had been generated as part of a separate project using MIDAS, a semi-automatic procedure that is described elsewhere (Freeborough et al., 1997). Each mask was extended to cover the intracranial region generously, and FSL FAST (<http://www.fmrib.ox.ac.uk/fsl>) applied to identify CSF within the pre-masked region. The original MIDAS mask was then extended by the resulting CSF to obtain a complete intracranial mask, as illustrated in Figure 4.5.

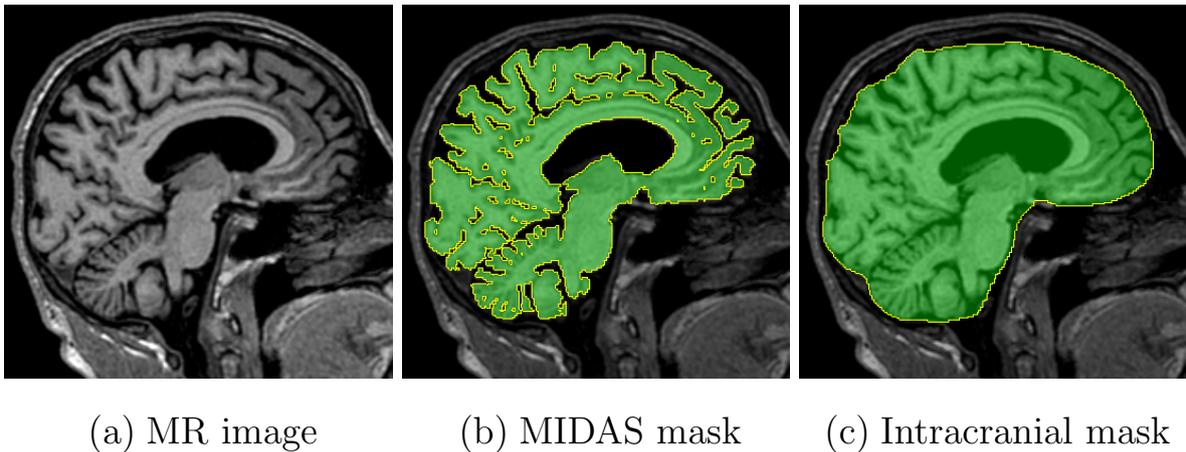


Figure 4.5: Illustration of the brain masking process in which (a) the MRI is processed using MIDAS, resulting in (b) a binary mask covering the intracranial white and grey matter. This mask is further processed to generate (c) the final intracranial mask required for segmentation.

One image for which a MIDAS mask was not available was excluded from further processing (see Appendix B). For the remaining 302 images, individual tissue probability maps for CSF,

grey matter and white matter were obtained using FSL FAST, as illustrated in Figure 4.6. These were combined into a single multi-spectral volume.

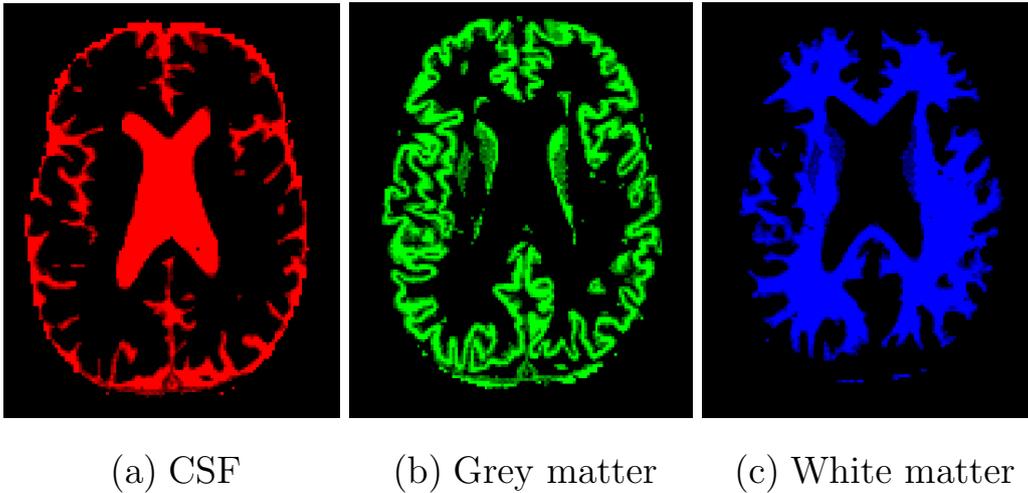


Figure 4.6: Illustration of the tissue classification results, showing individual tissue probability maps for (a) CSF, (b) grey matter, and (c) white matter.

4.4.2 Segmentation procedure

MAPER follows the multi-atlas segmentation approach described in Section 2.3.3. A target MRI was paired with each atlas to generate a set of atlas-based segmentations, which were then fused. The steps used to align the target MRI with a single atlas are summarised in Table 4.2.

Transformation model	Image data	Similarity criterion
Global (rigid)	mstprob	CC
Global (affine)	mstprob	CC
Coarse nonrigid (20 mm)	mstprob	CC
Detailed nonrigid (10, 5, 2.5 mm)	masked MRI	NMI

Table 4.2: MAPER steps for aligning a single atlas MRI with the target MRI. Image data consisted of the multi-spectral tissue probability map (mstprob), or MR intensities within the intracranial mask (masked MRI). Cross correlation (CC) or normalised mutual information (NMI) was used as the similarity criterion. For the nonrigid transformations, numbers indicate the control point spacings.

The transformation obtained from the final nonrigid registration step was applied to the atlas label volume using nearest-neighbour interpolation, resulting in a single atlas-based segmentation of the image in native MRI space. For a single target MRI, the registration and transformation

steps were performed using each atlas, producing 30 propagated label volumes in the space of the original image. The labels from these independent segmentations were combined using vote-rule decision fusion to produce a fused segmentation for the target image.

4.4.3 Tissue class masking

For FDG-PET image analysis, the grey matter portion within each cortical label is more relevant than the full label which covers both grey and white matter. Masked segmentations were therefore employed, in which all regions except ventricles, central structures, cerebellum and brainstem have been masked with a grey matter label, and the lateral ventricles with a CSF label. Typical examples of both the full and masked segmentations are shown in Figure 4.7.

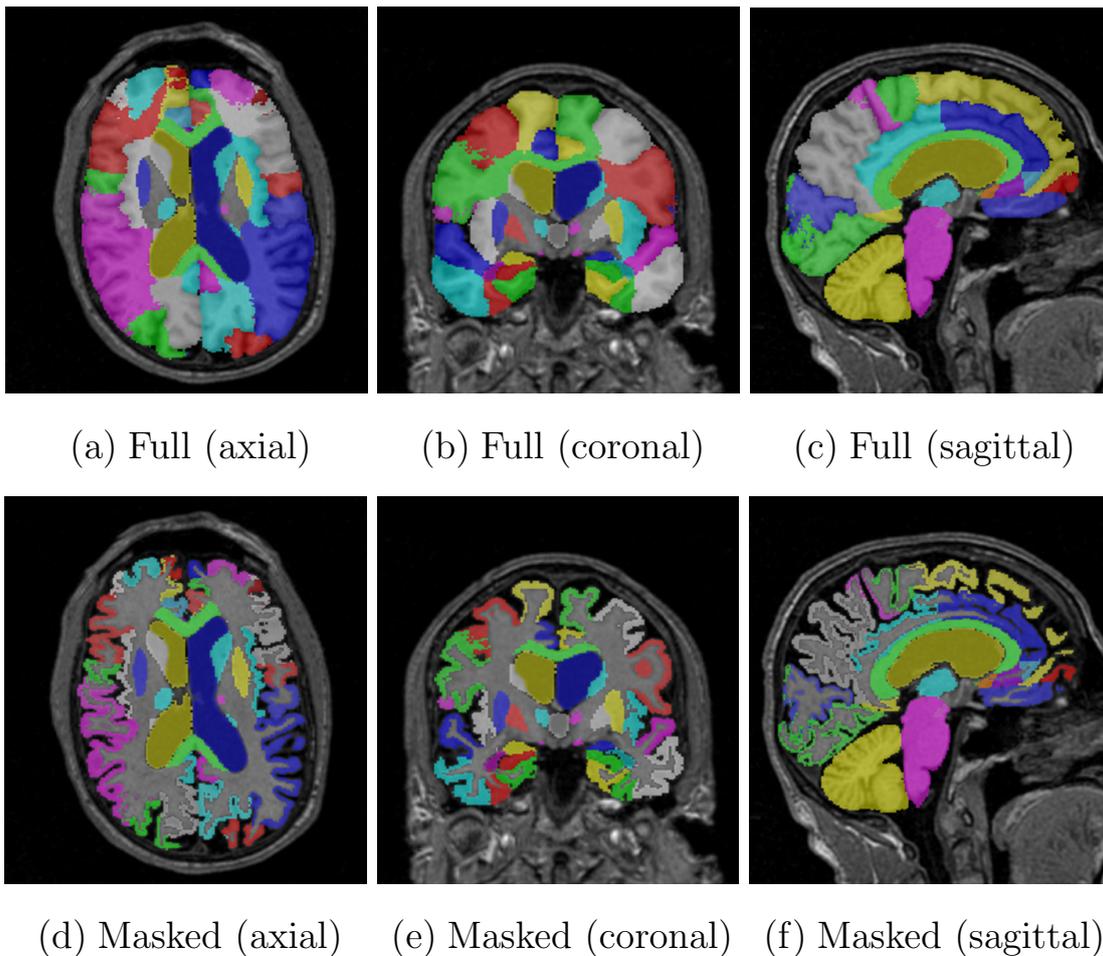


Figure 4.7: Typical examples of (a-c) a full segmentation, and (d-f) a masked segmentation shown overlaid onto the corresponding MR image.

4.5 FDG-PET intensity normalisation

When performing comparisons of FDG-PET images across subjects, intensity normalisation is required to account for inter-subject variability in overall radioactivity. Such variability can arise from both physiological differences in metabolic rate, as well as variations in, for example, the injected dose or time allowed for tracer uptake before scanning. An additional incentive for normalisation is provided by reports showing that relative values of the CMRgl are less variable than absolute values in both HC and AD patients (Wang et al., 1994; Minoshima et al., 1995).

FDG-PET intensity normalisation is often performed relative to the CGM. However, due to the nature of the disease, both MCI and AD patients have a lower CMRgl than HC across the whole brain. CGM normalisation therefore artificially scales up the CMRgl from patients, while scaling down those from HC. This results in under-estimation of the relative hypometabolism in patients compared to HC (Yakushev et al., 2008), as well as the semblance of hypermetabolism in relatively preserved regions of the brain, including the cerebellum, brainstem, basal ganglia, and sensorimotor cortex (Herholz et al., 2002). Recent work suggests that improved group discrimination can be achieved by using these relatively preserved regions for FDG-PET image intensity normalisation (Yakushev et al., 2009; Borghammer et al., 2009). Although normalisation relative to regions such as the cerebellum or sensorimotor cortex has been used in other previous studies, a key feature of the method proposed in Yakushev et al. (2009) is that preserved regions are extracted directly from the image data. The following subsections present results of investigations into FDG-PET intensity normalisation methods.

4.5.1 Additional image processing

FDG-PET image intensity normalisation is investigated using the SPM software described in Section 2.4. To perform voxel-wise comparisons, the images first had to be aligned with the MNI template described in Section 2.3.2. Each pre-processed ADNI MRI was linearly and non-linearly deformed (Ashburner and Friston, 2005) to the MNI template using the “Segment” module of SPM5. The resulting transformation parameters were applied to the corresponding

MRI-space FDG-PET image using tri-linear interpolation. This process is illustrated in Figure 4.8. Full inclusion of the cerebellum was ensured by extending the bounding box in the z-direction to cover -75 mm to +85 mm from the position of the anterior commissure in MNI space. Upon visual inspection, three images for which SPM Segment was unable to provide satisfactory transformations were excluded from further processing (see Appendix B).

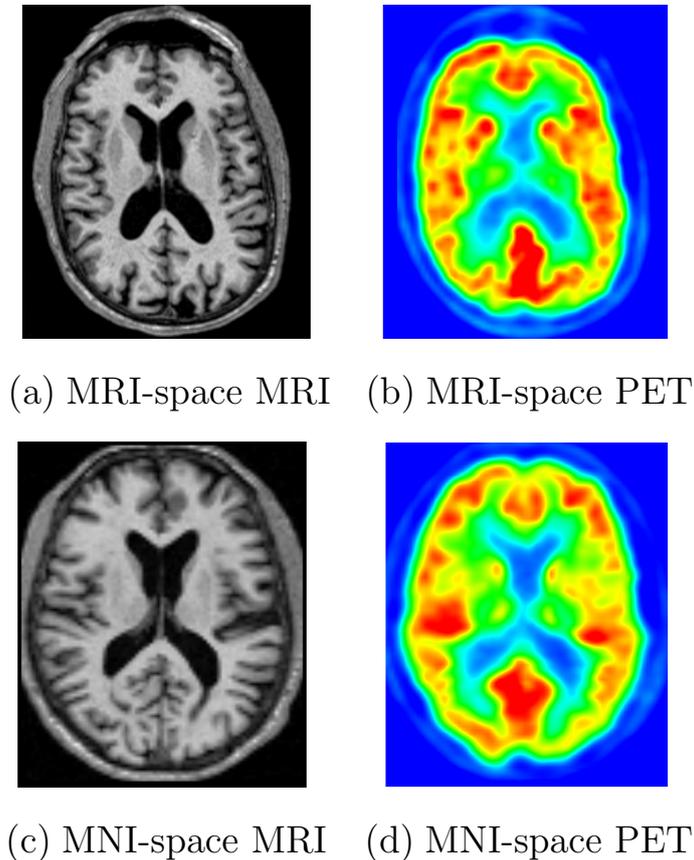


Figure 4.8: Example images showing alignment of MRI and FDG-PET to the MNI template.

The MNI-space FDG-PET images were smoothed to increase their signal-to-noise ratios, as well as compensate for any residual inter-subject variability remaining after the transformation to MNI space. Since the images had been acquired using one of 12 scanner models, they were first smoothed to a common isotropic spatial resolution of 8 mm full-width-at-half-maximum (FWHM) by applying scanner-specific Gaussian kernels. These kernels had been determined from the Hoffman brain phantom PET scans (Hoffman et al., 1990) acquired during the scanner certification process. The FWHM of these kernels are provided in Joshi et al. (2009), and summarised in Table 4.3. The FDG-PET images were then smoothed by an additional isotropic Gaussian kernel of 8 mm FWHM. The smoothing process is illustrated in Figure 4.9.

Scanner		FWHM	
Manufacturer	Model	In-plane (mm)	Axial (mm)
GE	Advance	5	3
GE	Discovery LS	-	-
GE	Discovery RX	5	4
GE	Discovery ST	4	3
Philips	Allegro	-	-
Philips	Gemini	3	3
Philips	Gemini GXL	-	-
Philips	Gemini TF	-	-
Siemens	Accel	2	3
Siemens	BioGraph	-	-
Siemens	Exact	-	-
Siemens	HR+	5	5

Table 4.3: In-plane and axial FWHM of the scanner-specific Gaussian kernels required to obtain an isotropic spatial resolution of 8 mm FWHM for the 12 ADNI PET scanner models from which images used in this work were acquired. Scanner models for which no FWHM values are provided did not require smoothing to reach the common lowest spatial resolution.

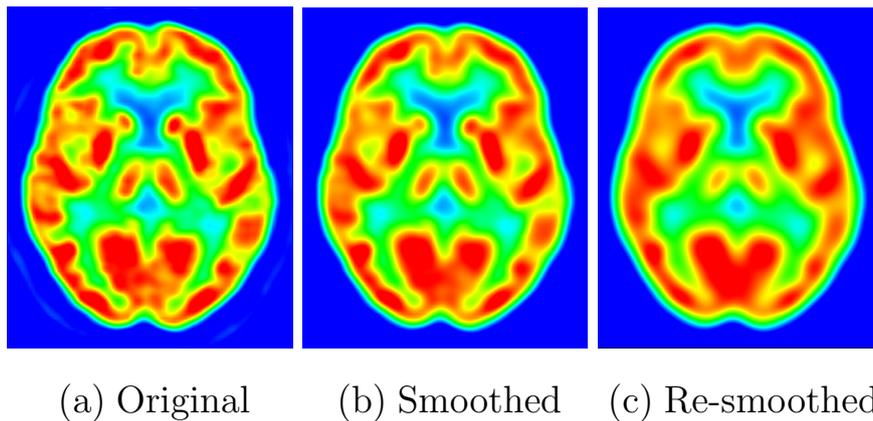


Figure 4.9: Typical example images illustrating the smoothing process, in which (a) the original MNI-space FDG-PET image is (b) smoothed with a scanner-specific Gaussian kernel, followed by (c) smoothing with a further isotropic Gaussian kernel.

4.5.2 Cerebral global mean intensity normalisation

The smoothed MNI-space FDG-PET images were first used to replicate the findings reported in [Langbaum et al. \(2009\)](#), in which SPM analyses were performed on ADNI-processed images. The aim was to verify that the processing applied as part of this work was comparable with that performed by the ADNI PET core. The ADNI-processed FDG-PET images were not used in this work because the intermediate re-orientation of each image, such that its anterior-posterior

axis was parallel to the AC-PC line, was not desirable for region-based analyses performed in native MRI space. The incorporation of unnecessary re-orientation steps would have resulted in degraded image quality due to the additional interpolation.

The SPM methods described in Section 2.4 were applied to identify voxels for which the FDG-PET signal intensities were significantly reduced in AD patients compared with HC. The two-sample t-test factorial design was applied, and inclusion of voxels outside the brain was avoided using a combination of relative threshold, explicit, and implicit masking. For relative threshold masking, only voxels for which all images exceeded 80% of their mean value were retained for analysis. The explicit mask specified was the SPM brain mask, which covers white and grey matter in MNI space. The implicit masking removed any voxels with “NaN” (not a number) values. The images were normalised for inter-subject global variations using proportional scaling to the CGM. The voxels identified as significantly ($p < 0.001$, uncorrected) reduced in AD patients compared with HC are shown in Figure 4.10.

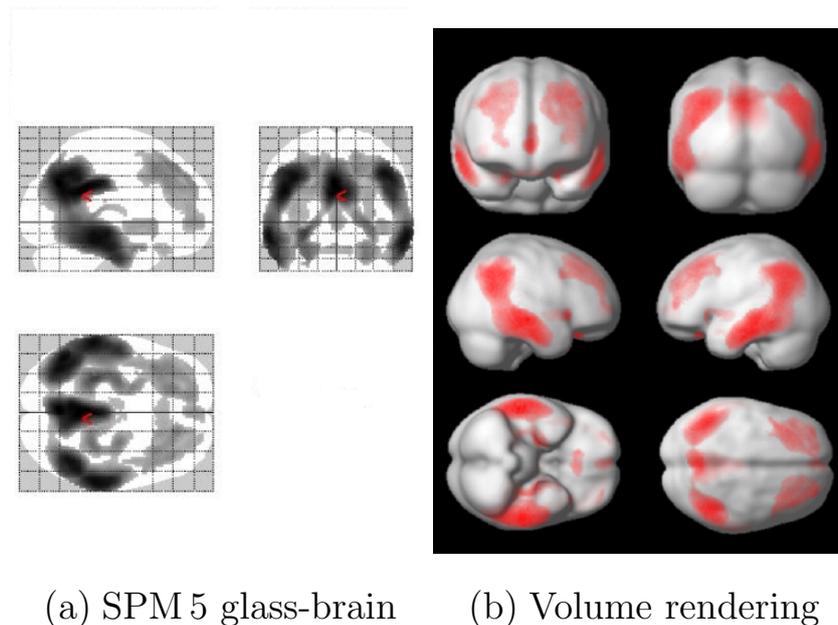


Figure 4.10: Voxels for which the FDG-PET signal intensity was significantly reduced in AD patients compared with HC when using CGM normalisation shown as (a) maximum intensity projections onto the SPM 5 glass-brain, and (b) a volume rendering. Red arrows on the glass-brain indicate the location of the voxel with the highest t-value ($t = 8.59$).

The same methods were applied to identify voxels with significantly ($p < 0.001$, uncorrected) reduced signal intensities in MCI patients compared with HC. These are shown in Figure 4.11.

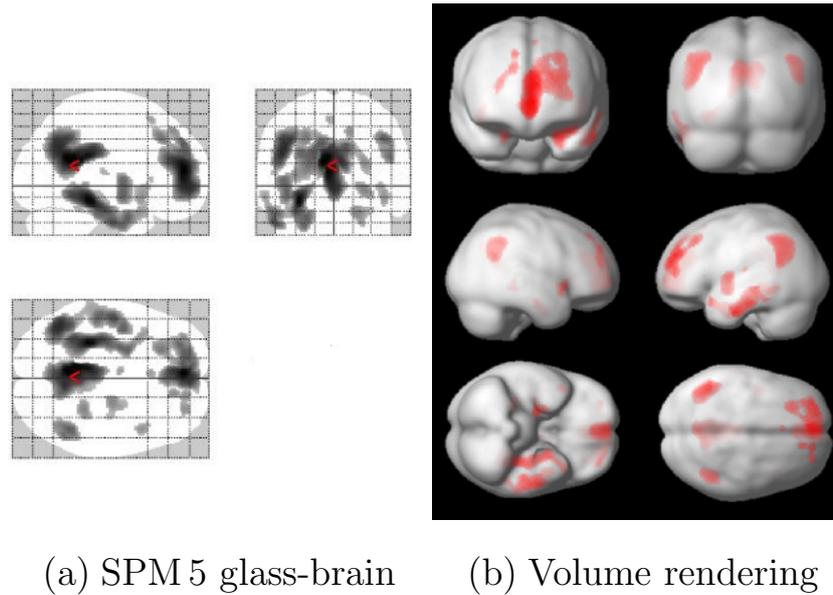


Figure 4.11: Voxels for which the FDG-PET signal intensity was significantly reduced in MCI patients compared with HC when using CGM normalisation shown as (a) maximum intensity projections onto the SPM 5 glass-brain, and (b) a volume rendering. Red arrows on the glass-brain indicate the location of the voxel with the highest t-value ($t = 5.57$).

The results of these two SPM analyses are comparable with those reported in [Langbaum et al. \(2009\)](#). Both AD and MCI patients display significantly reduced FDG-PET signal intensities in brain regions preferentially affected by AD, including the parietal and posterior temporal lobes, and posterior cingulate cortices. The reductions are of greater magnitude and spatial extent in AD patients compared with MCI patients.

4.5.3 Reference cluster intensity normalisation

The data-driven intensity normalisation approach proposed in [Yakushev et al. \(2009\)](#) was next investigated. Regions of the brain that are relatively preserved throughout AD were extracted from the ADNI FDG-PET images by identifying voxels in which there was a semblance of hypermetabolism in patients compared with HC. These regions form a so-called reference cluster which can be used for intensity normalisation. Voxels for which the FDG-PET signal intensities were significantly increased in the whole patient group (both AD and MCI patients) compared with the HC group were identified using the SPM methods described in Section 4.5.2. The

voxels identified as significantly ($p < 0.05$, FWE corrected for multiple comparisons at the voxel-level) increased are shown in Figure 4.12 (a). From these, a cluster was defined which contained the voxel with the highest t-value in the search space, as shown in Figure 4.12 (b).

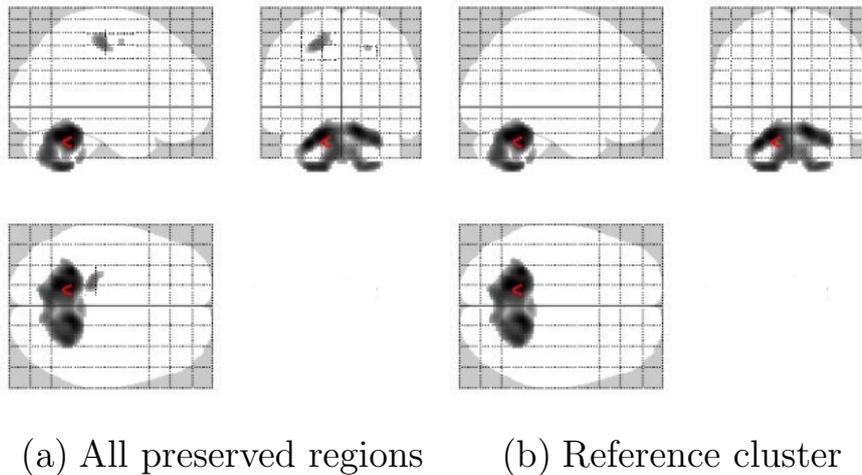


Figure 4.12: Voxels for which the FDG-PET signal intensity was significantly increased in AD and MCI patients compared with HC shown as maximum intensity projections onto the SPM 5 glass-brain. Both (a) all voxels, and (b) only voxels in the selected cluster are shown. Red arrows indicate the location of the voxel with the highest t-value ($t = 5.98$).

The results of this SPM analysis are comparable with those presented in [Yakushev et al. \(2009\)](#), with significant clusters observed bilaterally in regions of the cerebellum, brainstem and primary sensorimotor systems. The spatial extent of the significant regions is greater in both the cerebellum and sensorimotor cortex, possibly reflecting the potential for detecting more subtle changes when using such a large cohort. Apparent hypermetabolism was additionally observed in the putamen in [Yakushev et al. \(2009\)](#), which does not show up as significant in this study. This is most likely due to the different demographic characteristics of the two study populations. For example, the study reported in [Yakushev et al. \(2009\)](#) included no sMCI patients. The reference cluster identified using the ADNI FDG-PET images is located mostly in the cerebellum, but also extends to a small region of the brainstem, consistent with the findings of [Yakushev et al. \(2009\)](#).

Having defined the reference cluster for normalisation, the methods described in Section 4.5.2 were again applied to identify voxels for which the FDG-PET signal intensities were significantly reduced in AD patients compared with HC. Inter-subject global variations were now

accounted for by proportional scaling to the reference cluster mean. The voxels identified as significantly ($p < 0.001$, uncorrected) reduced in AD patients compared with HC, and MCI patients compared with HC are shown in Figures 4.13 and 4.14, respectively.

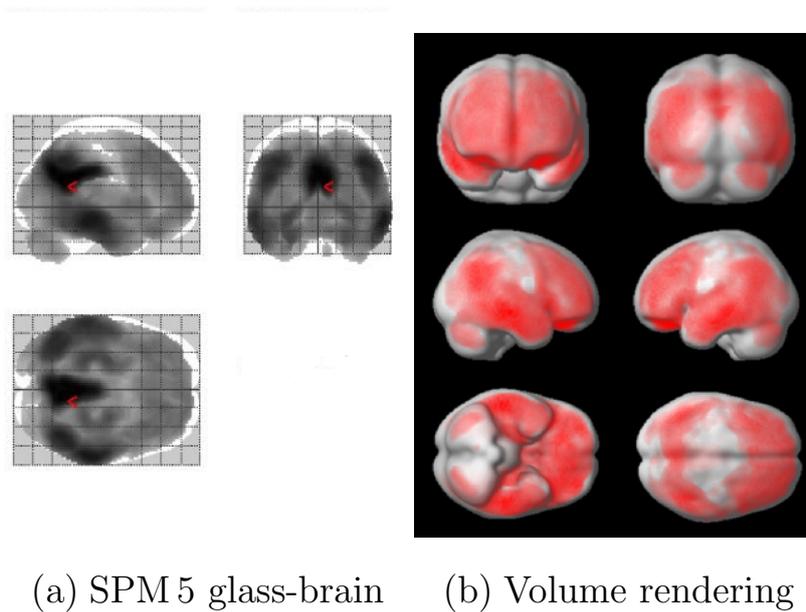


Figure 4.13: Voxels for which the FDG-PET signal intensity was significantly reduced in AD patients compared with HC when using reference cluster normalisation shown as (a) maximum intensity projections onto the SPM 5 glass-brain, and (b) a volume rendering. Red arrows on the glass-brain indicate the location of the voxel with the highest t-value ($t = 11.96$).

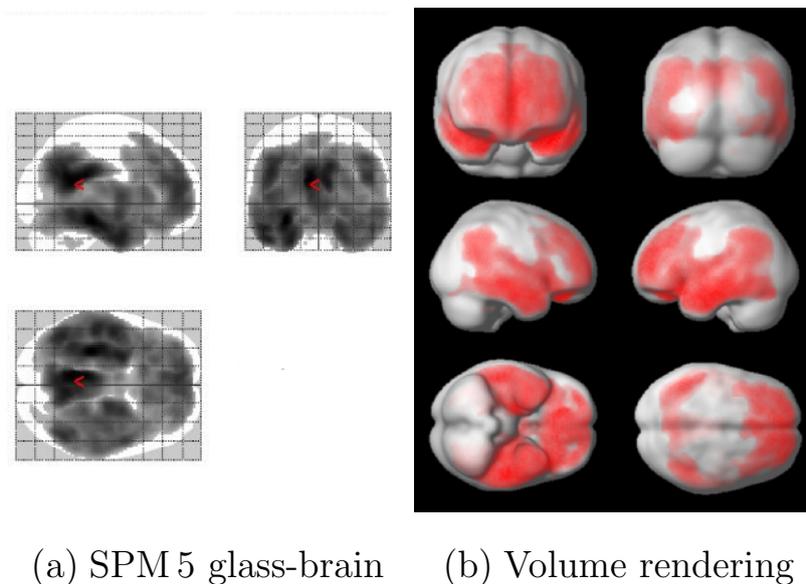


Figure 4.14: Voxels for which the FDG-PET signal intensity was significantly reduced in MCI patients compared with HC when using reference cluster normalisation shown as (a) maximum intensity projections onto the SPM 5 glass-brain, and (b) a volume rendering. Red arrows on the glass-brain indicate the location of the voxel with the highest t-value ($t = 7.24$).

The results of these two SPM analyses demonstrate that the reference cluster intensity normalisation method proposed in [Yakushev et al. \(2009\)](#) provides superior discrimination of the ADNI AD and MCI patients from the HC, compared with CGM normalisation. Both AD and MCI patients display significantly reduced FDG-PET signal intensities across most of the brain, and at higher significance levels than were observed for the CGM-normalised images.

The use of a normalisation region derived directly from the imaging data may introduce bias into classification studies. The validity of using the reference cluster derived in [Yakushev et al. \(2009\)](#) for normalisation of the ADNI FDG-PET images was therefore assessed. A MNI space image of the cluster used in [Yakushev et al. \(2009\)](#) was obtained from the author, and the FDG-PET signal intensity per mm^3 was determined from each image in both this independently-derived cluster, and in the ADNI-derived cluster. The intraclass correlation coefficient (ICC) between the two sets of values was calculated using the “one-way random” model implemented in SPSS. The “single measures” ICC was 0.95, a value high enough to suggest that the areas of the brain identified are reliably preserved across early AD and MCI.

4.6 Multi-region image analysis

Many image-based classification methods rely on voxel-wise comparisons in which each voxel is considered as a feature that may be exploited for group discrimination. As well as requiring that all images be transformed into a common space, these methods can fail to account for the spatial structure of the imaging data which results from correlations between neighbouring voxels. A feature selection process may therefore be required which, as well as reducing the dimensionality of the feature-space, requires that selected voxels form spatially contiguous regions, thus exploiting the underlying spatial structure. The multi-region analysis used in this work instead makes use of subject-specific anatomical segmentations into 83 regions that are automatically generated in the native space of each subject, rather than in the space of a single reference image. Exploiting the spatial structure of the imaging data in this way results in a greatly reduced number of available features for group discrimination and classification.

4.6.1 Region-based feature extraction

Each of the MRI-space FDG-PET images described in Section 4.3.4 was overlaid with its corresponding masked anatomical segmentation, and the FDG-PET signal intensity per mm^3 determined for all 83 regions. Global inter-subject variations were accounted for by normalisation to the signal intensity per mm^3 in the independently-derived reference cluster described in Section 4.5.3. This cluster was provided in MNI space, and the inverse transformation parameters obtained using SPM Segment, as described in Section 4.5.1, were used to transform the cluster into the native MRI space of each subject using tri-linear interpolation. Examples of the images required for regional FDG-PET feature extraction are shown in Figure 4.15.

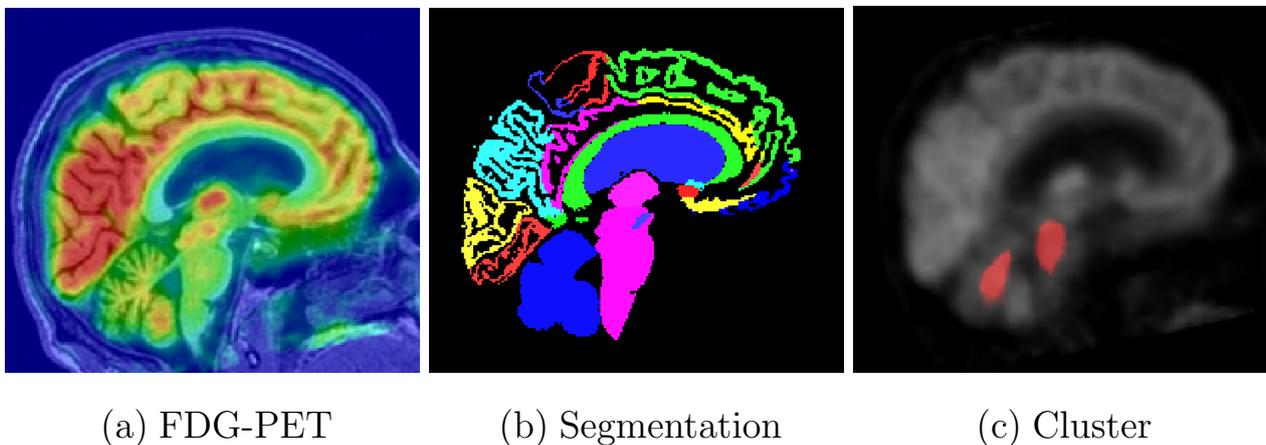


Figure 4.15: Typical examples of the images required for regional FDG-PET feature extraction, showing (a) the FDG-PET image overlaid onto its corresponding MRI, (b) the masked anatomical segmentation, and (c) the independently-derived normalisation cluster overlaid onto the FDG-PET image. All images are shown in native MRI space.

4.6.2 Region-based group differences

Boxplots based on the normalised signal intensity per mm^3 in two regions known to be affected in AD are shown in Figure 4.16. Comparisons by t-test were also performed between pairs of clinical groups to identify the regional features giving significant differences ($p < 0.01$, uncorrected for multiple comparisons). Regional t-values for comparisons between AD patients and HC, as well as between MCI patients and HC are shown in Figure 4.17.

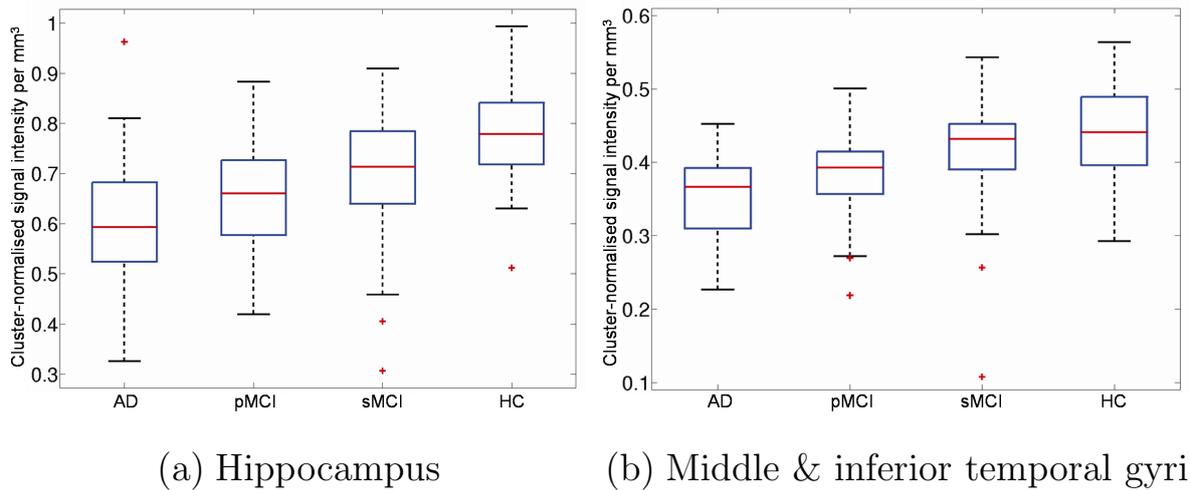


Figure 4.16: Boxplots showing group differences based on the normalised signal intensity per mm^3 in the (a) hippocampus and (b) middle and inferior temporal gyri. In each boxplot, the central red line represents the median, the edges of the blue box represent the 25th and 75th percentiles, and the black whiskers extend to the most extreme data points not considered outliers. Outliers are plotted individually in red for points outside of $\pm 1.5 \times$ interquartile range.

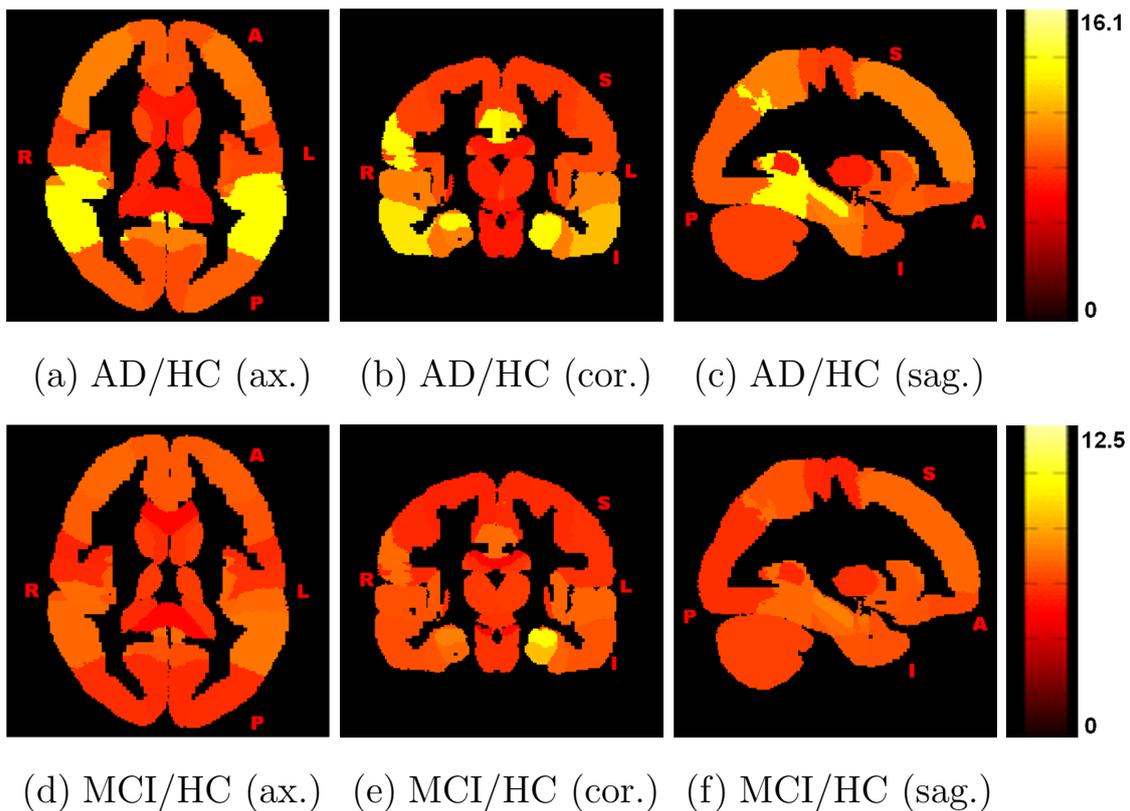


Figure 4.17: Regional t-values for comparisons between (a-c) AD patients and HC and (d-f) MCI patients and HC, superimposed onto a maximum probability brain atlas masked according to the same procedure as the anatomical segmentations. Axial (ax.), coronal (cor.) and sagittal (sag.) views are shown. A: anterior, P: posterior, S: superior, I: inferior, L: left, R: right.

The majority of regions differ significantly between AD patients and HC (74/83 regions), as well as between MCI patients and HC (54/83 regions). The magnitude and spatial extent of the significant regions is greater between AD patients and HC. The ten most significantly different regions between AD patients and HC are the bilateral posterior cingulate gyrus, hippocampus, posterior temporal lobe and parietal lobe, and left parahippocampal gyrus and middle and inferior temporal gyri. The ten most significantly different regions between MCI patients and HC are the bilateral hippocampus and parietal lobe, and left parahippocampal gyrus, amygdala, posterior temporal lobe, posterior cingulate gyrus, insula, and pre-subgenual frontal cortex. These findings are consistent with the voxel-based group differences described in Section 4.5.3.

4.7 Classification experiments

The potential use of multi-region FDG-PET information for image-based classification of the ADNI participants was next investigated. Other relevant classification studies based on imaging data acquired at baseline have been described in Section 3.4. The majority of research surrounding the classification of AD and MCI focuses on the discrimination of AD patients from HC and, in fewer cases, MCI patients from HC. In this work, all eight possible binary classification experiments (AD/HC, MCI/HC, pMCI/HC, sMCI/HC, AD/MCI, AD/pMCI, AD/sMCI, pMCI/sMCI) were performed, of which the potential to discriminate between pMCI and sMCI patients is the most interesting in terms of monitoring the progression of AD.

4.7.1 Methods

The normalised signal intensities per mm^3 in each of the 83 anatomically defined regions formed a feature vector for each of the 287 ADNI participants (71 AD, 62 pMCI, 85 sMCI, 69 HC). Two binary classifiers were investigated: a SVM classifier, and an algorithm which combines feature extraction using AdaBoost with a linear discriminant function classifier (Ada-LDA). Robust estimates of the performance of the two classifiers were obtained via the stratified repeated random sampling approach described in Section 3.3.2. The mean classification accuracy, sensi-

tivity and specificity for pairs of clinical groups were evaluated over 1,000 runs, in which 75% of the subjects were randomly selected for training, with the remaining 25% used for testing.

SVM classifier

The SVM classifier was applied using LIBSVM, an integrated software for support vector classification (<http://www.csie.ntu.edu.tw/~cjlin/libsvm>). Since the data were unlikely to be linearly separable, the soft-margin SVM formulation described in Section 3.2.2 was used. This is the “C-SVC” option in LIBSVM (Chang and Lin, 2011). The radial basis function kernel described in Section 3.2.2 was selected to map the data into a higher dimensional space.

As part of the training process, it was necessary to optimise the penalty parameter C , and kernel width γ . This was achieved by performing a grid-search using five-fold cross-validation. Various (C, γ) pairs were considered, and the pair resulting in the highest cross-validation accuracy was selected. The SVM classifier was then trained using the full set of training data, before having its performance assessed on the test data.

Ada-LDA classifier

The Ada-LDA classifier combines feature extraction using AdaBoost, described in Section 3.2.3, with the linear discriminant function described in Section 3.2.1. Although AdaBoost itself may be applied as a classifier, in this work it was employed solely as a feature selection tool, iteratively selecting the best unique features to pass to the linear discriminant function. A particular strength of this feature selection method is its consideration of previously selected features in the determination of the next, since features which individually provide good group discrimination may not necessarily work well in combination.

As part of the training process, it was necessary to determine the optimal number of features to pass to the linear discriminant function. This was again achieved using five-fold cross-validation, with the number of features resulting in the highest cross-validation accuracy selected. The

final classifier was then trained using the selected number of features from the full set of training data, before having its performance assessed on the test data.

4.7.2 Results

Classification results are presented in Table 4.4 for both the SVM and Ada-LDA classifiers. For each clinical group pair, unpaired t-tests between the distributions of classification results obtained from the 1,000 leave-25%-out repetitions were performed to assess the significance of differences in performance between the two classifiers. The SVM classifier achieves a significantly ($p < 0.05$) higher accuracy than the Ada-LDA classifier for the MCI/HC experiment, but no other differences in accuracy reach significance. All accuracies are significantly different from chance, as assessed using the permutation testing procedure described in Section 3.3.2.

	SVM			Ada-LDA		
	acc. (%)	sens. (%)	spec. (%)	acc. (%)	sens. (%)	spec. (%)
AD/HC	81.6 (0.2)	82.7 (0.3)	80.4 (0.3)	83.3 (0.2)	80.5 (0.3)	86.2 (0.3)
MCI/HC	70.2* (0.2)	73.8 (0.3)	62.3 (0.6)	63.6 (0.2)	58.5 (0.3)	74.4 (0.3)
pMCI/HC	71.8 (0.2)	71.8 (0.4)	71.8 (0.3)	73.1 (0.2)	72.8 (0.4)	73.4 (0.3)
sMCI/HC	63.7 (0.2)	62.6 (0.3)	63.0 (0.4)	59.6 (0.2)	58.1 (0.3)	61.3 (0.4)
AD/MCI	68.2 (0.2)	58.3 (0.6)	73.0 (0.3)	66.6 (0.2)	70.3 (0.3)	64.8 (0.3)
AD/pMCI	57.1 (0.2)	58.3 (0.6)	50.7 (0.5)	59.1 (0.3)	55.8 (0.4)	62.8 (0.4)
AD/sMCI	74.2 (0.2)	72.1 (0.3)	76.0 (0.3)	72.5 (0.2)	70.8 (0.3)	73.8 (0.3)
pMCI/sMCI	56.4 (0.2)	51.6 (0.6)	60.1 (0.4)	57.6 (0.2)	59.5 (0.4)	56.2 (0.3)

Table 4.4: Classification accuracy (acc.), sensitivity (sens.), and specificity (spec.) for the SVM and Ada-LDA classifiers, expressed as mean (standard error) over the 1,000 leave-25%-out repetitions. The single case where the difference in classifier performance reaches significance ($p < 0.05$) is indicated by an asterisk (*).

While the SVM classifier makes use of all 83 features, the Ada-LDA classifier instead employs feature selection to suppress redundancies. The first selected feature is maximally able to discriminate the clinical groups. For AD/HC, the features most commonly selected first across the 1,000 repetitions were the hippocampus and posterior cingulate cortices. The parahippocampal gyrus was most commonly selected first for MCI/HC and pMCI/HC, and the hippocampus for sMCI/HC. The middle and inferior temporal gyri were most commonly selected first for AD/MCI, AD/pMCI, and pMCI/sMCI, and the posterior temporal lobe for AD/sMCI.

4.7.3 Discussion

Multi-region features extracted from baseline FDG-PET images of ADNI participants have been used to achieve classification results comparable with other studies based on FDG-PET data, as well as with those obtained using data from MRI. For the discrimination of AD patients from HC, both the SVM and Ada-LDA classifiers achieve results that are in line with the studies reviewed in Section 3.4. Classification results may be converging on a glass ceiling for this task, since diagnostic consensus criteria themselves have an accuracy of around 90% (Ranginwala et al., 2008). Although higher accuracies have been reported in some studies, these tend to be in smaller, or more selective, patient cohorts. For FDG-PET in particular, it is important to consider the further confounding factor that approximately 10% of the ADNI AD patients have a pattern of glucose metabolism that is more consistent with frontotemporal dementia than with AD (Thiele et al., 2009; Jagust et al., 2010).

With the SVM classifier, good discrimination is achieved between MCI patients and HC (accuracy 70%), and slightly less good discrimination between AD and MCI patients (accuracy 68%). These results are again comparable with those of other published studies, such as Colliot et al. (2008). The Ada-LDA classifier performs less well, particularly in distinguishing MCI patients from HC (accuracy 64%). This is possibly because the heterogeneous nature of the MCI group means that a linear classifier is not optimal, even if it is applied to only a selected set of features.

In discussing the remaining comparisons, in which pMCI and sMCI patients are considered separately, it is important to note that clinical follow-up data are still being acquired for the ADNI participants. Subjects currently in the sMCI group may yet convert to AD. Even the three years of clinical follow-up that will eventually be available for these subjects is likely to be insufficient to allow the identification of all those who will develop AD in the future. Progression from MCI to AD is reported to occur at a rate of 10-15% per year (Petersen et al., 1999), with up to 80% of MCI patients developing AD over a six year period (Petersen, 2004).

Following sub-division of the MCI group, both classifiers achieved good discrimination between

pMCI patients and HC, as well as between AD and sMCI patients. However, neither classifier was able to achieve particularly good discrimination between sMCI patients and HC. The causes for memory impairment in sMCI patients are, perhaps, too heterogeneous in nature to be reliably distinguished from the changes associated with normal ageing. The relatively poor performance of both classifiers in distinguishing AD from pMCI patients is to be expected, since many of the changes associated with AD will have already begun in pMCI patients. The discrimination between pMCI and sMCI patients is also relatively poor with both classifiers, but significantly different from chance. This result corroborates other published studies, such as [Cuingnet et al. \(2011\)](#). The discrimination of pMCI from sMCI patients is the most clinically interesting, and subsequent studies will aim to improve on these results by incorporating follow-up FDG-PET data, or information obtained using other modalities.

The aim of this work was to investigate the utility of multi-region FDG-PET information acquired at baseline for image-based classification of AD and MCI patients. Two readily available classifiers which exhibit different behaviours were therefore applied. There was little appreciable difference in their performance, but the Ada-LDA classifier provides valuable information about the features that best distinguish between various clinical groups. The features selected first by the Ada-LDA classifier all have highly significant regional t-values, as shown in Section [4.6.2](#). AdaBoost was applied for feature selection, rather than a simple t-test, because the features selected by AdaBoost are complementary, in the sense that previously selected features are considered in the determination of the next.

The findings presented here support the use of FDG-PET, in conjunction with other neuroimaging methods, clinical assessments and measures of non-imaging biomarkers, as a tool for the early diagnosis of AD and for monitoring its progression. However, an important consideration of the described regional FDG-PET analysis approach is its requirement for MR imaging data. Structural imaging, either with MRI or CT, is routinely used in clinical practice to exclude focal brain lesions that could lead to a clinical picture mimicking AD. Both MRI and FDG-PET are mentioned in the revised AD diagnostic criteria ([McKhann et al., 2011](#); [Albert et al., 2011](#); [Sperling et al., 2011](#)) as providing potentially useful biomarkers. The recent development of hybrid MRI-PET technology means that the simultaneous acquisition of both modalities could

become a practical solution for dementia imaging in the future. For example, one such system has been approved for use in clinical practice in both Europe and the USA, and its clinical application in oncology has already been demonstrated (Drzezga et al., 2012). The requirement for MR data has the key benefit that regional volumes and volume changes are also available for each patient, and these data could potentially be combined with the FDG-PET information.

4.8 Conclusion

This chapter has presented a multi-region analysis of baseline FDG-PET imaging data from the ADNI study. Whole-brain segmentations into 83 anatomically defined regions were automatically generated in the native MRI space of each subject, and used to extract regional signal intensities from the corresponding FDG-PET images. Image intensity normalisation methods were investigated, and an independently-derived cluster of relatively preserved regions was used to normalise the FDG-PET images prior to region-based analyses of group differences and binary classification experiments. The classification results are comparable with other studies, but it is expected that they might be improved by the inclusion of additional data. The work presented in the following chapter will focus on the incorporation of follow-up FDG-PET data.

Chapter 5

Multi-region longitudinal FDG-PET for classification

Work in this chapter has, in part, been presented in:

K. R. Gray, R. Wolz, R. A. Heckemann, P. Aljabar, A. Hammers, D. Rueckert, for the Alzheimer's Disease Neuroimaging Initiative. Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease. Neuroimage, 60(1):221–229, 2012.

and:

K. R. Gray, R. Wolz, R. A. Heckemann, A. Hammers, and D. Rueckert. Classification of ADNI subjects based on longitudinal analysis of the hippocampal FDG-PET signal. (Abstract) Alzheimer's Association International Conference on Alzheimer's Disease (ICAD'10). Alzheimer's and Dementia, 6(4, Supplement):S288–S289, 2010.

5.1 Introduction

This chapter presents a regional analysis of longitudinal FDG-PET imaging data from the ADNI study. Both cross-sectional and longitudinal FDG-PET and MR imaging data were used by [Hinrichs et al. \(2011\)](#) in their investigations into the application of kernel combination methods for multi-modality AD classification. They observed that longitudinal FDG-PET

features performed relatively poorly in distinguishing AD patients from HC, compared with using the raw signal intensities at either timepoint. They suggested that two-year changes in FDG-PET signal intensity alone are not sufficient for image-based classification of individual patients. However, highly significant group differences between AD or MCI patients and HC have been reported by [Chen et al. \(2010\)](#) in their longitudinal analysis of 12-month metabolic declines in ADNI subjects. These two studies suggest that while longitudinal FDG-PET data alone may not be sufficient for AD classification, they may provide valuable complementary information which can enhance the results achievable using cross-sectional FDG-PET. This work investigates the value of combining cross-sectional and longitudinal multi-region FDG-PET information for classification. Whole-brain segmentations into 83 anatomically defined regions were automatically generated in the native MRI space of each subject for both baseline and 12-month images. These were used to extract regional signal intensities from the corresponding FDG-PET images, and their combined use for image-based AD classification was investigated. The effect on classification of the subjects' gender and age at time of scanning was additionally investigated.

The ADNI participants whose imaging data are used in this work are first described in Section 5.2, followed by the image acquisition and pre-processing steps in Section 5.3. Preliminary studies involving only the hippocampal region are presented in Section 5.4. These studies found that increases in classification accuracy could be achieved by combining cross-sectional and longitudinal FDG-PET features. The additional benefit of using multi-region information was then investigated. The extension of the multi-region MRI segmentation procedure described in Section 4.4 to the follow-up images is presented in Section 5.5, followed by multi-region analyses of group differences in Section 5.6, and binary classification experiments in Section 5.7. Longitudinal FDG-PET features were found to provide complementary information which enhanced classification performance when used in conjunction with cross-sectional features, resulting in classification results in line with the current state-of-the-art. This work demonstrates that information extracted from serial FDG-PET through regional analysis can accurately discriminate diagnostic groups, even at the early symptomatic stages of AD. This finding may be usefully applied for AD diagnosis and prognosis.

5.2 Imaging data

Clinical and imaging data used in this chapter were obtained from the ADNI database, which has been described in Section 4.2. Baseline and 12-month FDG-PET and 1.5 T MR images were available to download for 321 ADNI participants (as of March 2011). A number of images were excluded from the analyses presented in this chapter. These exclusions are summarised in Figure 5.1, and will be explained in further detail later in the chapter.

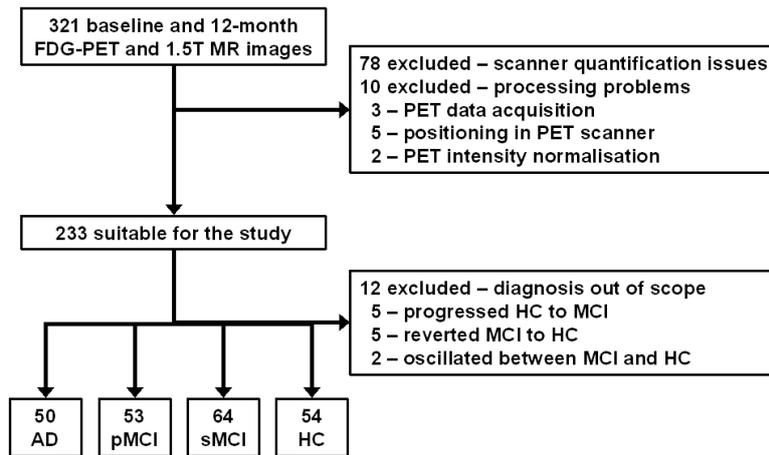


Figure 5.1: Summary of exclusions, showing that of the 321 baseline and 12-month FDG-PET and MRI available to download, 221 were suitable for the analyses presented in this chapter.

The MCI patients were divided into pMCI and sMCI based on changes in clinical status occurring over 19 ± 10 (range 6 – 48) months. ADNI subject identifiers for all excluded images are listed in Appendix B, along with the reasons for their exclusion. Groupwise characteristics are provided in Table 5.1 for the 221 subjects whose imaging data were used in this work.

	N(F)	Age (mean \pm sd)	Δ_T (months) (mean \pm sd)	CDR (%)			MMSE (mean \pm sd)	Δ_{MMSE} (mean \pm sd)
				0	0.5	1		
AD	50 (20)	75.8 \pm 6.1	11.5 \pm 0.8	0	36	64	23.5 \pm 2.0	-2.76 \pm 3.96
pMCI	53 (20)	75.1 \pm 7.1	11.5 \pm 1.1	0	100	0	26.7 \pm 1.7	-1.79 \pm 2.57
sMCI	64 (18)	76.6 \pm 7.1	11.8 \pm 0.8	0	100	0	27.5 \pm 1.7	0.05 \pm 1.79
HC	54 (18)	75.2 \pm 4.6	11.5 \pm 0.9	100	0	0	28.9 \pm 1.2	0.30 \pm 1.47

Table 5.1: Clinical and demographic information for the study population ($N = 221$). For each clinical group, the total number of subjects (N) and number of females (F) are shown, along with the average age, average time between baseline and 12-month FDG-PET scans (Δ_T), baseline CDR distribution, average baseline MMSE score, and average change in MMSE score over the 12-month follow-up period (Δ_{MMSE}). Standard deviation (sd) values are also provided.

The mean age at baseline and mean time between baseline and 12-month FDG-PET scans do not vary significantly ($p > 0.01$) on t-test between the clinical groups.

5.3 Image acquisition and pre-processing

The anatomical segmentations required for regional sampling were automatically generated in the native MRI space of each subject and timepoint (see Section 5.5). It was therefore necessary to co-register each FDG-PET image with its corresponding MRI. FDG-PET image intensity normalisation was performed using a cluster of relatively preserved regions that were provided in MNI space and then transformed into the native MRI space of each subject and timepoint. These image acquisition and pre-processing steps are summarised in Figure 4.2.

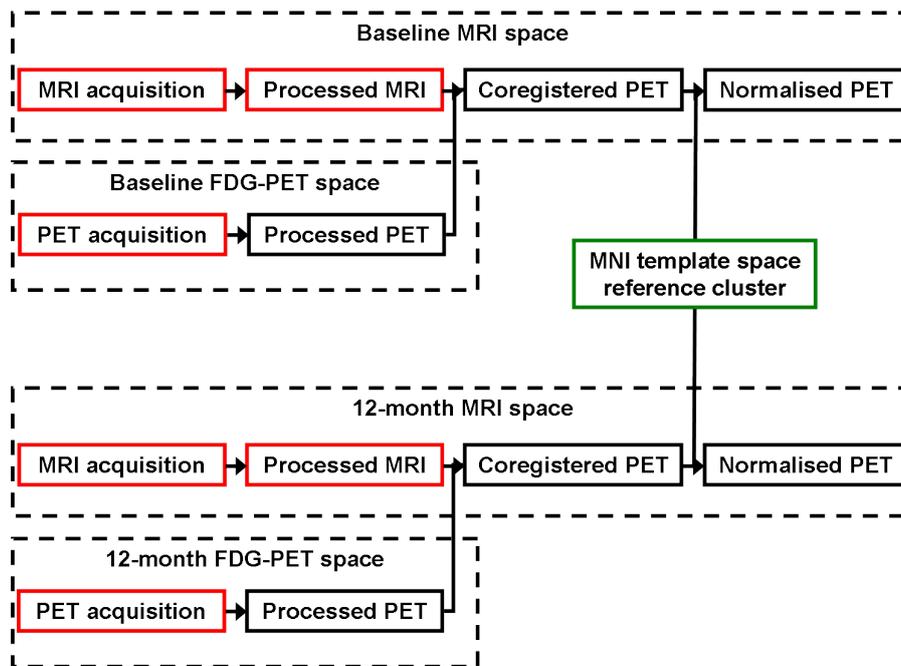


Figure 5.2: Summary of the FDG-PET and MR image acquisition and pre-processing steps. Those outlined in red were performed by ADNI investigators, while those outlined in black were performed as part of this work. The reference cluster outlined in green was provided in MNI space by [Yakushev et al. \(2009\)](#).

The FDG-PET images were acquired as described in Section 4.3.1. Baseline and 12-month FDG-PET images were available for 321 ADNI participants. For this work, all those acquired using either the Siemens HRRT or BioGraph HiRez scanners ($n = 78$) were excluded, again due to the scanner quantification issues described in Section 4.3.1. The 243 remaining baseline

FDG-PET scans were downloaded from the LONI image data archive in their original DICOM or ECAT format. These were pre-processed to produce 30-minute statics as described in Section 4.3.2. Three images were excluded due to missing timeframe information in their ECAT headers, and five because the positioning of the subject in the scanner was such that either the cerebellum or frontal cortex was partially excluded from the field of view (see Appendix B).

Baseline and 12-month T1-weighted 1.5 T MRI scans for all 235 subjects with corresponding pre-processed FDG-PET images were downloaded from the LONI image data archive in NIfTI format. These had been acquired and pre-processed as described in Section 4.3.3. For each subject and timepoint, the pre-processed FDG-PET image was affinely co-registered with the corresponding pre-processed MR image, and re-sampled to the higher resolution of the MRI. Registration was performed as described in Section 4.3.4.

FDG-PET image intensity normalisation was performed using the reference cluster method described in Section 4.5.3. The independently-derived cluster from Yakushev et al. (2009) was again used for normalisation to avoid the introduction of bias. This cluster was provided in MNI space, and transformed into the baseline and 12-month MRI space of each subject. The pre-processed MR images were linearly and non-linearly deformed to the MNI template using SPM Segment, as described in Section 4.5.1. The resulting inverse transformation parameters were then applied to transform the MNI-space cluster into the native MRI space of each subject and timepoint using tri-linear interpolation. The cluster was also re-sampled to the higher resolution of the MRI. Two images for which SPM Segment was unable to provide satisfactory transformations were additionally excluded from the analysis (see Appendix B).

5.4 Preliminary studies

Before proceeding with a multi-region analysis of the longitudinal ADNI FDG-PET imaging data, preliminary studies involving only the hippocampal region were performed. Section 5.4.1 describes the automatic generation of hippocampal segmentations for both baseline and 12-month MR images. Section 5.4.2 describes the subsequent extraction of hippocampal FDG-

PET signal intensities at both imaging timepoints, as well as investigations of group differences. Binary classification experiments are then described in Section 5.4.3.

5.4.1 Hippocampal segmentation

The 12-month MR images were affinely co-registered with their corresponding baseline MR images. Rigid registration parameters were estimated, and used as a starting point for the estimation of a twelve-parameter affine registration, using normalised mutual information as the similarity criterion. The 12-month MR images and their corresponding co-registered 12-month FDG-PET images were then transformed into the space of the baseline MR images using linear interpolation. The follow-up MRI and both baseline and follow-up FDG-PET images were thus all aligned with the baseline MRI, as illustrated in Figure 5.3. Affine registration was used to obtain brain-wide alignment while maintaining local changes such as atrophy.

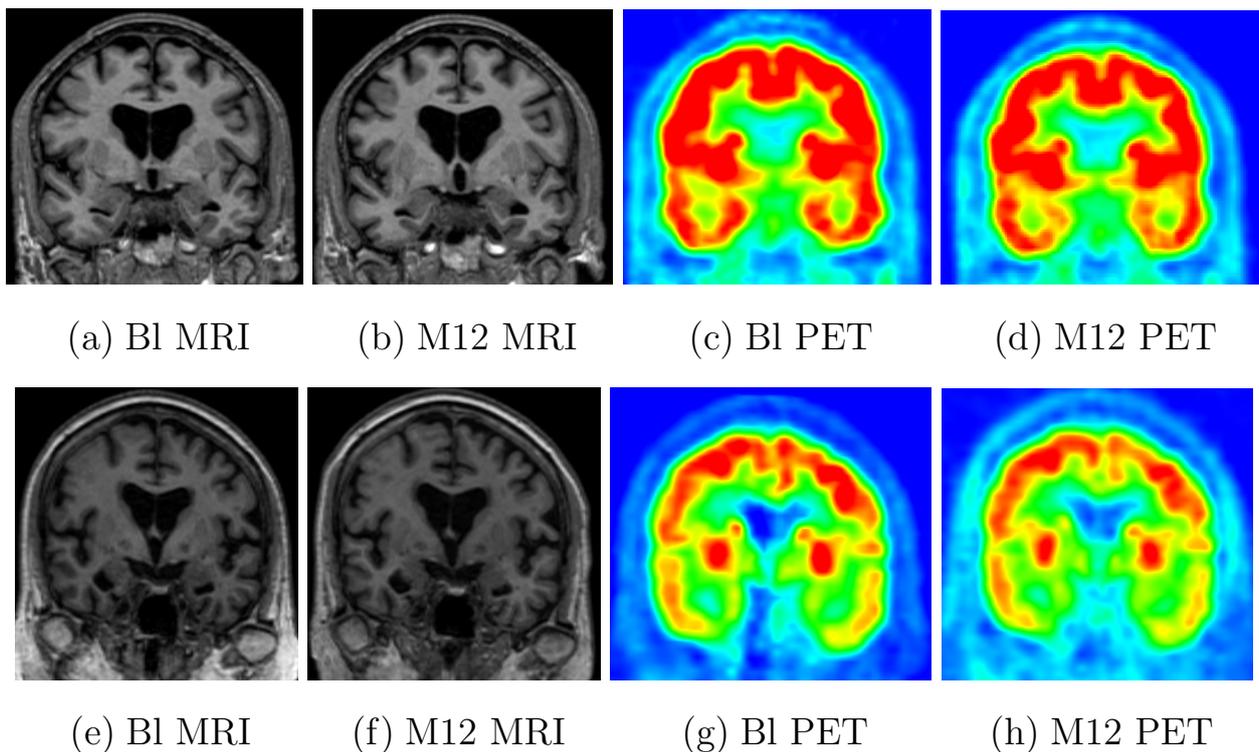


Figure 5.3: Typical examples showing the affine alignment to baseline (Bl) MRI space of the 12-month (M12) MRI and FDG-PET images at both timepoints. Images are shown for a HC (a-d), as well as an AD patient (e-h). FDG-PET images are shown after intensity normalisation.

Hippocampal segmentations for both the baseline and follow-up images were generated simul-

taneously in baseline MRI space using multi-atlas segmentation followed by a 4D graph-cut method. These segmentations were prepared as part of a project involved with measuring hippocampal atrophy, described in [Wolz et al. \(2010\)](#). Since both timepoints were segmented simultaneously, the segmentations were both consistent and sensitive to atrophy. Typical examples of the hippocampal segmentations are shown in [Figure 5.4](#).

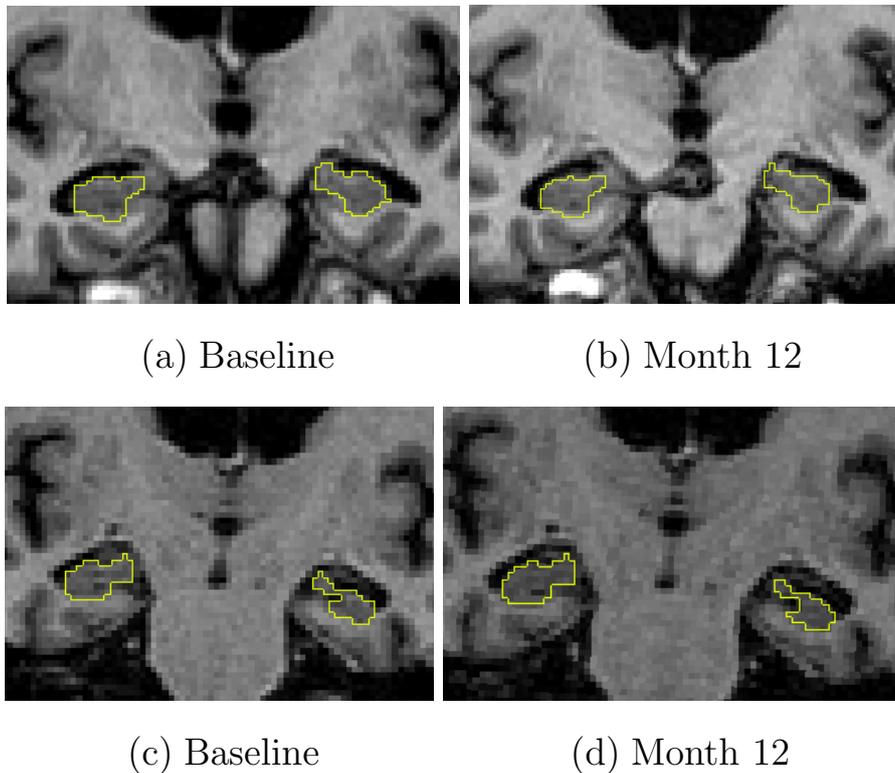


Figure 5.4: Typical examples of the hippocampal segmentations at both timepoints for a HC (a-b), as well as an AD patient (c-d).

5.4.2 Image analysis

Each of the MRI-space FDG-PET images described in [Section 5.4.1](#) was overlaid with its corresponding segmentation, and the signal intensity per mm^3 in the hippocampus determined at each timepoint. Global inter-subject variations were accounted for by normalisation to the signal intensity per mm^3 in the independently-derived reference cluster described in [Section 5.3](#). Normalised hippocampal signal intensities were thus extracted from both the baseline and 12-month FDG-PET images, and the relative changes in signal intensity over the 12-month follow-up period were additionally determined. These feature data are illustrated in [Figure 5.5](#).

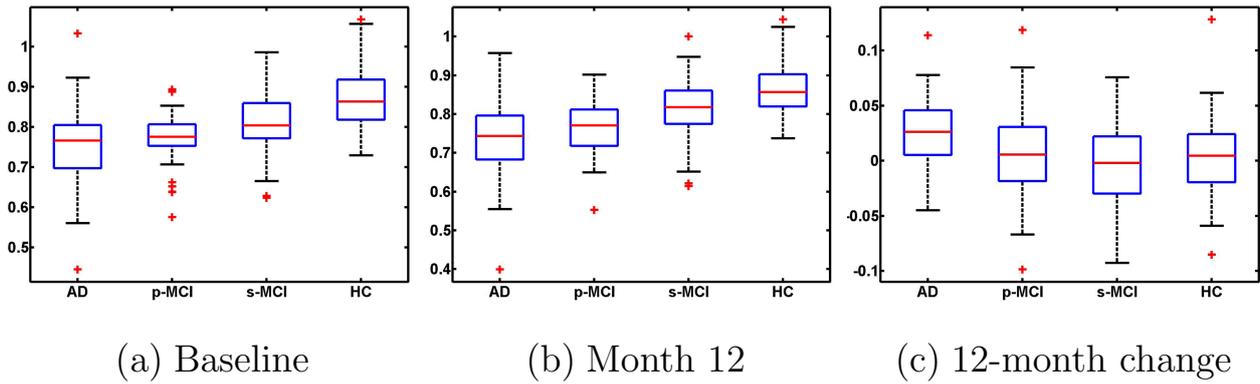


Figure 5.5: Boxplots showing group differences based on the normalised FDG-PET signal intensity per mm^3 in the hippocampus at (a) baseline and (b) month 12, as well as based on (c) the intensity change over the follow-up period.

The FDG-PET signal intensity in small structures such as the hippocampus can be influenced by the partial volume effect. This is a consequence of the limited PET detector resolution, and results in an apparent loss of intensity in structures that are smaller than twice the scanner’s FWHM resolution (Hoffman et al., 1979). The partial volume effect can result in the underestimation of FDG-PET signal intensities in small hippocampi. It therefore works in the favour of classification studies, since signal intensities in the smaller hippocampi of AD patients appear even further reduced compared with those in the larger hippocampi of HC. It is possible to correct for the effects of partial volume (Rousset et al., 2007; Aston et al., 2002), but this is generally not feasible as part of routine clinical practice. Partial volume correction is therefore not applied for the studies presented in this thesis.

5.4.3 Classification experiments

The potential use of the hippocampal FDG-PET features for classification was next investigated. Two sets of cross-sectional features had been extracted for each subject (hippocampal signal intensities at each timepoint). Longitudinal features had also been evaluated as the relative changes in signal intensity over the 12-month follow-up period. Two additional feature sets which combined the cross-sectional and longitudinal information were formed by concatenating the longitudinal change features with the signal intensities at either timepoint. Classification rates were assessed between four clinically relevant pairs of diagnostic groups (AD/HC,

pMCI/HC, AD/sMCI, pMCI/sMCI) using the linear discriminant function classifier described in Section 3.2.1. This simple linear classifier was sufficient for these preliminary investigations in which only a small number of features were available. Robust estimates of classifier performance were obtained via the stratified repeated random sampling approach described in Section 3.3.2. The mean classification accuracy, sensitivity and specificity for pairs of groups were evaluated over 1,000 runs in which 75% of the subjects were randomly selected for training, with the remaining 25% used as test data. For each group pair, unpaired t-tests between the distributions of classification results obtained from the 1,000 leave-25% out runs were performed to assess the significance of differences in performance between the five feature sets. Results are shown in Table 5.2.

		Bl	M12	Ch	Bl & ch	M12 & ch
AD/HC	Acc (%)	76.2 (0.3)	80.4 (0.2)	69.5 (0.3)	81.0 (0.2)	81.1 (0.2)
	Sens (%)	77.5 (0.4)	75.8 (0.4)	68.3 (0.4)	75.7 (0.4)	75.5 (0.4)
	Spec (%)	75.1 (0.4)	84.3 (0.3)	70.6 (0.4)	85.5 (0.3)	85.7 (0.3)
pMCI/HC	Acc (%)	70.9 (0.3)	74.1 (0.2)	44.1 (0.3)	75.6 (0.3)	76.2 (0.3)
	Sens (%)	74.6 (0.4)	75.6 (0.3)	42.9 (0.4)	79.1 (0.3)	79.6 (0.4)
	Spec (%)	67.8 (0.4)	77.8 (0.4)	45.0 (0.4)	72.6 (0.4)	73.3 (0.4)
AD/sMCI	Acc (%)	62.3 (0.3)	70.1 (0.2)	68.2 (0.3)	69.5 (0.3)	71.3 (0.2)
	Sens (%)	58.2 (0.5)	67.6 (0.4)	75.6 (0.4)	70.7 (0.4)	71.3 (0.4)
	Spec (%)	65.3 (0.4)	74.8 (0.3)	62.7 (0.4)	68.6 (0.4)	69.9 (0.3)
pMCI/sMCI	Acc (%)	58.3 (0.3)	60.1 (0.2)	50.9 (0.3)	60.6 (0.3)	61.3 (0.3)
	Sens (%)	60.7 (0.5)	61.7 (0.4)	50.4 (0.4)	58.5 (0.5)	59.8 (0.5)
	Spec (%)	56.6 (0.4)	62.2 (0.4)	51.2 (0.4)	62.1 (0.4)	61.5 (0.4)

Table 5.2: Classification accuracy (acc), sensitivity (sens) and specificity (spec) expressed as mean (standard error) over the 1,000 leave-25%-out runs. Results are provided for baseline signal intensities (Bl), 12-month signal intensities (M12), relative changes in signal intensity over 12 months (Ch), baseline signal intensities concatenated with 12-month changes (Bl & ch), and 12-month signal intensities concatenated with 12-month changes (M12 & ch).

Significant ($p < 0.05$) increases in classification accuracy were achieved when using 12-month hippocampal signal intensities compared with using baseline signal intensities for all clinical group pairs except pMCI/sMCI. Classification based on longitudinal information alone had significantly ($p < 0.05$) lower accuracy compared with using either of the two cross-sectional feature sets, except for AD/sMCI. Consistent increases in accuracy were achieved when combining longitudinal information with 12-month data, compared with its combination with baseline data, although differences do not reach significance. Classification based on the combination

of longitudinal data with 12-month signal intensities was consistently improved compared with using 12-month signal intensities alone, although again differences do not reach significance.

5.4.4 Discussion

Similarly to [Hinrichs et al. \(2011\)](#), it was observed that the relative changes in signal intensity over 12 months alone did not provide particularly good classification performance. Increases in classification accuracy were achieved by combining the cross-sectional and longitudinal features, with the highest accuracies achieved for all group pairs when using the combination of 12-month signal intensities and changes over 12 months. Although not all differences in classification performance reached significance, these results suggest that longitudinal FDG-PET features may provide some complementary information which can enhance classification when used in conjunction with cross-sectional FDG-PET features. The additional benefit of using multi-region information was therefore investigated.

5.5 Multi-region segmentation

Automatic whole-brain segmentations into 83 anatomical regions had been prepared in the native space of each baseline MRI using MAPER as described in Section 4.4. To obtain comparable segmentations for the follow-up images, the baseline segmentations were nonrigidly propagated to the native space of their corresponding 12-month MRI. Although incorrectly labelled voxels in the baseline segmentation were thereby propagated to the follow-up image, this method produced consistent segmentations for each subject. Intra-subject segmentation consistency is important for measuring longitudinal change ([Crum et al., 2001](#)) since uncorrelated errors lead to greater measurement uncertainty. The MAPER segmentation procedure was therefore not used directly to generate independent segmentations for the follow-up images.

The intracranial portion of each 12-month MRI was determined by rigid propagation of the baseline intracranial mask that had been used for brain extraction during the MAPER segmen-

tation procedure described in 4.4. This process is illustrated in Figure 5.6.

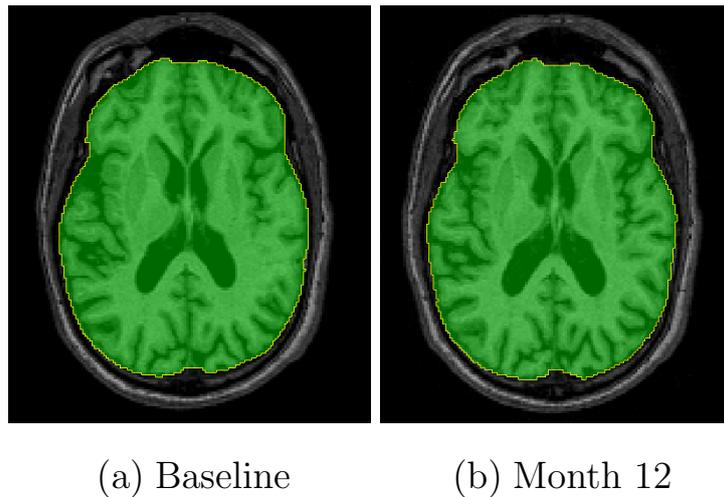


Figure 5.6: An illustration of the rigid propagation of the intracranial brain mask from (a) baseline MRI space to (b) 12-month MRI space.

The rigidly aligned intracranial-masked baseline and 12-month MRI were then affinely aligned, followed by a series of nonrigid registrations. Nonrigid registration was performed using control point spacings of 10, 5, and 2.5 mm. The unmasked baseline anatomical segmentation was nonrigidly propagated to 12-month MRI space using nearest neighbour interpolation. Individual tissue probability maps for CSF, grey matter and white matter were obtained for the 12-month MRI using FSL FAST, and the segmentation masked as described in Section 4.4.3. Typical examples of the masked segmentations at both timepoints are shown in Figure 5.7.

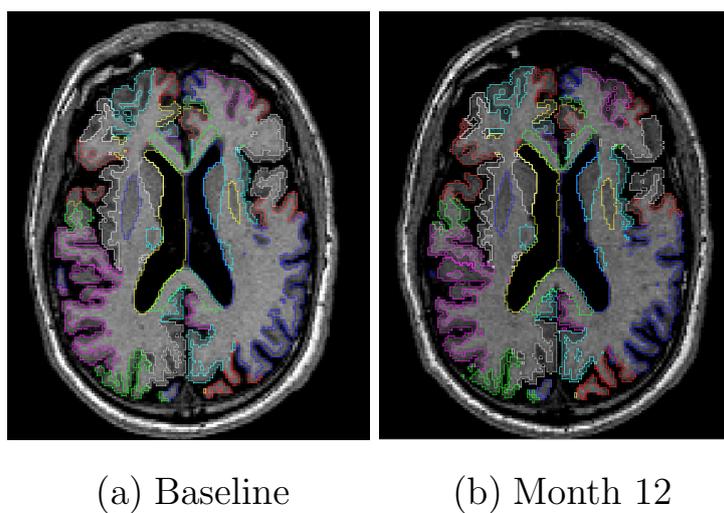


Figure 5.7: Typical examples of the masked multi-region segmentations in (a) baseline MRI space, and (b) 12-month MRI space.

5.6 Multi-region image analysis

Each of the native MRI-space FDG-PET images described in Section 5.3 was overlaid with its corresponding masked anatomical segmentation. The FDG-PET signal intensity per mm^3 was determined for each of the 83 anatomically defined regions at both imaging timepoints. Global variations in the cerebral metabolic rate of glucose between subjects were accounted for by normalisation to the signal intensity per mm^3 in the independently-derived reference cluster also described in Section 5.3. Normalised regional signal intensities were thus extracted from both the baseline and 12-month FDG-PET images. Additionally, the regional changes in FDG-PET signal intensity over the 12-month follow-up period were determined.

To verify that the regional features to be used for classification were biologically plausible, t-tests were performed to identify the features which gave significant differences between diagnostic groups ($p < 0.01$, uncorrected for multiple comparisons). Both sets of cross-sectional features (baseline and 12-month regional signal intensities) were considered, as well as the regional changes in signal intensity over the 12-month follow-up period. Regional t-values between AD patients and HC based on the cross-sectional features indicated significant differences across most of the brain. This finding was consistent with the voxel-wise t-tests reported in [Yakushev et al. \(2009\)](#). The most significantly different regions between groups included those known to be affected in AD for all three feature sets, consistent with previous voxel-wise t-tests performed on the ADNI FDG-PET data ([Langbaum et al., 2009](#); [Chen et al., 2010](#)).

The majority of regions differed significantly between AD patients and HC for both baseline intensities (65/83 regions), and 12-month intensities (73/83 regions). In addition to more regions reaching significance, significance levels were also higher in the 12-month data than in the baseline data. Far fewer regions reached significance for the change features (26/83 regions), and significance levels were lower than for either of the cross-sectional feature sets. These results are illustrated in Figure 5.8, and similar patterns were consistently observed across the remaining three clinical group pairs (pMCI/HC, AD/sMCI, pMCI/sMCI). For these pairs of groups, fewer regions reached significance than between AD patients and HC, and at reduced significance levels. The fewest significant regions and lowest significance levels were

found between pMCI and sMCI patients.

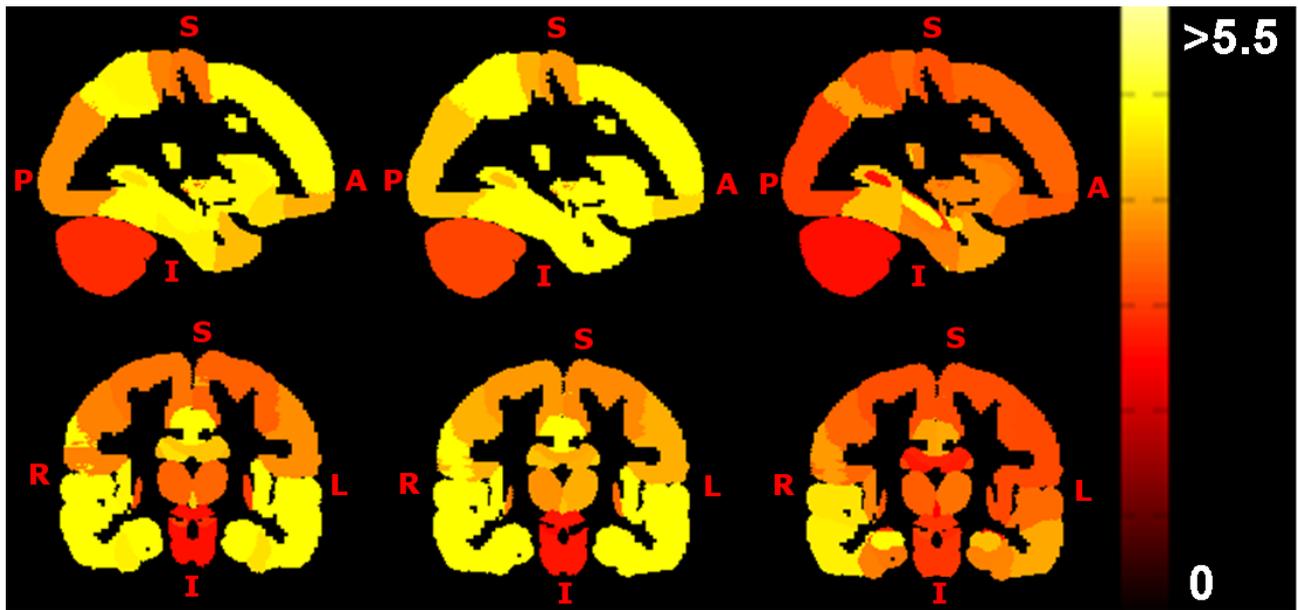


Figure 5.8: Regional t-values for comparisons between AD patients ($n = 50$) and HC ($n = 54$) superimposed onto sagittal (top row) and coronal (bottom row) slices of a maximum probability brain atlas which has been masked according to the same procedure as the anatomical segmentations. The feature sets tested are, from left to right: baseline signal intensities; 12-month signal intensities; changes in signal intensity over 12 months. To allow all three feature sets to be visualised using the same colour scale, so that their spatial patterns may be compared, all t-values greater than 5.5 have been set to the maximum value. A: anterior, P: posterior, S: superior, I: inferior, R: right, L: left.

For both sets of cross-sectional features, the five most significantly different regions between AD patients and HC were the bilateral hippocampus, left parietal lobe, left posterior temporal lobe, and right posterior cingulate gyrus. However, only one of these regions (right hippocampus) was among the five most significantly different regions for the change features, along with the right amygdala, right middle and inferior temporal gyri, right posterior part of the superior temporal gyrus, and right posterior temporal lobe. For the remaining three group pairs, the five most significantly different regions for each of the three feature sets contained some combination of the regions identified between AD patients and HC, with the parahippocampal gyrus also identified in some cases.

Interestingly, the amygdala was consistently identified among the five most significantly different regions for the change features, but not for either of the cross-sectional feature sets. It was the only region reaching significance for the change features between pMCI and sMCI patients.

5.7 Classification experiments

The potential use of combining multi-region cross-sectional and longitudinal FDG-PET data for image-based AD classification was next investigated. Two sets of cross-sectional features had been extracted for each subject (regional signal intensities at each of the timepoints). Longitudinal features had also been evaluated as the relative changes in signal intensity over the 12-month follow-up period. Two additional feature sets which combined the cross-sectional and longitudinal information were formed by concatenating the longitudinal change features with the signal intensities at either timepoint.

5.7.1 Methods

Classification rates were assessed between four clinically relevant pairs of diagnostic groups (AD/HC, pMCI/HC, AD/sMCI, pMCI/sMCI) using the SVM classifier described in Section 3.2.2. The SVM classifier was applied using the soft-margin formulation from LIBSVM with a radial basis function kernel, as described in Section 4.7.1. Robust estimates of classifier performance were obtained via the stratified repeated random sampling approach described in Section 3.3.2. The mean classification accuracy, sensitivity and specificity for pairs of groups were evaluated over 1,000 runs in which 75% of the subjects were randomly selected for training, with the remaining 25% used as test data. For each group pair, unpaired t-tests between the distributions of classification results obtained from the 1,000 leave-25% out runs were performed to assess the significance of differences in performance between the five feature sets.

To demonstrate that classification was truly based on disease-specific imaging information, rather than the intrinsic age and gender information captured in the images, classification was additionally performed after accounting for these effects. A linear regression step was incorporated into the classification procedure for every clinical group pair such that, for each of the 1,000 repetitions, a regression model was estimated from the training data, and the SVM trained on the residuals. The regression model was then applied to the test data, and the SVM tested on the resulting residuals.

5.7.2 Results

For all four clinical group pairs, highly significant ($p < 0.001$) increases in classification accuracy were achieved when using 12-month signal intensities compared with using baseline signal intensities. Classification based on the longitudinal information alone had significantly ($p < 0.05$) lower accuracy compared with using either of the two cross-sectional feature sets. Highly significant ($p < 0.001$) increases in classification accuracy were achieved when combining longitudinal information with 12-month data, compared with its combination with baseline data. In addition, classification based on the combination of longitudinal data with 12-month signal intensities was significantly ($p < 0.05$) improved compared with using 12-month signal intensities alone. All classification accuracies were significantly different from chance, as assessed using the permutation testing procedure described in Section 3.3.2.

The above results are illustrated as boxplots in Figure 5.9, with numerical results provided in Table 5.3. ROC curves, as described in Section 3.3.1, for classification based on the best-performing combined feature set (longitudinal change concatenated with 12-month signal intensities) are displayed in Figure 5.10.

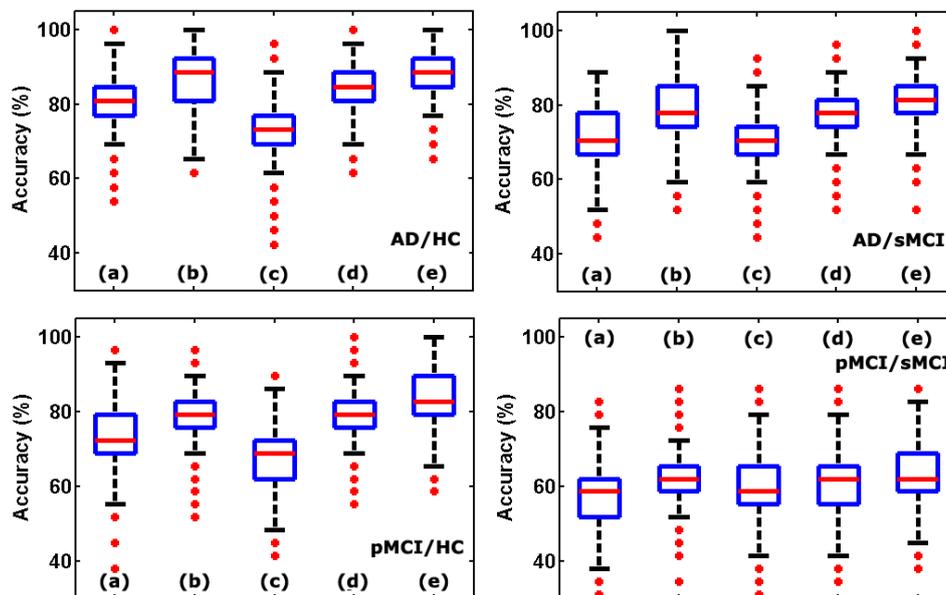


Figure 5.9: Classification accuracies for the four clinical group pairs based on the five feature sets studied. From left to right for each boxplot: (a) baseline signal intensities, (b) 12-month signal intensities, (c) change over 12 months, (d) combined baseline intensities and change, (e) combined 12-month intensities and change.

		Bl	M12	Ch	Bl & ch	M12 & ch
AD/HC	Acc (%)	80.9 (0.2)	86.1 (0.2)	73.7 (0.3)	85.9 (0.2)	88.4 (0.2)
	Sens (%)	79.6 (0.3)	81.2 (0.3)	66.4 (0.4)	81.2 (0.3)	83.2 (0.3)
	Spec (%)	82.3 (0.3)	91.0 (0.3)	81.0 (0.4)	90.6 (0.3)	93.6 (0.3)
pMCI/HC	Acc (%)	70.7 (0.3)	79.2 (0.3)	70.0 (0.3)	77.7 (0.3)	81.3 (0.2)
	Sens (%)	70.5 (0.4)	77.2 (0.4)	63.0 (0.4)	75.1 (0.4)	79.8 (0.3)
	Spec (%)	71.0 (0.4)	81.4 (0.3)	77.5 (0.4)	80.4 (0.3)	82.9 (0.3)
AD/sMCI	Acc (%)	72.7 (0.3)	79.3 (0.2)	67.2 (0.3)	78.5 (0.2)	83.5 (0.2)
	Sens (%)	65.6 (0.4)	77.9 (0.3)	56.6 (0.5)	73.9 (0.4)	79.9 (0.3)
	Spec (%)	78.5 (0.3)	80.5 (0.3)	75.8 (0.3)	82.2 (0.3)	86.4 (0.3)
pMCI/sMCI	Acc (%)	58.4 (0.3)	62.3 (0.3)	58.4 (0.3)	61.8 (0.3)	63.1 (0.3)
	Sens (%)	51.5 (0.4)	53.2 (0.4)	50.2 (0.5)	52.5 (0.5)	52.2 (0.5)
	Spec (%)	64.8 (0.5)	70.8 (0.4)	68.5 (0.4)	70.5 (0.4)	73.2 (0.4)

Table 5.3: Classification accuracy (acc), sensitivity (sens) and specificity (spec) expressed as mean (standard error) over the 1,000 leave-25%-out runs. Results are provided for baseline signal intensities (Bl), 12-month signal intensities (M12), relative changes in signal intensity over 12 months (Ch), baseline signal intensities concatenated with 12-month changes (Bl & ch), and 12-month signal intensities concatenated with 12-month changes (M12 & ch).

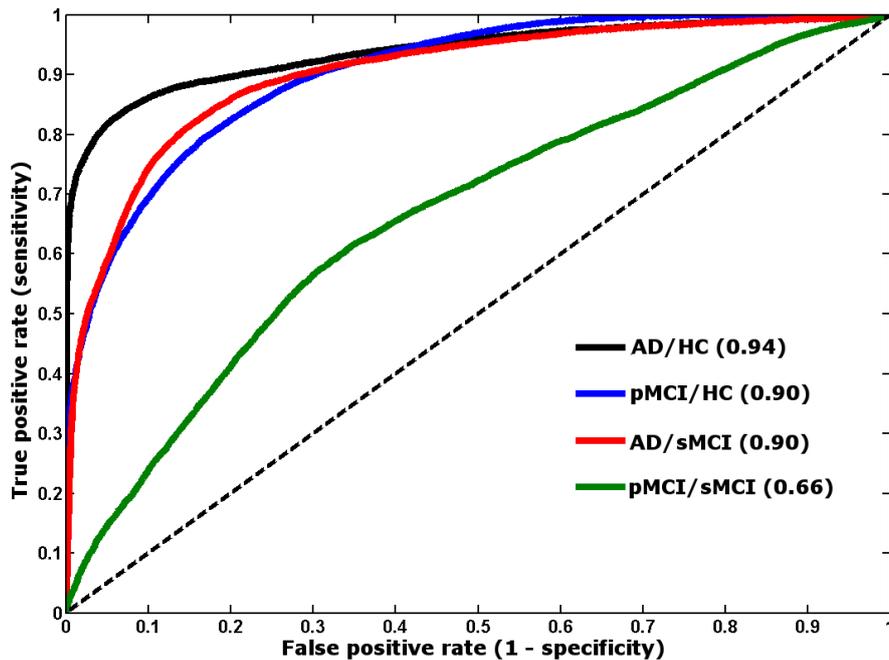


Figure 5.10: ROC curves for the combined feature set of relative changes concatenated with 12-month signal intensities. AUC values for each clinical group pair are provided in brackets.

Linear regression for age and gender had no significant effect on the classification accuracy for the majority of experiments. Classification results after linear regression are shown in Table 5.4, in which cases where regression did significantly affect the accuracy are indicated.

	Bl	M12	Ch	Bl & ch	M12 & ch
AD/HC	81.2 (0.2)	85.9 (0.2)	73.6 (0.3)	84.8 (0.2)	87.1 (0.2)
pMCI/HC	73.7 (0.3)*	80.2 (0.3)	68.9 (0.3)	78.4 (0.3)	81.6 (0.2)
AD/sMCI	74.2 (0.3)	78.8 (0.3)	67.8 (0.3)	79.1 (0.2)	83.5 (0.3)
pMCI/sMCI	61.5 (0.3)*	62.2 (0.3)	58.9 (0.3)	62.7 (0.3)	64.0 (0.3)

Table 5.4: Classification accuracy (%) after regression for age and gender expressed as mean (standard error) over the 1,000 leave-25%-out runs. Results are provided for baseline signal intensities (Bl), 12-month signal intensities (M12), relative changes in signal intensity over 12 months (Ch), baseline signal intensities concatenated with 12-month changes (Bl & ch), and 12-month signal intensities concatenated with 12-month changes (M12 & ch). Cases where the regression significantly ($p < 0.05$) affected the accuracy are indicated by asterisks (*). In both these cases, regression improved the accuracy.

5.7.3 Discussion

Classification results in line with the current state-of-the-art have been achieved by using a combination of cross-sectional and longitudinal multi-region FDG-PET information. For the most commonly reported classification task of separating AD patients from HC, the accuracy of 88% is comparable with other recent classification results based on multi-modality imaging and non-imaging data (for example, [Hinrichs et al. \(2011\)](#): 92% accuracy, [Zhang et al. \(2011\)](#): 93% accuracy), and also with the results of high-dimensional pattern recognition methods applied to cross-sectional MR imaging data (for example, [Cuingnet et al. \(2011\)](#): up to 81% sensitivity, 95% specificity). As explained in Section 4.7.3, classification results may be converging on a glass ceiling of around 90% for this task. The less commonly reported, but more clinically significant, task of separating pMCI from sMCI patients was also investigated. The accuracy of 63% is encouraging compared with the most directly comparable studies based on MR imaging data (for example, [Cuingnet et al. \(2011\)](#): up to 62% sensitivity, 69% specificity), [Wolz et al. \(2010\)](#): 64% accuracy). However, as explained in Section 4.7.3, to properly assess the utility of any classification method in predicting progression, longer clinical follow-up would be required than is currently available for the ADNI participants.

Similarly to [Hinrichs et al. \(2011\)](#), it was observed that the percentage change in signal intensity over 12 months alone did not provide particularly good classification performance between AD patients and HC (74% accuracy). Although the longitudinal change data alone appear insuffi-

cient for matching state-of-the-art classification performance, the results of experiments based on combined feature sets show that they provide some complementary information which can enhance classification when used in conjunction with cross-sectional features. This suggestion is supported by the t-test results displayed in Section 5.6, which show that the pattern of regional significances differs between cross-sectional and longitudinal data. For example, the amygdala was identified among the best five features for group discrimination only for the longitudinal data. The two cross-sectional feature sets, on the other hand, had similar patterns of regional significances, although improved group discrimination was achieved with the 12-month data. This is to be expected, since patients that are 12 months further along the disease process should be more easily distinguished from HC. Rather than relying on t-tests to explore the relative feature importances, a direct visualisation of the SVM weight vector would be desirable. However, because of the nonlinearity of the kernel used, it was not possible to map the weight vectors learned in the transformed feature space back to the original feature space.

The aim of this work has not been to introduce a novel classification approach, but instead to use a readily available SVM classifier and simple feature combination approach (direct concatenation) to demonstrate the utility of longitudinal FDG-PET information for improving classification amongst four clinically relevant pairs of diagnostic groups. All classification experiments were also performed after accounting for the effects of age and gender by linear regression. The lack of significant effect on accuracy observed in the majority of cases indicates that the classification results were truly based on disease-specific imaging information, rather than the intrinsic age and gender information also captured in the images.

This work demonstrates that information extracted from serial FDG-PET through regional analysis can accurately discriminate diagnostic groups, even at the early symptomatic stages of AD. This finding may be usefully applied in the diagnosis of AD, predicting disease course in MCI patients, and in the selection of participants for clinical trials. Importantly, the utility of serial regional FDG-PET for patient classification is demonstrated in a realistic multi-centre setting. Although the use of longitudinal data for the clinical diagnosis of AD is not necessarily practical, its use for stratification of pMCI versus sMCI patients could still be valuable. For clinical trial recruitment, it may be acceptable to use longitudinal information acquired over

12 months to gain additional certainty about whether a candidate fits the selection criteria.

5.8 Conclusion

This chapter has presented a multi-region analysis of longitudinal FDG-PET imaging data from the ADNI study. Whole-brain segmentations into 83 anatomically defined regions were automatically generated in the native MRI space of each subject for both baseline and 12-month images. These were used to extract regional signal intensities from the corresponding FDG-PET images, and their combined use for image-based AD classification was investigated. The effect on classification of the subjects' gender and age at scan were additionally considered. The classification results are in line with the current state-of-the-art, but there are several areas for further research. For example, a more sophisticated method could be used for data combination, and the information from both the FDG-PET and MR images could be combined. The work presented in the following chapter will focus on these issues, as well as investigating the incorporation of non-imaging data such as CSF biomarkers and genetic information.

Chapter 6

Random forest-based similarities for multi-modality classification

Work in this chapter has, in part, been presented in:

K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, and D. Rueckert. Random forest-based manifold learning for classification of imaging data in dementia. (In Proceedings) Machine Learning in Medical Imaging Workshop at 14th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'11). Lecture Notes in Computer Science, 7009:159–166, 2011.

6.1 Introduction

This chapter presents a multi-modality classification framework evaluated using imaging and biological data from the ADNI study. As described in Section 1.5, changes in multiple neuroimaging and biological measures may provide complementary information for the diagnosis and prognosis of AD. Automated classification of individual patients based on multiple biomarkers could provide valuable support for clinicians, when considered alongside cognitive assessment scores and traditional visual image analysis. Two independent studies investigating multi-modality classification using multi-kernel learning have reported that classification based

on multi-modality data was superior to that based on any individual modality (Zhang et al., 2011; Hinrichs et al., 2011).

This work presents an alternative approach for multi-modality classification, based on pairwise similarity measures derived from random forest classifiers. The similarities were used to construct a manifold representation from labelled training data and then to infer the diagnostic labels of test data mapped into this space. This method facilitates the incorporation of multi-modality data, since similarities derived from several datasets may be readily combined to generate an embedding that simultaneously encodes information from all features. Classification based on the combination of regional MRI volumes, voxel-based FDG-PET signal intensities, CSF biomarker measures, and ApoE allele status is superior to that based on any individual modality for comparisons between AD patients and HC, as well as between MCI patients and HC. Results are comparable with those obtained using multi-kernel learning in Zhang et al. (2011) and Hinrichs et al. (2011). Since random forest classifiers extend naturally to multi-class problems, the framework described here could be used for other applications in the future, such as the differential diagnosis of AD.

Manifold learning techniques based on pairwise similarities between images have been applied in a variety of neuroimaging studies. For example, Laplacian eigenmaps (Belkin and Niyogi, 2003) have been used to generate an embedding of brain MR images based on similarities derived from overlaps of their structural segmentations (Aljabar et al., 2008). Isomap (Tenenbaum et al., 2000) has also been used to estimate the manifold structure of brain MR images, using distance measures based on nonrigid transformations between image pairs (Gerber et al., 2009). A framework for fusing manifold learning steps based on multiple pairwise similarity measures has been presented in Aljabar et al. (2010). The method described here uses random forests to derive consistent pairwise similarity measures for multiple modalities, thus facilitating the combination of different types of feature data.

The ADNI participants whose imaging and biological data are used in this chapter are first described in Section 6.2. Details of the multi-modality classification framework are then presented in Section 6.3. The results of preliminary experiments based on the combination of voxel-based

FDG-PET and region-based MR imaging data are presented in Section 6.4, followed by results based on the combination of imaging and biological data in Section 6.5.

6.2 Imaging and biological data

Clinical, imaging, and biological data used in this chapter were obtained from the ADNI database, which has been described in Section 4.2. Baseline FDG-PET and MR imaging data from the 287 participants described in Section 4.2 were used for the preliminary experiments presented in Section 6.4. ApoE genotype information was available for all 287 participants, but baseline CSF measures were only available for a subset of 147. Groupwise characteristics are provided in Table 6.1 for this subset of participants, whose imaging and biological data were used for the experiments presented in Section 6.5. The MCI patients in this subset were divided into pMCI and sMCI based on changes in clinical status occurring over 20 ± 11 (range 6 – 36) months. The mean age at baseline does not vary significantly ($p > 0.01$) on t-test between the clinical groups.

	N(F)	Age	MMSE Score	CDR (%)		
		(mean \pm std. dev.)	(mean \pm std. dev.)	0	0.5	1
AD	37(14)	76.8 \pm 6.6	23.5 \pm 2.0	0	49	51
pMCI	34(12)	76.1 \pm 7.3	26.5 \pm 1.7	0	100	0
sMCI	41(12)	75.7 \pm 6.5	27.3 \pm 1.8	0	100	0
HC	35(12)	74.5 \pm 5.2	28.9 \pm 1.2	100	0	0

Table 6.1: Clinical and demographic information for the subset of participants for whom CSF data were available. For each clinical group, the total number of subjects (N) and number of females (F) are shown, along with the average age, average MMSE score, and CDR distribution.

The four imaging and biological feature sets used for evaluation of the multi-modality classification framework presented in this chapter are described in the following subsections.

6.2.1 Region-based MRI features

Automatic whole-brain segmentations into 83 anatomical regions had already been prepared in the native space of each baseline MRI using MAPER, as described in Section 4.4. Since AD

is characterised by cortical neuronal loss, the masked segmentations described in Section 4.4.3 were employed for feature extraction. Regional volumes were computed for all 83 structures, and normalised by the total intracranial volume (ICV) computed from the masks described in Section 4.4.1. ICV normalisation has been shown to substantially reduce variation, remove gender-related differences (Whitwell et al., 2001), and eliminate inaccuracies arising from scaling or voxel size errors remaining after phantom correction of the MRI (Clarkson et al., 2009).

6.2.2 Voxel-based FDG-PET features

Voxel-based features were extracted from the FDG-PET images to enable demonstration that the proposed multi-modality classification approach could readily combine different types of features. The FDG-PET images were aligned with the MNI template and smoothed, as described in Section 4.5.1. Global inter-subject variations were accounted for by normalisation to the signal intensity per mm^3 in the independently-derived reference cluster described in Section 4.5.3. The SPM brain mask, which covers white and grey matter in MNI space, was thresholded at 50% probability, and applied to each normalised FDG-PET image to exclude voxels outside the brain. Signal intensities were then extracted from all remaining voxels, resulting in 239,304 voxel-based features per image.

6.2.3 Biological CSF and ApoE genotype features

The ADNI Biomarker Core, based at the University of Pennsylvania, provides biological data for the study participants. These data include CSF measures of $A\beta$, tau and phosphorylated tau (ptau), as well as ApoE genotype information determined from a blood sample. Details of the biofluid collection and processing are provided in Trojanowski et al. (2010). The genetic feature data for each participant consist of a single categorical variable describing their ApoE genotype. This categorical feature takes one of five possible values: (ϵ_3, ϵ_3) , (ϵ_3, ϵ_4) , (ϵ_4, ϵ_4) , (ϵ_2, ϵ_3) , (ϵ_2, ϵ_4) . Groupwise CSF measures and genotype information are provided in Table 6.2 for the 147 participants for whom biological data were available.

	CSF measures (pg/mL)			ApoE genotype (%)				
	$A\beta$	tau	ptau	(ϵ_3, ϵ_3)	(ϵ_3, ϵ_4)	(ϵ_4, ϵ_4)	(ϵ_2, ϵ_3)	(ϵ_2, ϵ_4)
AD	141 \pm 43	127 \pm 64	46 \pm 24	24	41	30	0	5
pMCI	152 \pm 47	96 \pm 41	37 \pm 12	41	41	18	0	0
sMCI	167 \pm 55	105 \pm 81	35 \pm 22	39	39	12	10	0
HC	208 \pm 56	69 \pm 28	26 \pm 16	48	23	0	26	3

Table 6.2: Groupwise CSF measures and genetic information for the 147 participants for whom biological data were available. CSF measures are expressed as mean \pm standard deviation, and genetic information is expressed as the proportion of each diagnostic group possessing each of the five ApoE allele pairs.

6.3 Multi-modality classification framework

A schematic overview of the proposed similarity-based multi-modality classification approach is shown in Figure 6.1.

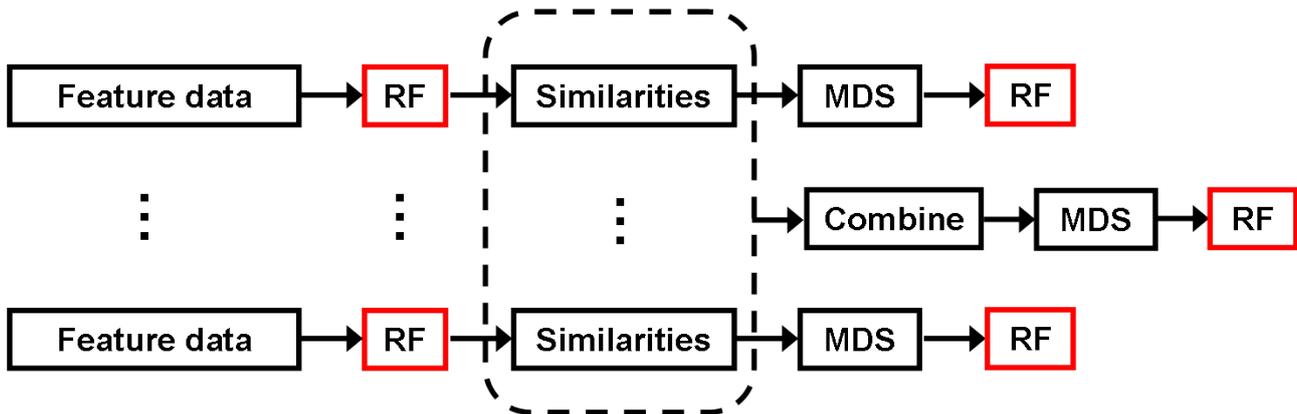


Figure 6.1: Schematic overview of the proposed multi-modality classification approach. Each random forest (RF) step provides a classification result whose performance will be reported. Random forests are used both to derive the similarity matrices for each feature set, and also to perform the single- and multi-modality classification experiments.

A random forest classifier, as described in Section 3.2.4, was applied to the feature data from each modality independently. Single-modality classification results for comparison were thus obtained, as well as pairwise similarity measures between subjects. These similarities were used to construct single-modality manifold representations from labelled training data and then to infer the diagnostic labels of test data mapped into this space. Similarities from multiple modalities were additively combined to generate an embedding that simultaneously encodes information from all features. Multi-modality classification was then performed using coordinates from this joint embedding.

Classical MDS was applied as described in Section 3.2.4 to learn the manifolds on which to perform classification. MDS is commonly used to provide low-dimensional visualisations of similarity relationships, including those derived from random forests (Hastie et al., 2011). Random forest-derived similarities have been successfully applied in unsupervised clustering tasks, for example those involving high-dimensional genetic or tissue microarray data (Shi and Horvath, 2006; Shi et al., 2005). Here, random forests are used to derive *supervised* similarities, with the aim of generating manifolds that are optimal for the task of clinical group discrimination.

The multi-modality classification framework described in this chapter has been implemented using the R package for random forests. This is a port of Leo Breiman and Adele Cutler's original Fortran code, by Andy Liaw and Matthew Wiener (Liaw and Wiener, 2002) (<http://cran.r-project.org/web/packages/randomForest>).

6.4 Combining FDG-PET and MR imaging data

The multi-modality classification framework described in Section 6.3 was first applied for the combination of voxel-based FDG-PET and region-based MR imaging data. These data comprised baseline images acquired from the 287 ADNI participants described in Section 4.2. Details of the methodology and results obtained from both single- and multi-modality classification experiments are presented in the following subsections.

6.4.1 Classification methodology

Classification performance was assessed between two clinically relevant pairs of diagnostic groups (AD/HC, MCI/HC) based on both single- and multi-modality imaging information. The mean classification accuracy, balanced accuracy, sensitivity and specificity for both group pairs were evaluated using the stratified ten-fold cross-validation method described in Section 3.3.2, with the same folds used for all experiments. Classification performance is commonly reported in terms of accuracy, but here statistical comparisons between experiments are per-

formed based on the balanced accuracy. This provides a more meaningful performance metric for groups of unequal sizes, such as the MCI and HC subject groups in this work.

Before performing classification experiments, the number of trees grown in each forest, t , and the number of features randomly selected at each tree node, d , had to be selected. Stable estimates of the out-of-bag classification error were consistently observed for $t \gtrsim 1,000$, and $t = 5,000$ was therefore used for all experiments. The value of d was consistently observed to have little effect on the out-of-bag classification error estimate. The value $d = \sqrt{D}$ was therefore used for all experiments, following the recommendation of [Liaw and Wiener \(2002\)](#).

6.4.2 Single-modality classification results

A random forest classifier was applied to the feature data from each modality independently, and the single-modality classification results obtained are presented in [Table 6.3](#).

	AD vs. HC		MCI vs. HC	
	MRI	FDG-PET	MRI	FDG-PET
Acc. (%)	84.4 (1.6)	87.9 (2.6)	64.4 (3.0)	63.9 (2.3)
Bacc. (%)	84.4 (3.3)	87.9 (4.3)	67.5 (4.9)	66.6 (3.7)
Sens. (%)	83.2 (3.5)	92.0 (3.9)	59.3 (4.5)	59.1 (2.8)
Spec. (%)	85.5 (3.0)	83.8 (4.7)	75.7 (5.2)	74.1 (4.6)

Table 6.3: Single-modality classification accuracy (acc.), balanced accuracy (bacc.), sensitivity (sens.) and specificity (spec.) based on the application of a random forest classifier to the original imaging data from 287 ADNI participants. Results are expressed as mean (standard error).

As described in [Section 3.2.4](#), estimates of the relative importances of the various features for classification may be extracted from the random forest. Feature importances for the two imaging modalities are shown in [Figure 6.2](#) for both clinical group pairs. The most important features for MRI include volumes of the hippocampus, amygdala, and other medial temporal lobe structures. The most important features for FDG-PET include signal intensities of voxels located in the posterior cingulate gyrus, parietal lobe, posterior temporal lobe, and around the hippocampus.

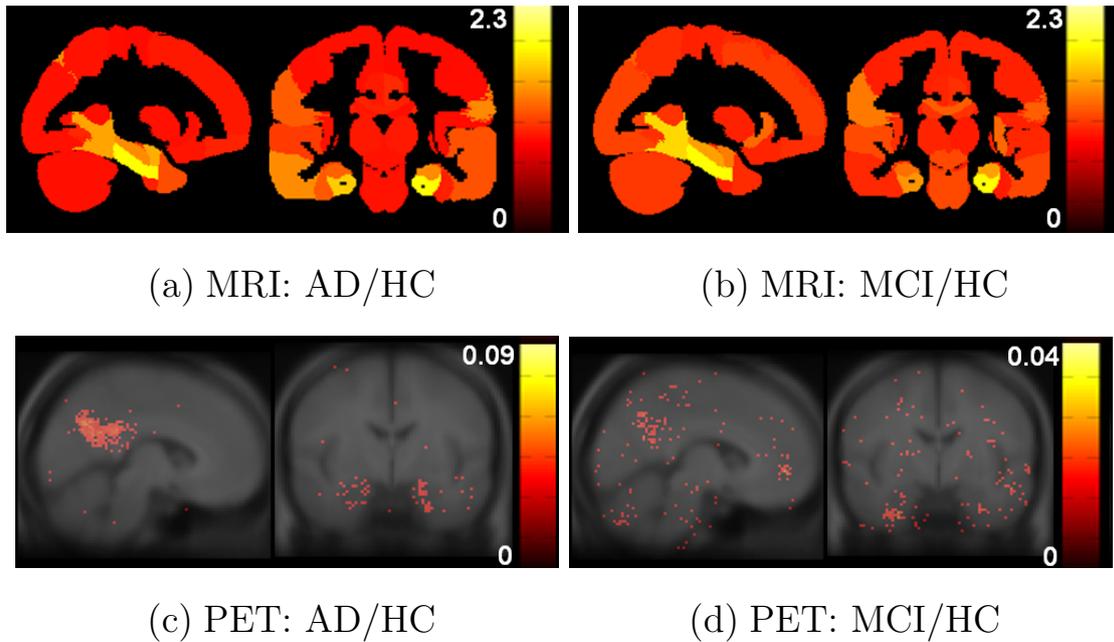


Figure 6.2: Feature importances for distinguishing between clinical groups using (a-b) region-based MRI, and (c-d) voxel-based FDG-PET. For MRI, regional feature importances are superimposed onto sagittal and coronal slices of a maximum probability brain atlas which has been masked in the same way as the anatomical segmentations. For FDG-PET, important voxels are overlaid onto sagittal and coronal slices of a MNI-space average MR image.

6.4.3 Single-modality similarity-based classification results

The random forest classifiers described in Section 6.4.2 were additionally used to derive pairwise similarity measures for each of the modalities, as described in Section 3.2.4. After applying MDS to the similarity matrix for each modality, the eigenvectors corresponding to the 25 largest-valued eigenvalues were used in generating the embeddings for classification. The value of 25 was empirically determined to ensure that zero-valued eigenvalues were not included, while capturing the maximum possible amount of information. Examples of the similarity matrices are shown later in Section 6.5.3.

A random forest classifier was applied to the embedded feature data from each of the two modalities independently, and the single-modality classification results obtained are presented in Table 6.4. The balanced accuracies achieved based on the embedding coordinates do not differ significantly from those achieved using the original imaging data (paired t-test, $p > 0.05$).

	AD vs. HC		MCI vs. HC	
	MRI	FDG-PET	MRI	FDG-PET
Acc. (%)	87.2 (2.0)	87.8 (2.6)	64.8 (3.0)	65.3 (1.9)
Bacc. (%)	87.2 (2.9)	87.8 (4.0)	65.2 (5.3)	65.3 (3.4)
Sens. (%)	87.5 (3.2)	91.8 (2.9)	64.8 (3.9)	65.3 (2.9)
Spec. (%)	86.9 (2.6)	83.8 (5.1)	65.5 (6.7)	65.2 (3.8)

Table 6.4: Single-modality classification accuracy (acc.), balanced accuracy (bacc.), sensitivity (sens.) and specificity (spec.) based on the application of a random forest classifier to the embedded imaging data from 287 ADNI participants. Results are expressed as mean (standard error).

6.4.4 Multi-modality similarity-based classification results

To generate a combined embedding that simultaneously incorporated information from both the FDG-PET and MR imaging data, the similarity matrices from the individual modalities were additively combined, and MDS applied to the resulting joint similarity matrix. A random forest classifier was then applied to the embedded feature data, and the multi-modality classification results obtained are presented in Table 6.5. Results based on the joint embedding out-performed the corresponding application to the separate embedding coordinates. The improvement in balanced accuracy was significant for the MCI/HC experiment (paired t-test, $p < 0.05$).

For comparison, the FDG-PET and MR imaging features were combined by simple concatenation, and a random forest classifier applied. The performance based on this feature set is also shown in Table 6.5. The balanced accuracy based on this feature set does not significantly differ from that based solely on the FDG-PET imaging features (paired t-test, $p > 0.05$).

	AD vs. HC		MCI vs. HC	
	Joint embedding	Concatenation	Joint embedding	Concatenation
Acc. (%)	90.0 (2.6)	87.9 (2.6)	75.5 (2.2)	64.3 (2.4)
Bacc. (%)	89.4 (3.6)	87.9 (4.3)	74.7 (3.0)	66.9 (3.9)
Sens. (%)	88.9 (3.4)	92.0 (3.9)	76.9 (3.2)	59.8 (3.2)
Spec. (%)	89.8 (3.8)	83.8 (4.7)	72.4 (4.5)	74.0 (4.6)

Table 6.5: Multi-modality classification accuracy (acc.), balanced accuracy (bacc.), sensitivity (sens.) and specificity (spec.) based on the application of a random forest classifier to the jointly embedded imaging data from 287 ADNI participants. Classification results based on the application of a random forest classifier to the concatenated imaging features are also shown. Results are expressed as mean (standard error).

6.4.5 Discussion

Classification based on the joint embedding constructed using both FDG-PET and MR imaging data out-performed classification based on either modality alone. This supports previous suggestions that there is complementary information between MRI and FDG-PET which can be exploited to produce a more powerful combined biomarker for AD (Walhovd et al., 2010; Landau et al., 2010). The classification accuracy for discriminating AD patients from HC based on the joint embedding is comparable with other state-of-the-art machine learning techniques which use either single-modality (Cuingnet et al., 2011) or multi-modality (Zhang et al., 2011; Hinrichs et al., 2011) imaging data. The accuracy for discriminating MCI patients from HC based on the joint embedding is also comparable with state-of-the-art multi-modality methods. For example, Zhang et al. (2011) report an accuracy of 76% using multi-kernel learning, but based on the combination of MRI, FDG-PET and CSF biomarkers. The lack of significant difference between classification performance based on the original imaging data and that based on the embedding coordinates for each individual modality is expected, since a random forest is already a nonlinear classifier. The motivation for the embedding step was to facilitate the incorporation of multi-modality data. It was shown that a simple concatenation of the FDG-PET and MR imaging features does not optimally combine these data, as this did not improve classification performance compared with the single modalities.

In the context of a neuroimaging application, one of the key benefits of random forests is that they provide estimates of the importances of the features for classification (Langs et al., 2011). This is valuable because it allows verification that the high-dimensional imaging features that contribute most to the classifier correspond to regions or structures with a biologically plausible connection to pathology. In this work, the most important features for discriminating between clinical groups correspond with those known to be visibly affected in AD on both FDG-PET and structural MR imaging (Hampel et al., 2008; Patwardhan et al., 2004). The important features for distinguishing between AD patients and HC are localised to affected regions, with the more challenging distinction between MCI patients and HC requiring features spread across a larger portion of the brain. The motivation for extracting voxel-based features from the FDG-PET

and region-based features from the MRI was to demonstrate that these two different types of imaging features could be readily combined using the proposed method.

6.5 Combining imaging and biological data

Following the success of the preliminary experiments described in Section 6.4, the multi-modality classification framework described in Section 6.3 was next applied for the combination of imaging and biological data. These data comprised ApoE genotype information and baseline FDG-PET, MRI and CSF measures acquired from the 147 ADNI participants described in Section 6.2. As well as extending the application of the method to biological data, improvements were made to the cross-validation and embedding steps, and an additional classification experiment (pMCI/sMCI) was included. Details of the methodology and results obtained from both single- and multi-modality classification experiments are presented in the following subsections.

6.5.1 Classification methodology

Classification performance was assessed between three clinically relevant pairs of diagnostic groups (AD/HC, MCI/HC, pMCI/sMCI) based on both single- and multi-modality imaging and biological data. Robust estimates of classifier performance were obtained using the stratified repeated random sampling approach described in Section 3.3.2, with the same folds used for all experiments. The mean accuracy, balanced accuracy, sensitivity and specificity were evaluated over 100 runs in which 75% of the data were selected for training, with the remaining 25% used as test data. Values of $t = 5,000$ and $d = \sqrt{D}$ were selected as described in Section 6.4.1.

6.5.2 Single-modality classification results

A random forest classifier was applied to the feature data from each modality independently, and the single-modality classification results obtained are presented in Table 6.6. As described in Section 3.2.4, estimates of the relative importances of the various features for classification

may be extracted from the random forest. Feature importances for the two imaging modalities are shown in Figure 6.3 for all three clinical group pairs. The most important features for both MRI and FDG-PET are as described in Section 6.4.2.

		CSF	MRI	FDG-PET	Genetic
AD/HC	Acc. (%)	76.5 (0.8)	81.6 (0.8)	86.0 (0.7)	72.6 (0.9)
	Bacc. (%)	76.8 (1.3)	81.8 (1.3)	86.0 (1.2)	72.7 (1.3)
	Sens. (%)	73.0 (1.3)	79.8 (1.3)	86.8 (1.1)	71.3 (1.3)
	Spec. (%)	80.5 (1.3)	83.8 (1.3)	85.1 (1.3)	74.1 (1.4)
MCI/HC	Acc. (%)	63.1 (0.8)	66.9 (0.9)	66.5 (0.8)	73.8 (0.5)
	Bacc. (%)	63.8 (1.4)	68.9 (1.3)	66.9 (1.3)	60.7 (0.9)
	Sens. (%)	62.0 (1.1)	63.7 (1.2)	65.7 (1.1)	94.7 (0.5)
	Spec. (%)	65.5 (1.7)	74.0 (1.4)	68.1 (1.5)	26.6 (1.2)
pMCI/sMCI	Acc. (%)	52.9 (1.0)	55.1 (1.0)	52.6 (1.0)	47.3 (0.9)
	Bacc. (%)	53.5 (1.6)	55.5 (1.7)	53.1 (1.7)	42.4 (2.4)
	Sens. (%)	58.1 (1.6)	59.1 (1.8)	57.4 (2.0)	32.1 (2.1)
	Spec. (%)	48.8 (1.5)	51.9 (1.6)	48.8 (1.5)	52.6 (2.7)

Table 6.6: Single-modality classification accuracy (acc.), balanced accuracy (bacc.), sensitivity (sens.) and specificity (spec.) based on the application of a random forest classifier to the original imaging and biological data from 147 ADNI participants. Results are expressed as mean (standard error).

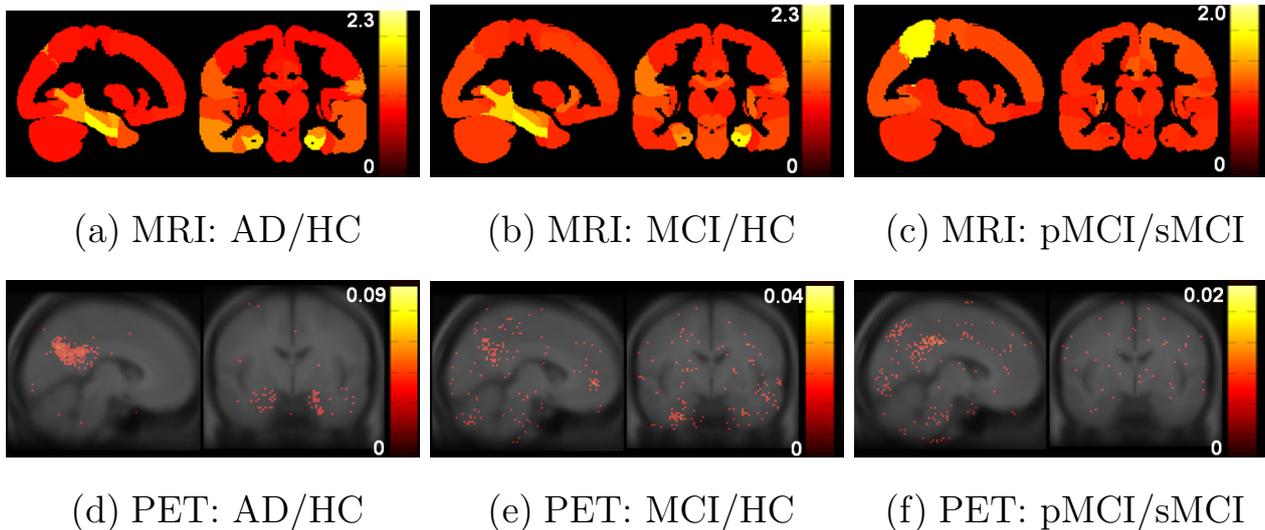


Figure 6.3: Feature importances for distinguishing between clinical groups using (a-c) region-based MRI, and (d-f) voxel-based FDG-PET. For MRI, regional feature importances are superimposed onto sagittal and coronal slices of a maximum probability brain atlas which has been masked in the same way as the anatomical segmentations. For FDG-PET, important voxels are overlaid onto sagittal and coronal slices of a MNI-space average MR image.

6.5.3 Single-modality similarity-based classification results

The random forest classifiers described in Section 6.5.2 were additionally used to derive pairwise similarity measures for each of the four modalities, as described in Section 3.2.4. Examples of the resulting similarity matrices are shown in Figure 6.4. The categorical nature of the genetic information results in similarity matrices which have an almost binary structure.

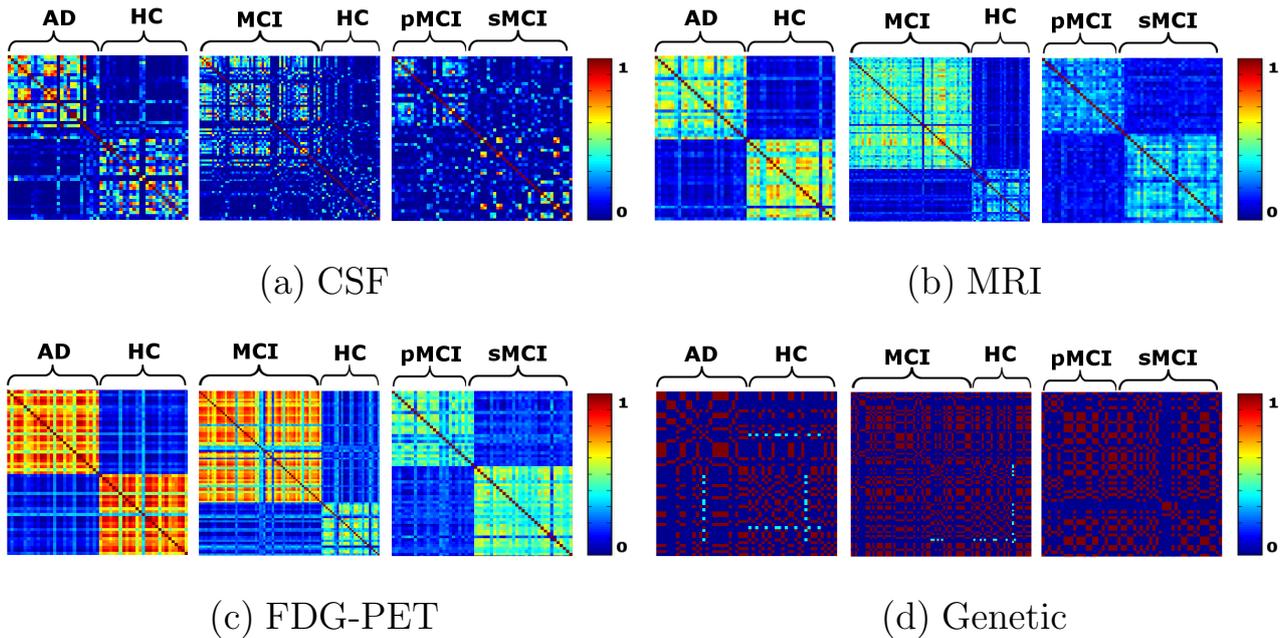


Figure 6.4: Similarity matrices for all three clinical group pairs based on (a) CSF biomarker measures, (b) regional MRI volumes, (c) voxel-based FDG-PET signal intensities and (d) categorical ApoE genotype information. The matrices are symmetric, and each entry represents the similarity between a pair of subjects based on the input feature data.

MDS was applied to each similarity matrix, and a goodness-of-fit value of 90% was used to determine an appropriate dimensionality for the resulting embeddings, as described in Section 3.2.4. A random forest classifier then was applied to the embedded feature data from each of the four modalities independently, and the single-modality classification results obtained are presented in Table 6.7, along with the dimensionality of each embedding.

No consistent differences were observed between the balanced accuracies based on the embedded imaging and biological feature data shown in Table 6.7, and those based on the original feature data shown in Table 6.6.

		CSF	MRI	FDG-PET	Genetic
AD/HC	Acc. (%)	76.1 (0.8)	82.5 (0.7)	86.4 (0.7)	72.6 (0.9)
	Bacc. (%)	76.3 (1.3)	82.1 (1.4)	86.5 (1.2)	72.7 (1.3)
	Sens. (%)	72.8 (1.3)	88.6 (1.2)	85.8 (1.2)	71.3 (1.3)
	Spec. (%)	79.8 (1.4)	75.6 (1.5)	87.1 (1.3)	74.1 (1.4)
	k	13	22	9	2
MCI/HC	Acc. (%)	61.7 (0.8)	67.3 (1.0)	53.5 (0.7)	73.8 (0.5)
	Bacc. (%)	61.7 (1.3)	69.1 (1.4)	60.2 (1.2)	60.7 (0.9)
	Sens. (%)	61.6 (1.1)	64.3 (1.3)	42.3 (1.1)	94.7 (0.5)
	Spec. (%)	61.8 (1.5)	73.9 (1.4)	78.0 (1.3)	26.6 (1.2)
	k	25	47	35	2
pMCI/sMCI	Acc. (%)	52.1 (1.0)	58.4 (1.0)	53.0 (1.0)	43.5 (0.9)
	Bacc. (%)	52.7 (1.7)	58.3 (1.7)	52.8 (1.7)	41.2 (2.4)
	Sens. (%)	57.9 (1.6)	56.9 (1.6)	50.6 (1.8)	27.4 (2.0)
	Spec. (%)	47.5 (1.7)	59.7 (1.8)	54.9 (1.6)	55.0 (2.7)
	k	21	38	35	1

Table 6.7: Single-modality classification accuracy (acc.), balanced accuracy (bacc.), sensitivity (sens.) and specificity (spec.) based on the application of a random forest classifier to the embedded imaging and biological data from 147 ADNI participants. Results are expressed as mean (standard error). The mean dimensionality of each embedding (k) is also shown.

6.5.4 Multi-modality similarity-based classification results

A joint similarity matrix \mathbf{S} was defined as a linear combination of the similarity matrices from each of the four modalities \mathbf{S}_i . Each modality was assigned a weighting factor α_i , such that

$$\mathbf{S} = \sum_{i=1}^4 \alpha_i \mathbf{S}_i, \text{ where } \sum_{i=1}^4 \alpha_i = 1.$$

To ensure the best combination of the four modalities for classification, the α_i parameters were optimised as part of the training process. This was achieved by performing a grid-search within the training data, and selecting the set of parameters resulting in the highest cross-validated accuracy. The classifier was then trained using this set of parameters, before having its performance assessed on the test data. For each of the classification experiments, the distribution of parameters selected over the 100 runs is illustrated in Figure 6.5.

MDS was applied to the joint similarity matrix constructed using information from all four modalities, and a goodness-of-fit value of 90% was again used to determine an appropriate

dimensionality for the resulting embedding. A random forest classifier was applied to the embedded feature data, and the multi-modality classification results obtained are presented in Table 6.8.

	Acc. (%)	Bacc. (%)	Sens. (%)	Spec. (%)
AD/HC	89.0 (0.7)	89.0 (1.2)	87.9 (1.2)	90.0 (1.1)
MCI/HC	74.6 (0.8)	72.7 (0.8)	77.5 (1.0)	67.9 (1.7)
pMCI/sMCI	58.0 (0.9)	57.9 (1.7)	57.1 (1.8)	58.7 (1.5)

Table 6.8: Multi-modality classification accuracy (acc.), balanced accuracy (bacc.), sensitivity (sens.) and specificity (spec.) based on the application of a random forest classifier to the jointly embedded imaging and biological data from 147 ADNI participants. Results are expressed as mean (standard error).

The balanced accuracies based on multi-modality classification are significantly (paired t-test, $p < 0.01$) higher than those based on any individual modality for both the AD/HC and MCI/HC experiments. For the pMCI/sMCI experiment, however, the balanced accuracy based on multi-modality classification is not significantly different from that based on MRI information alone.

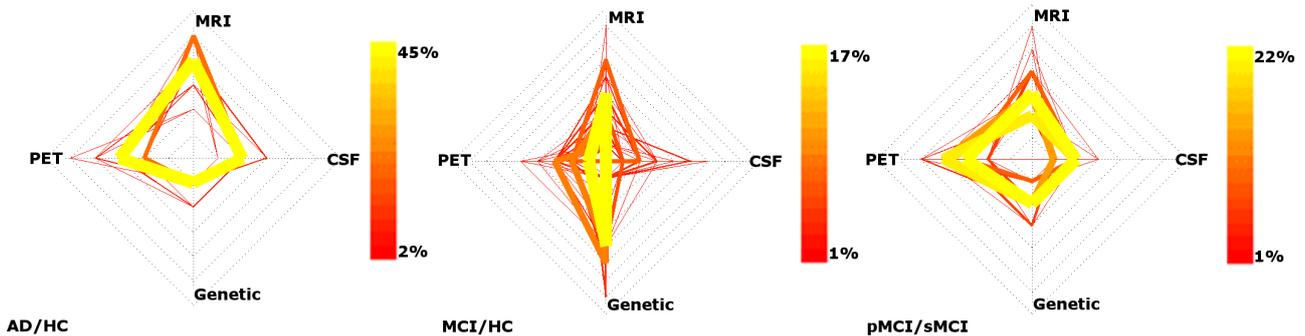


Figure 6.5: Cobweb plots showing the distribution of parameters selected over the 100 leave-25%-out runs for all three classification experiments. The four spokes of each plot represent the four modalities, and each coloured line connecting the four spokes represents a set of parameter values. The colour and weight of each line represents the percentage of runs in which the associated parameter set was selected.

6.5.5 Discussion

Classification based on the joint embedding constructed using information from all four modalities was superior to classification based on any individual modality for comparisons between AD patients and HC, as well as between MCI patients and HC. This lends further support to previous suggestions that there is some complementary information between these neuroimaging and

biological measures which can be exploited to produce a more powerful combined biomarker for AD and MCI (Walhovd et al., 2010; Landau et al., 2010).

In terms of accuracy, 89% classification was achieved between AD patients and HC, and 75% between MCI patients and HC. These results are comparable with the 92% and 93% accuracies reported between AD patients and HC in Hinrichs et al. (2011) and Zhang et al. (2011), respectively, as well as the 76% accuracy reported between MCI patients and HC in Zhang et al. (2011). Both of these studies employ a kernel combination framework for multi-modality classification, and this work proposes an alternative approach which achieves comparable results. Random forests provide consistent pairwise similarity measures for multiple modalities, thus facilitating the combination of different types of feature data. This has been demonstrated using datasets in which the number of features differ by several orders of magnitude. Since random forest classifiers extend naturally to multi-class problems, the framework described here could be used for other applications in the future, such as the differential diagnosis of AD. In addition, the implementation of random forests used in this work could be easily modified to produce uncertainty information about the predicted diagnostic labels. This could be achieved by having each leaf node store a probabilistic distribution of labels, rather than a point estimate. Criminisi et al. (2012), for example, describes this and other extensions to the original random forests algorithm, and presents a unified model of random decision forests for classification, regression, density estimation, manifold learning, and semi-supervised learning.

In Section 6.4, comparable multi-modality classification results were achieved using only information extracted from the two imaging modalities (accuracies of 90% for AD/HC and 76% for MCI/HC). The lack of improvement over these results is likely to be attributable to the considerable reduction in size of the subject group as a result of the requirement for CSF biomarker information. A total of 287 ADNI participants had both FDG-PET and MR imaging data available, but CSF biomarker measures were additionally available for only 147. A more robust form of cross-validation was also applied in this section, which uses stratified repeated random sampling as opposed to the single round of ten-fold cross-validation employed in Section 6.4.

The motivation for the embedding step was to facilitate the incorporation of multi-modality

data. A goodness-of-fit parameter of 90% was used to determine an appropriate dimension for the embeddings which would reduce noise in the data. This method of determining the dimensionality of the embeddings was more principled than the approach employed in Section 6.4, in which the 25 largest-valued eigenvalues were retained. The lack of significant difference between classification performance based on the original feature data and that based on the embedding coordinates for each individual modality was expected, since a random forest is already a nonlinear classifier. However, a difference was observed for the comparison between MCI patients and HC based on the voxel-based FDG-PET features. This may be due to the inhomogeneity of the MCI group, which comprises both pMCI and sMCI patients. It is possible that the high-dimensional voxel-based FDG-PET features are sensitive to differences in the pattern of glucose metabolism between these two groups, resulting in a reduced classification performance based on the associated embedding coordinates.

Random forests are ensemble-based classifiers that are often applied to high-dimensional datasets. Here, random forests are also applied to low-dimensional biological data so that consistent pairwise similarity measures may be obtained for all modalities. In the case of a single feature, such as the categorical genetic information, a random forest reduces to bootstrap aggregation.

Visualisation of the parameters selected to combine similarities for multi-modality classification (Figure 6.5) provides some interesting insights into the relationships among the modalities. The figure indicates the optimum way in which to combine MRI, FDG-PET, CSF and genetic information within the framework described. For distinguishing between AD patients and HC, for example, it appears that FDG-PET and MR imaging features provide the most complementary information. This indication supports the hypothetical temporal model of biomarker dynamics shown in Figure 1.10, in which FDG-PET and MRI measures show the greatest difference between AD patients and cognitively normal individuals. For distinguishing MCI patients from HC, genetic information appears to have a relatively high importance. However, the optimum modality weightings for distinguishing between these groups are less stable than those for distinguishing between AD patients and HC. This may be because the heterogeneity of the MCI group makes their selection dependent on the proportions of pMCI and sMCI patients in the training set. The figure suggests an interesting avenue for further research, in that estimates

of inter-modality correlations could help to determine the amount of complementary information between them. This could facilitate decisions on how to acquire the maximum amount of diagnostically relevant information for a patient using a minimum number of assessments.

The classification performance between pMCI and sMCI patients is not significantly improved by combining multi-modality information in this study. The results presented in Chapter 5 have shown that incorporating longitudinal information can be beneficial to improve the ability to distinguish between these two groups. This is another avenue for future research. It is also important to consider, however, that progression from MCI to AD occurs at a rate of 10-15% per year (Petersen et al., 1999), with up to 80% of MCI patients developing AD over a six year period (Petersen, 2004). Longer clinical follow-up is therefore required to properly assess the utility of any classification method in separating pMCI from sMCI patients.

6.6 Conclusion

This chapter has presented a framework for multi-modality classification based on pairwise similarity measures derived from random forests. Random forests provide consistent pairwise similarity measures for multiple modalities, thus facilitating the combination of different types of feature data. The approach has been evaluated using imaging and biological data from the ADNI study, including voxel-based FDG-PET and region-based MR imaging data, CSF biomarker measures, and categorical ApoE genotype information. Classification based on multiple modalities has been shown to out-perform that based on any individual modality. This finding supports previous suggestions that there is some complementary information between neuroimaging and biological measures which can be exploited to produce a more powerful combined biomarker for AD and MCI.

Several areas for further research have been identified. Methodologically, the approach is generalisable, in that the manifold learning and classification steps could be performed using alternative algorithms, and similarities could be extracted or combined using more sophisticated metrics. Additionally, longitudinal data could be incorporated, as well as clinical and neuropsych-

chological information. Since random forest classifiers extend naturally to multi-class problems, the framework described here could be used for other applications in the future, such as the differential diagnosis of AD. In addition, the implementation of random forests used in this work could be modified to produce uncertainty information about the predicted diagnostic labels. This may be more useful to clinicians than a simple binary prediction.

Chapter 7

Early identification of Alzheimer's disease

Work in this chapter will, in part, be presented in:

K. R. Gray, R. Wolz, R. A. Heckemann, D. Rueckert, and A. Hammers. Structural differences in cognitively normal elderly individuals with abnormal amyloid biomarkers: detection using volumetric MRI in ADNI and AIBL. (Abstract) Alzheimer's Association International Conference (AAIC'12), 2012 (Accepted).

7.1 Introduction

This chapter presents findings of early signs of neurodegeneration in cognitively normal individuals at high risk of developing AD. The ability to identify individuals in the pre-symptomatic stages of AD is desirable for early diagnosis, intervention, counselling, and drug discovery. As described in Section 1.5, pathological changes in the brain precede cognitive symptoms by several years. Disease-specific imaging biomarkers could potentially detect pre-clinical disease. The majority of efforts aimed at early detection have either focused on MCI patients, or required serial imaging over 12 months (for example, [Schott et al. \(2010\)](#)).

In this work, multi-region analysis of MR images acquired at a single timepoint was used to identify early signs of neurodegeneration in cognitively normal elderly individuals with evidence of cortical β -amyloid deposition. Imaging and biological data used in this chapter were obtained from cognitively normal participants in both the ADNI study, and the Australian Imaging, Biomarkers & Lifestyle Flagship Study of Ageing (AIBL; <http://www.aibl.csiro.au>). Section 7.2 describes these two cohorts. Section 7.3 presents details of the MRI acquisition, pre-processing, and anatomical segmentation procedure for both subject groups. Section 7.4 describes the assignment of the participants to risk groups, based on CSF $A\beta$ in ADNI, and neocortical-to-cerebellar ratios of amyloid deposition on PiB-PET imaging in AIBL. Section 7.5 then presents the results of comparisons between regional MR volumes in the high- and low-risk sub-sets of each group.

Reduced volumes in temporo-parietal and orbito-frontal regions in high-risk individuals were observed in both cohorts. These differences could be indicative of very early changes associated with AD. The consistency of regional differences observed in the two independent groups suggests that volumetric MRI can reveal structural brain changes that precede the onset of clinical symptoms. It may therefore be useful in identifying early signs of neurodegeneration in healthy elderly individuals, potentially providing a useful early screening tool, or outcome measure for clinical trials.

7.2 Imaging and biological data

Clinical, imaging and biological data used in this chapter were obtained from cognitively normal participants enrolled in both the ADNI and AIBL studies. Groupwise characteristics of these subjects are presented in Table 7.1. The mean age and MMSE score do not vary significantly ($p > 0.01$) on t-test between the ADNI and AIBL groups. Further information about each cohort is provided in the following subsections, and groupwise characteristics according to amyloid-based risk status are provided later in Section 7.4.

	N(F)	Age (mean \pm std. dev.)	MMSE Score (mean \pm std. dev.)	CDR (%)			M18 conversion
				0	0.5	1	
ADNI	109(55)	75.8 \pm 5.2	29.1 \pm 1.0	100	0	0	2 (2%)
AIBL	119(63)	73.2 \pm 7.2	28.8 \pm 1.2	91	8	1	4 (3%)

Table 7.1: Clinical and demographic information for the study populations. For both ADNI and AIBL, the total number of HC subjects (N) and number of females (F) are shown, along with the average age, average MMSE score, and CDR distribution. The number and percentage of subjects who had progressed to either MCI or AD within 18 months (M18 conversion) are also provided.

7.2.1 ADNI participants

The ADNI database has been described in Section 4.2. Although 3 T MR images are available for a sub-set of ADNI participants, 1.5 T MR images are available for all. Baseline 1.5 T MR images and CSF measures of $A\beta$ were available to download for 109 cognitively normal ADNI participants (as of January 2012). Up to 48 months of clinical follow-up data were available for these participants, during which eight subjects had progressed to either MCI or AD. Two subjects progressed during their first 18 months of clinical follow-up.

7.2.2 AIBL participants

The AIBL study aims to improve understanding of the pathogenesis, early clinical manifestation, and diagnosis of AD, as well as to identify diet and lifestyle factors that influence disease development. It is a longitudinal study of 1,112 subjects, including 211 AD patients, 133 MCI patients, and 768 HC. Details of the methodology have been presented in [Ellis et al. \(2009\)](#).

Core funding for AIBL was provided by the Australian Commonwealth Scientific Industrial and Research Organisation (CSIRO). This was supplemented by contributions from several leading researchers and research organisations located in the Australian cities of Melbourne and Perth. AIBL participants have been enrolled at one of five sites in these two cities. At baseline, all participants underwent a screening interview, had comprehensive cognitive testing, gave blood for biomarker analysis, and completed health and lifestyle questionnaires. Approximately one quarter of participants underwent neuroimaging with PiB-PET and structural MRI. Partici-

pants were recruited between late 2006 and mid 2008, and are being followed up at 18-month intervals, as described in [Ellis et al. \(2010\)](#). Clinical and imaging data from the AIBL study are available to download from the LONI image data archive. Baseline PiB-PET and MR images were available for 119 cognitively normal AIBL participants (as of January 2012). Four subjects had progressed to either MCI or AD during their first 18 months of clinical follow-up.

7.3 MRI acquisition and anatomical segmentation

The aim of this work was to perform comparisons between regional MR volumes extracted from groups of HC differing in amyloid-based risk status for the development of AD. Multi-region anatomical segmentations were therefore required for both the ADNI and AIBL subject groups. For the ADNI participants, automatic whole-brain segmentations into 83 anatomical regions had already been prepared in the native space of each baseline MRI using MAPER, as described in Section 4.4. Details of the acquisition and anatomical segmentation of the AIBL MR images are presented in the following subsections, as well as computation of the regional MR volumes.

7.3.1 AIBL MRI acquisition

Baseline T1-weighted MRI scans for the 119 cognitively normal AIBL participants were downloaded from the LONI image data archive in NIfTI format. As described in [Ellis et al. \(2010\)](#), the MRI parameters used in the ADNI study had been adopted for the neuroimaging stream of the AIBL study. However, in contrast with the multi-centre ADNI MRI acquisition, AIBL MR images had all been acquired at one of two sites, using either a 1.5 T ($n = 39$) or 3 T ($n = 80$) Siemens MRI scanner.

7.3.2 AIBL MRI anatomical segmentation

Automatic whole-brain segmentations of the baseline AIBL MR images were prepared in native MRI space using MAPER as described for the ADNI group in Section 4.4. Additional image

processing was first required for bias field correction and brain extraction.

Bias field correction

Although the AIBL MR images had been acquired and processed according to the ADNI procedure described in Section 4.3.3, some images appeared inhomogeneous upon visual review. Three images were randomly selected, and corrected for field inhomogeneity using the EM segmentation tool from NiftySeg (<http://niftyseg.sourceforge.net>). The corrected images were reviewed both directly and after subtraction from the originals. This showed that substantial amounts of typical field bias had been removed, as illustrated in Figure 7.1. The bias correction procedure was therefore applied to all images.

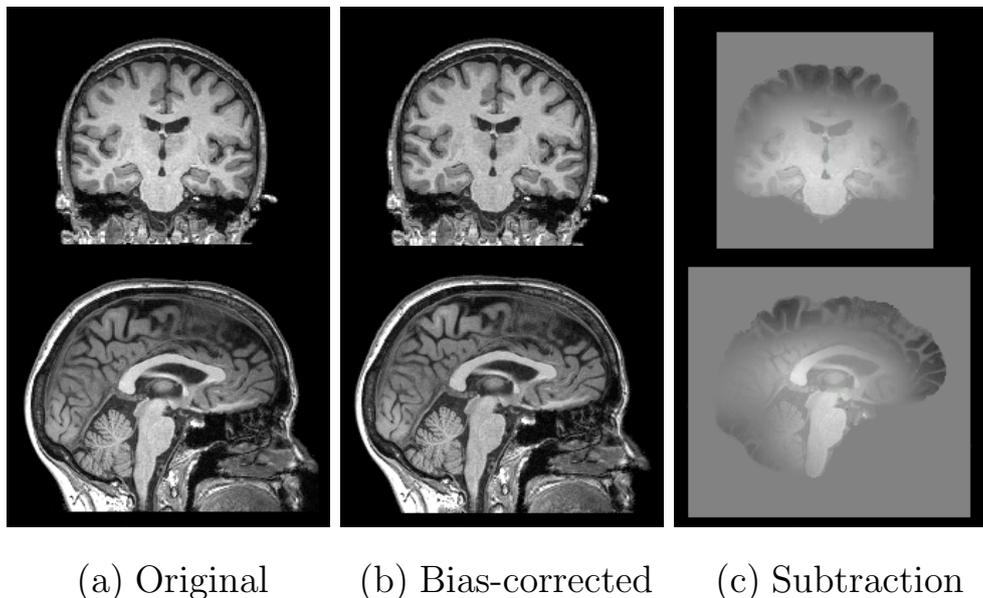


Figure 7.1: Illustration of the bias correction procedure applied to the AIBL MR images, showing (a) the original MRI, (b) the bias-corrected MRI, and (c) the corrected image after subtraction from the original.

Brain extraction

A coarse brain extraction was performed on each of the bias-corrected MR images using FSL BET (Smith, 2002). This was iteratively refined using a multi-level, multi-resolution method named PINCRAM (Pyrimidal INtraCRAnial Masking), developed by Rolf Heckemann. At a

single iteration, PINCRAM follows a similar approach to the multi-atlas segmentation method described in Section 2.3.3. The atlases comprised 39 randomly selected baseline ADNI MR images and their corresponding intracranial masks, which had been generated as described in Section 4.4.1. The multi-atlas brain masking procedure employed at a single PINCRAM iteration is illustrated in Figure 7.2.

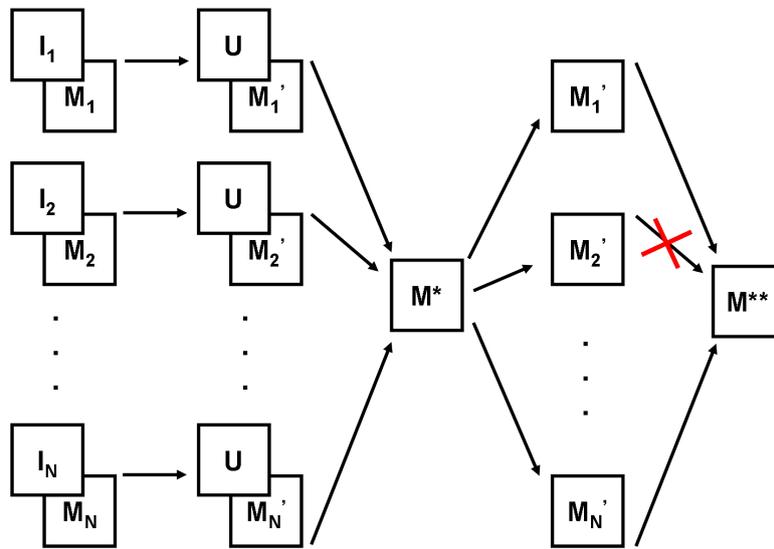


Figure 7.2: Schematic representation of a single PINCRAM iteration. A set of atlas anatomical images I_i is registered to the unseen image U , and the corresponding atlas intracranial masks M_i propagated to the unseen image. The transformed masks M'_i are then fused to generate the intermediate consensus mask M^* . Overlap measures between the transformed masks and this consensus mask are used to determine a sub-set of masks most similar to the consensus. This sub-set of masks are fused to generate the final consensus mask M^{**} for the iteration.

A refined extraction of the bias-corrected MR image was performed after each PINCRAM iteration, based on the consensus brain mask generated. Four iterations of the multi-atlas brain masking procedure illustrated in Figure 7.2 were employed. As the PINCRAM procedure progresses, more detailed transformations are applied to align the atlases with the target MR image, and smaller sub-sets of atlases are retained for generation of the consensus brain mask. Details of the four iterations used in this work are summarised in Table 7.2.

Images illustrating the refinement of the BET-based brain extraction achieved using the iterative PINCRAM procedure are shown in Figure 7.3. The final consensus brain mask was used to perform the detailed brain extraction required for multi-atlas segmentation using MAPER.

Iteration	Transformation model	Image resolution	Number of atlases retained
1	Global affine	4 mm	31
2	Coarse nonrigid (20 mm)	4 mm	27
3	Moderate nonrigid (8 mm)	2 mm	17
4	Detailed nonrigid (3 mm)	full	11

Table 7.2: Details of the four PINCRAM iterations employed to generate a refined brain extraction of the bias-corrected MRI. As the PINCRAM procedure progresses, the transformation models used to register the atlases to the target become more detailed, the images undergo less smoothing, and a smaller sub-set of atlases is retained for use in the next iteration. For the nonrigid transformations, numbers in brackets indicate the control point spacings.

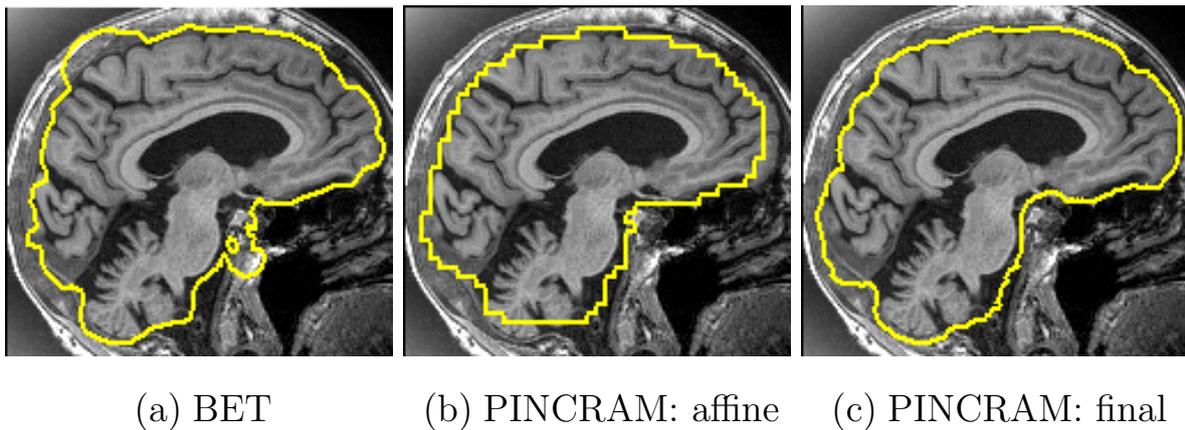


Figure 7.3: Illustration of the PINCRAM procedure, showing (a) the initial brain mask obtained using BET, (b) the brain mask obtained following the affine PINCRAM iteration, and (c) the brain mask obtained following the final PINCRAM iteration. Masks are shown overlaid onto the corresponding bias-corrected MR images.

7.3.3 Computation of regional MRI volumes

MRI volumes were computed for all 83 anatomically defined regions for both the ADNI and AIBL groups. To reduce inter-individual variation, regional volumes were normalised by the total ICV computed from the masks described in Sections 4.4.1 and 7.3.2. This has been shown to substantially reduce variation, remove gender-related differences (Whitwell et al., 2001), and eliminate inaccuracies arising from scaling or voxel size errors remaining after phantom correction of the MRI (Clarkson et al., 2009).

The MAPER segmentation procedure has been shown to produce consistent results across field strengths of 1.5 T and 3 T in ADNI (Heckemann et al., 2011). However, significant differences ($p < 0.05$) in volume were observed in over half of the 83 regions between the 1.5 T and 3 T

AIBL MR images, both before and after accounting for the effects of subject age using a linear regression model. These differences could be attributed to a systematic difference in image processing between scanners, or in the demographic characteristics of the subjects. The AIBL MR images acquired using the two scanners were therefore not pooled in this work, but divided into two groups which were analysed independently (see Section 7.4.2).

7.4 Amyloid-based risk status

The cognitively normal ADNI and AIBL participants were assigned to risk groups for the development of AD based on evidence of their cortical β -amyloid deposition. This was assessed using CSF $A\beta_{1-42}$ measurements in ADNI, and neocortical-to-cerebellar ratios of amyloid deposition on PiB-PET imaging in AIBL. As described in Section 1.5.4, these two measures show a highly significant inverse correlation (Fagan et al., 2006). Low CSF $A\beta$ measures and high neocortical-to-cerebellar ratios of amyloid deposition on PiB-PET imaging are associated with a high risk for the development of AD.

A CSF $A\beta_{1-42}$ cutoff of 192 pg/mL has been estimated from an autopsy-confirmed sample as the best discriminator between AD patients and HC (Shaw et al., 2009). This has been applied to distinguish between normal and abnormal CSF $A\beta$ measures among ADNI participants (Schott et al., 2010). An analogous cutoff of 1.5 on a neocortical-to-cerebellar ratio of amyloid deposition on PiB-PET imaging has also been found useful in ADNI (Weigand et al., 2011). Details of the assignment of the ADNI and AIBL participants to risk groups for the development of AD based on these two analogous cutoffs are presented in the following subsections.

7.4.1 ADNI participants

The 109 cognitively normal ADNI participants were assigned to high- and low-risk groups for the development of AD based on CSF measures of $A\beta_{1-42}$ (high risk \leq 192 pg/mL). These data had been acquired as described in Section 6.2.3, and are summarised in Figure 7.4.

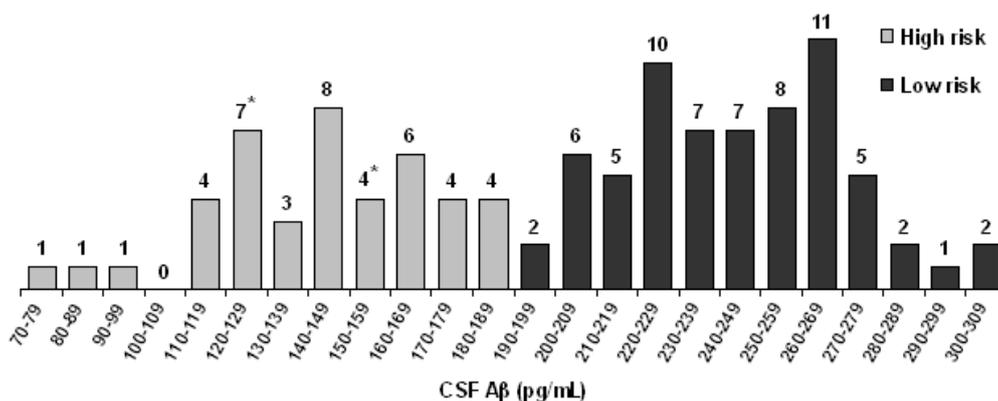


Figure 7.4: Summary of the ADNI CSF A β measures. Ranges containing values associated with the two subjects who progressed to either MCI or AD within 18 months are indicated by asterisks (*).

Following sub-division of the ADNI participants, 39% were classified as having a high risk for the development of AD. The mean ages of the high- and low-risk sub-sets (76.6 ± 5.1 and 75.3 ± 5.2 , respectively) were not significantly different (t-test, $p > 0.01$). The two subjects who had progressed to either MCI or AD within 18 months were both assigned to the high-risk group based on their CSF A β measures.

7.4.2 AIBL participants

The 119 cognitively normal AIBL participants were assigned to high- and low-risk groups for the development of AD based on neocortical-to-cerebellar ratios of amyloid deposition on PiB-PET imaging (high risk > 1.5). The following subsections describe the PiB-PET image acquisition and processing, as well as computation of the neocortical-to-cerebellar ratios.

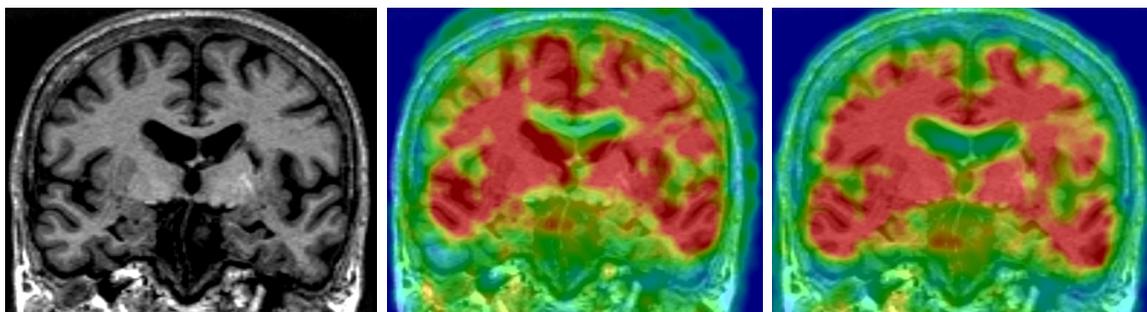
AIBL PiB-PET acquisition and pre-processing

Baseline pre-processed PiB-PET images for the 119 cognitively normal AIBL participants were downloaded from the LONI image data archive in NIfTI format. These had been acquired according to a standard protocol using a Philips Allegro PET scanner at both neuroimaging sites. A 30-minute dynamic scan consisting of six 5-minute frames was acquired, beginning approximately 40 minutes after the intravenous injection of ~ 370 MBq of PiB. Data were

corrected for both scatter and measured attenuation, determined using a transmission scan, and images were iteratively reconstructed. The six frames were summed to produce a single 30-minute static image, and these summed images were available to download. PiB-PET imaging was performed according to the methodology described in [Pike et al. \(2007\)](#).

Co-registration of PiB-PET with MRI

Each of the 119 summed baseline PiB-PET images was affinely co-registered with its corresponding native space baseline MR image, and re-sampled to the higher resolution of the MRI using linear interpolation. Registration was performed as described in Section 4.3.4, and is illustrated in Figure 7.5.



(a) Native space MRI (b) Native space PET (c) MRI-space PET

Figure 7.5: Illustration of PET-MRI co-registration, showing (a) the native space MR image overlaid with (b) the native space PiB-PET image, and (c) the MRI-space PiB-PET image.

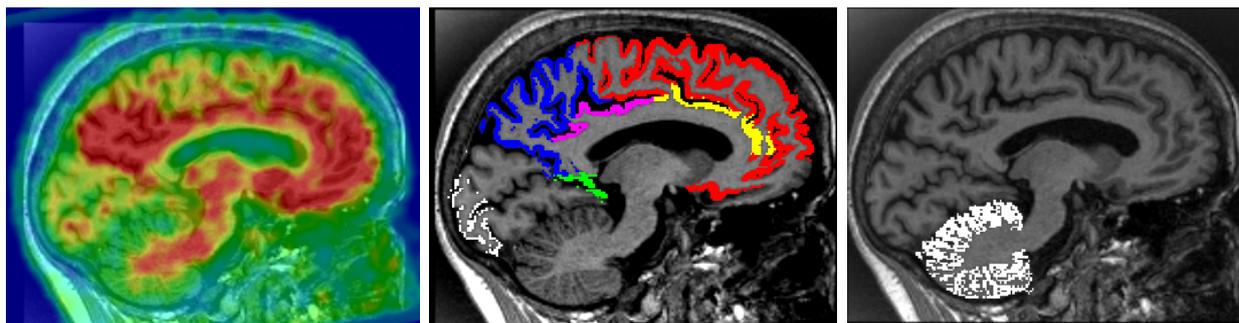
Neocortical-to-cerebellar ratios of amyloid deposition

The extent of cortical β -amyloid deposition was assessed on PiB-PET imaging using the SUVR analysis method described in Section 1.4.1. Each of the MRI-space PiB-PET images was overlaid with its corresponding masked anatomical segmentation, and the SUVR computed between a composite neocortical region and the cerebellar grey matter, similarly to the method described in [Rowe et al. \(2010\)](#).

The neocortical SUV was computed as the volume-weighted average PiB-PET signal intensity per mm^3 across six regions: frontal, superior parietal, lateral temporal, lateral occipital, anterior

cingulate, and posterior cingulate. These regions were derived from the 83 region anatomical segmentations as follows. The frontal region comprised the middle, inferior and superior frontal gyri; anterior, lateral and posterior orbital gyri; precentral and straight gyri; subgenual and pre-subgenual frontal cortices; and the subcallosal area. The superior parietal region comprised the postcentral gyrus, and superior parietal gyrus. The lateral temporal region comprised the lateral part of the anterior temporal lobe; anterior and central parts of the superior temporal gyrus; middle and inferior temporal gyri, and the posterior temporal lobe.

The cerebellar grey matter was identified by masking the full label with the binary maximum-probability grey matter map generated using FAST as described in Section 4.4.1. The cerebellar grey matter provides a suitable reference region due to the lack of PiB accumulation in either AD patients or HC (Klunk et al., 2004). Examples of the images required for computation of neocortical-to-cerebellar ratios of amyloid deposition are shown in Figure 7.6.



(a) PiB-PET

(b) Neocortical regions

(c) Reference region

Figure 7.6: Examples of the images required for assessment of the extent of cortical β -amyloid deposition, showing (a) the PiB-PET image, (b) the composite neocortical regions, and (c) the cerebellar grey matter. Images are overlaid onto the corresponding MRI. Neocortical regions shown comprise frontal (red), superior parietal (blue), lateral temporal (green), lateral occipital (white), anterior cingulate gyrus (yellow), and posterior cingulate gyrus (purple).

The 119 cognitively normal AIBL participants were divided according to scanner model (1.5 T, $n = 39$; 3 T, $n = 80$) as explained in Section 7.3.3. Participants were then assigned to high- and low- risk groups for the development of AD based on their PiB-PET SUVR measures (high risk > 1.5). These are summarised in Figure 7.7 for the 39 participants scanned using the 1.5 T MR scanner. Following sub-division of these participants, 21% were classified as having a high risk for the development of AD. The mean ages of the high- and low-risk sub-sets (74.0 ± 6.9

and 69.3 ± 6.2 , respectively) were not significantly different (t-test, $p > 0.01$). One of the four subjects who had progressed to either MCI or AD within 18 months was scanned using the 1.5 T MR scanner, and assigned to the low-risk group based on their PiB-PET SUVR measure.

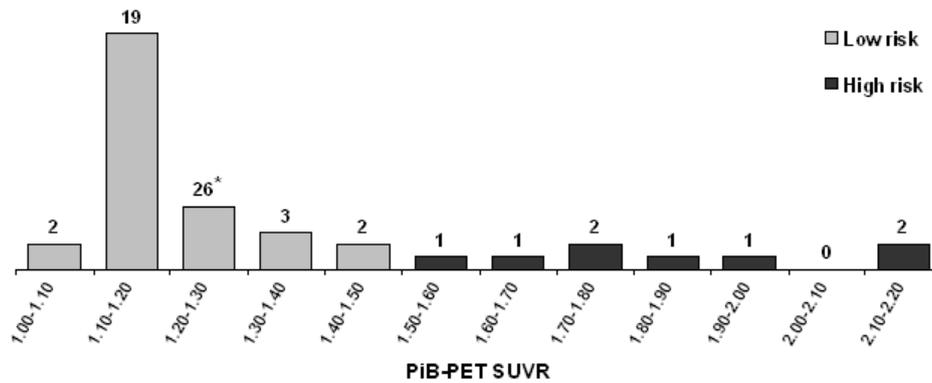


Figure 7.7: Summary of the AIBL PiB-PET SUVR measures for participants scanned using the 1.5 T MR scanner. Ranges containing values associated with subjects who progressed to either MCI or AD within 18 months are indicated by asterisks (*).

PiB-PET SUVR measures are summarised in Figure 7.8 for the 80 participants scanned using the 3 T MR scanner. Following sub-division of these participants, 34% were classified as having a high risk for the development of AD. The mean age of the high-risk sub-set (77.7 ± 5.8) was significantly higher than that of the low-risk sub-set (72.6 ± 7.1) (t-test, $p < 0.01$). For this reason, a linear regression model was applied to the MR volumes from both ADNI and AIBL to account for the effects of age. Three of the four subjects who had progressed to either MCI or AD within 18 months were scanned using the 3 T MR scanner, and all three were assigned to the high-risk group based on their PiB-PET SUVR measures.

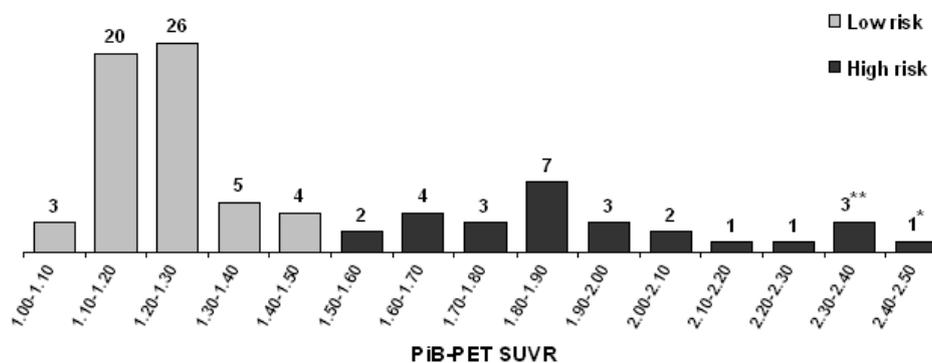


Figure 7.8: Summary of the AIBL PiB-PET SUVR measures for participants scanned using the 3 T MR scanner. Ranges containing values associated with subjects who progressed to either MCI or AD within 18 months are indicated by asterisks (*).

7.5 Volumetric differences between risk groups

Comparisons by t-test were performed between the ICV-normalised, age-corrected MR volumes in the high- and low-risk sub-sets of each group (ADNI, 1.5 T AIBL, 3 T AIBL). Correction for multiple comparisons was performed using the P plot graphical method described in Turkheimer et al. (2001), which is based on estimating the number of “true” null hypotheses in a set of statistics. This estimate can be used to sharpen existing multiple comparison correction approaches, such as the procedure described in Hochberg (1988).

The P plot method takes account of the fact that regional MR volumes are not entirely independent, and that differences are therefore expected in more than one brain region. It is a less conservative approach than the Bonferroni-like corrections described in Section 2.4.3, and has greater power for detecting true differences.

For a set of N hypotheses, the associated p -values are ranked in descending order, and a P plot constructed as illustrated in Figure 7.9. Points corresponding to true null hypotheses (large p) are expected to lie approximately along a straight line passing through the origin, while points corresponding to false hypotheses should deviate to the right. The slope, β , of the straight line fitted to the points with large p -values is used to estimate the number of “true” null hypotheses, $N_0 = (1/\beta) - 1$.

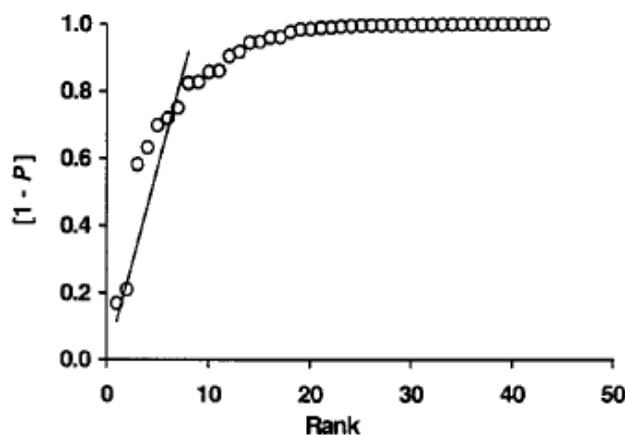


Figure 7.9: Illustration of a P plot based on p -values computed from t-tests on 43 regions (Turkheimer et al., 2001). The estimated slope of the straight line, β was used to estimate the number of “true” null hypotheses, $N_0 = 8$.

7.5.1 ADNI participants

Regional t-values between the high- and low-risk sub-sets of the cognitively normal ADNI participants are illustrated in Figure 7.10. Five of 83 regions across the brain were significantly smaller in the high-risk sub-set ($p < 0.05$, uncorrected). These regions are listed in Table 7.3, and outlined in blue on Figure 7.10.

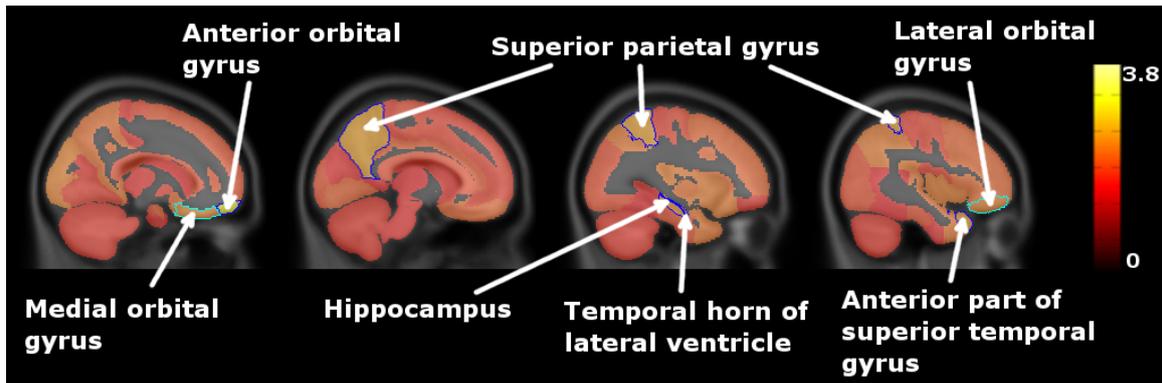


Figure 7.10: Regional t-values between age-corrected MR volumes in the high- and low-risk sub-sets of the cognitively normal ADNI participants are shown superimposed onto a maximum probability brain atlas which has been masked in the same way as the anatomical segmentations. The t-value map is overlaid onto a MNI-space average MR image for reference. Sagittal slices are shown viewed from the right side of the brain, moving towards the left. Regions showing a significant difference in age-corrected volume ($p < 0.05$, uncorrected) are outlined in blue. Since the significance threshold of $p < 0.05$ is somewhat arbitrary, regions with $0.05 < p < 0.1$ are additionally outlined in cyan. All outlined regions were smaller in the high-risk sub-set.

Region	p -value
Anterior orbital gyrus (right)	< 0.0005
Anterior part of superior temporal gyrus (left)	0.025
Superior parietal gyrus (left)	0.026
Temporal horn of lateral ventricle (left)	0.026
Hippocampus (left)	0.049
Medial orbital gyrus (right)	0.076
Lateral orbital gyrus (left)	0.095

Table 7.3: Regions showing a significant ($p < 0.05$, uncorrected) difference in age-corrected MR volume between the high- and low-risk sub-sets of the cognitively normal ADNI participants, along with the corresponding p -values. Since the significance threshold of $p < 0.05$ is somewhat arbitrary, regions are additionally listed down to a significance level of $p < 0.1$. All regions listed were smaller in the high-risk sub-set.

One region (right anterior orbital gyrus) remained significantly smaller in the high-risk sub-set after correction for multiple comparisons using the P plot graphical method ($p < 0.05$) described

in Section 7.5. The temporal horn of the left lateral ventricle was among the significantly smaller regions. This structure is expected to be enlarged in high-risk individuals. However, as shown in Figure 7.10, it is a very small region. The result could thus be attributed to subject motion during scanning or inaccuracies in the automatic segmentation procedure.

7.5.2 AIBL 1.5 T participants

Regional t-values between the high- and low-risk sub-sets of the cognitively normal AIBL participants scanned using the 1.5 T MR scanner are illustrated in Figure 7.11. Five of 83 regions across the brain were significantly smaller in the high-risk sub-set ($p < 0.05$, uncorrected). These regions are listed in Table 7.4, and outlined in blue on Figure 7.11. One region (left occipitotemporal gyrus) remained significantly smaller in the high-risk sub-set after correction for multiple comparisons using the P plot graphical method ($p < 0.05$).

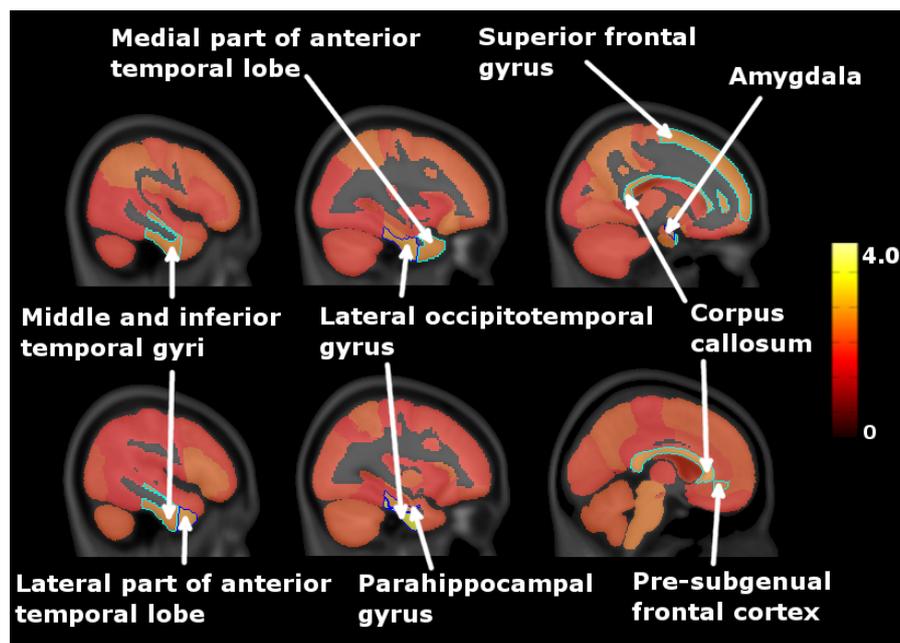


Figure 7.11: Regional t-values between age-corrected MR volumes in the high- and low-risk sub-sets of the cognitively normal AIBL participants scanned using the 1.5 T MR scanner are shown superimposed onto a maximum probability brain atlas which has been masked in the same way as the anatomical segmentations. The t-value map is overlaid onto a MNI-space average MR image for reference. The top row shows sagittal slices of the brain viewed from the right, and the bottom row shows slices from the left. Regions showing a significant difference in age-corrected volume ($p < 0.05$, uncorrected) are outlined in blue. Again, regions with $0.05 < p < 0.1$ are additionally outlined in cyan. All outlined regions were smaller in the high-risk sub-set, apart from the corpus callosum.

Region	p -value
Lateral occipitotemporal gyrus (left)	< 0.0005
Amygdala (right)	0.005
Lateral occipitotemporal gyrus (right)	0.021
Lateral part of anterior temporal lobe (left)	0.023
Parahippocampal gyrus (left)	0.037
Corpus callosum (spans midline)	0.050
Medial part of anterior temporal lobe (right)	0.056
Superior frontal gyrus (right)	0.058
Middle and inferior temporal gyri (right)	0.062
Middle and inferior temporal gyri (left)	0.066
Pre-subgenual frontal cortex (left)	0.089

Table 7.4: Regions showing a significant ($p < 0.05$, uncorrected) difference in age-corrected MR volume between the high- and low-risk sub-sets of the cognitively normal AIBL participants scanned using the 1.5 T MR scanner, along with the corresponding p -values. Again, regions are additionally listed down to a significance level of $p < 0.1$. All regions listed were smaller in the high-risk sub-set, apart from the corpus callosum.

7.5.3 AIBL 3 T participants

Regional t -values between the high- and low-risk sub-sets of the cognitively normal AIBL participants scanned using the 3 T MR scanner are illustrated in Figure 7.12. Four of 83 regions across the brain were significantly smaller in the high-risk sub-set ($p < 0.05$, uncorrected). These regions are listed in Table 7.5, and outlined in blue on Figure 7.12.

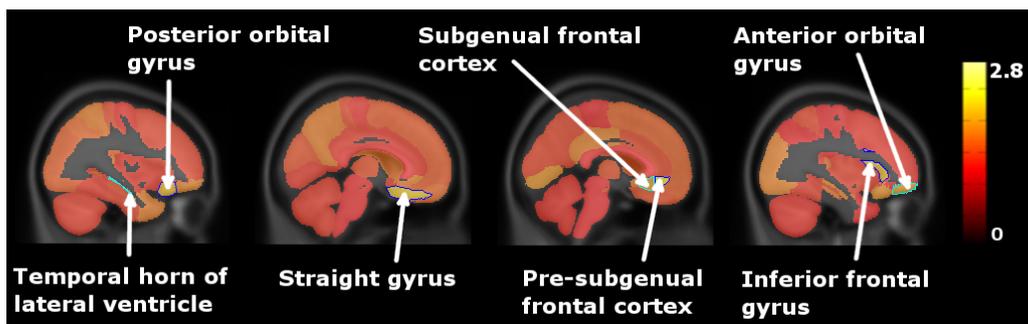


Figure 7.12: Regional t -values between age-corrected MR volumes in the high- and low-risk sub-sets of the cognitively normal AIBL participants scanned using the 3 T MR scanner are shown superimposed onto a maximum probability brain atlas which has been masked in the same way as the anatomical segmentations. The t -value map is overlaid onto a MNI-space average MR image for reference. Sagittal slices are shown viewed from the right side of the brain, moving towards the left. Regions showing a significant difference in age-corrected volume ($p < 0.05$, uncorrected) are outlined in blue. Again, regions with $0.05 < p < 0.1$ are additionally outlined in cyan. All outlined regions were smaller in the high-risk sub-set, apart from the temporal horn of the right lateral ventricle.

No regions remained significantly smaller in the high-risk sub-set after correction for multiple comparisons using the P plot graphical method ($p < 0.05$).

Region	p -value
Pre-subgenual frontal cortex (left)	0.004
Straight gyrus (right)	0.020
Inferior frontal gyrus (left)	0.025
Posterior orbital gyrus (right)	0.036
Subgenual frontal cortex (left)	0.059
Temporal horn of lateral ventricle (right)	0.060
Anterior orbital gyrus (left)	0.062

Table 7.5: Regions showing a significant ($p < 0.05$, uncorrected) difference in age-corrected MR volume between the high- and low-risk sub-sets of the cognitively normal AIBL participants scanned using the 3 T MR scanner, along with the corresponding p -values. Again, regions are additionally listed down to a significance level of $p < 0.1$. All regions listed were smaller in the high-risk sub-set, apart from the temporal horn of the right lateral ventricle.

7.5.4 Discussion

Multi-region analysis of MR images acquired at a single timepoint has been used to identify early signs of neurodegeneration in cognitively normal elderly individuals with evidence of cortical β -amyloid deposition. Reduced volumes in temporo-parietal and orbito-frontal regions in high-risk individuals were observed in two independent cohorts from the ADNI and AIBL studies. These differences could be indicative of very early changes associated with the development of AD. The similarity of regional differences observed in the two independent groups suggests that volumetric MRI can reveal structural brain changes that precede the onset of clinical symptoms. Variations in significant regions between subject groups can be attributed to the different demographic characteristics of the participants concerned. For example, the AIBL HC groups include individuals with subjective, but not objective, memory impairments.

Participants were assigned to risk groups for the development of AD based on evidence of their cortical β -amyloid deposition. This was assessed based on cutoff values on CSF measures of $A\beta$ in ADNI (high risk ≤ 192 pg/mL), and neocortical-to-cerebellar ratios of amyloid deposition on PiB-PET imaging in AIBL (high risk > 1.5). Similar proportions of participants were assigned to the high-risk sub-set of each group (ADNI, 39%; 1.5 T AIBL, 21%; 3 T AIBL, 34%). It

has been previously suggested that the cutoff values on CSF $A\beta$ and PiB-PET SUVR are analogous, and that measures of CSF $A\beta$ may be transformed into calculated PiB units of cortical β -amyloid (Weigand et al., 2011).

As discussed in Section 1.5.2, many studies have described the structural brain changes associated with MCI and AD. Far fewer studies have investigated the potential for detecting structural changes in the pre-symptomatic stages of the disease. It is important to consider that evidence of cortical β -amyloid deposition in cognitively normal elderly individuals does not necessarily indicate an increased risk of developing AD. The volumetric differences observed may therefore reflect the amyloid status of participants, rather than directly their risk for disease development.

Tondelli et al. (2012) used voxel-based morphometry and shape analyses of MR imaging data to show reduced volumes in temporo-parietal and orbito-frontal regions of the brain in cognitively normal subjects who were subsequently diagnosed with MCI or AD, in comparison with subjects who remained cognitively normal over ten years. These results were based on MR images, taken at least four years before the onset of any cognitive symptoms, from 8 pre-clinical AD patients, 32 pre-clinical MCI patients, and 40 HC. The results presented in this chapter are consistent with those of Tondelli et al. (2012), supporting the suggestion that these structural changes may be predictive of the future development of AD. Further clinical follow-up data will become available for both the ADNI and AIBL participants over the next few years, and one interesting area for future research would involve performing an analysis similar to that of Tondelli et al. (2012) based on the large ADNI and AIBL cohorts.

Schott et al. (2010) compared whole-brain, hippocampal and ventricular volumes and atrophy rates between cognitively normal ADNI participants, stratified into high- and low-risk based on CSF measures of $A\beta$. They reported significantly increased whole-brain loss, ventricular expansion, and hippocampal atrophy in high-risk individuals, but no significant volumetric differences. The work presented in this chapter has shown that a multi-region analysis approach can be used to detect significant volumetric differences in certain regions of the brain.

7.6 Conclusion

This chapter has presented findings of early signs of neurodegeneration in cognitively normal individuals at high risk of developing AD. Multi-region analysis of MR images acquired at a single timepoint was used to identify early signs of neurodegeneration in cognitively normal elderly individuals with evidence of cortical β -amyloid deposition. Reduced volumes in temporo-parietal and orbito-frontal regions in high-risk individuals were observed in both cohorts. The similarity of regional differences observed in two independent groups suggests that volumetric MRI can reveal structural brain changes that precede the onset of clinical symptoms. It may therefore be useful in identifying early signs of neurodegeneration in healthy elderly individuals, potentially providing a useful early screening tool, or outcome measure for clinical trials.

Future work will aim to determine whether the observed group differences can be translated into markers for the prediction of future cognitive decline in individual patients. In addition, it would be interesting to perform a similar study using multi-region FDG-PET signal intensities rather than MR volumes, since it is hypothesised that changes in metabolism can be detected on FDG-PET before corresponding structural changes are visible on MRI ([Aisen et al., 2010](#)).

Chapter 8

Overall conclusion

8.1 Contributions

The research presented in this thesis contributes to the growing body of literature surrounding the image-based classification of MCI and AD. Imaging biomarkers for AD are important for improved diagnosis and monitoring, as well as drug discovery. Automated image-based classification of individual patients could provide valuable diagnostic support for clinicians, when considered alongside cognitive assessment scores and traditional visual image analysis. This research has investigated machine learning methods aimed at the early identification of AD, and prediction of progression in patients with MCI. A comparison of the key classification results from Chapters 4, 5 and 6 is presented in Table 8.1.

	Base. PET		Long. PET		PET, MRI		PET, MRI, CSF, Gen.	
	N/N	acc	N/N	acc	N/N	acc	N/N	acc
AD/HC	71/69	81.6%	50/54	88.4%	71/69	90.0%	37/35	89.0%
MCI/HC	147/69	70.2%	-	-	147/69	75.5%	75/35	74.6%
pMCI/HC	62/69	71.8%	53/54	81.3%	-	-	-	-
AD/sMCI	71/85	74.2%	50/64	83.5%	-	-	-	-
pMCI/sMCI	62/85	56.4%	53/64	63.1%	-	-	34/41	58.0%

Table 8.1: Comparison of key classification results from Chapters 4, 5 and 6. The number of subjects in each diagnostic group (N/N) and classification accuracy (acc) are shown for baseline (base.) FDG-PET, longitudinal (long.) FDG-PET, FDG-PET combined with MRI, and the combination of FDG-PET, MRI, CSF and genetic (gen.) data.

Investigations of the potential utility of multi-region FDG-PET features for image-based classification of AD and MCI were described in Chapters 4 and 5. In particular, attempts were made to distinguish between MCI patients who subsequently progressed to AD and those who remained stable. Chapter 4 demonstrated that regional information extracted from FDG-PET images acquired at a single timepoint can be used to achieve classification results in line with those obtained using data from MRI, or biomarkers obtained invasively from the CSF. Chapter 5 then demonstrated the additional benefit of incorporating longitudinal FDG-PET information for classification. By combining cross-sectional and longitudinal multi-region FDG-PET features, classification results in line with the current state-of-the-art were achieved. The findings described in these chapters support the use of FDG-PET for the early diagnosis of AD and for monitoring its progression.

Chapter 6 presented a multi-modality classification framework in which manifolds are constructed based on pairwise similarity measures derived from random forest classifiers. Similarities from multiple modalities were combined to generate an embedding that simultaneously encoded information about all the available features. Multi-modality classification was then performed using coordinates from this joint embedding. Random forests provide consistent pairwise similarity measures for multiple modalities, thus facilitating the combination of different types of feature data. Classification results based on the combination of regional MRI volumes, voxel-based FDG-PET signal intensities, CSF biomarker measures, and ApoE allele status are comparable with those obtained in other studies using multi-kernel learning. Since random forest classifiers extend naturally to multi-class problems, the framework described could be used for other applications in the future, such as the differential diagnosis of AD.

Novel findings of early signs of neurodegeneration in cognitively normal individuals at high risk of developing AD were presented in Chapter 7. Multi-region analysis of MR images acquired at a single timepoint was used to show volumetric differences in cognitively normal individuals differing in amyloid-based risk status for the development of AD. Reduced volumes in temporoparietal and orbito-frontal regions in high-risk individuals from two independent cohorts could be indicative of very early changes associated with AD. These findings suggest that volumetric MRI can reveal structural brain changes that precede the onset of clinical symptoms. It may

therefore be useful in identifying early signs of neurodegeneration in healthy elderly individuals, potentially providing a useful early screening tool, or outcome measure for clinical trials.

8.2 Future work

The work presented in Chapters 4, 5 and 6 has been focused on the image-based classification of AD and MCI. As explained in Chapter 4, classification results for distinguishing between AD patients and HC may be converging on a glass ceiling since the diagnostic consensus criteria themselves have an accuracy of around 90% (Ranginwala et al., 2008). One interesting area for further research could be to compare the subjects that are mis-classified using different machine learning methods based on the various available modalities. It would also be beneficial to further investigate the case of discriminating pMCI from sMCI patients, since this is one of the most clinically interesting. As explained in Chapter 4, the three years of clinical follow-up that will eventually be available for the MCI patients are likely to be insufficient to allow the identification of all those who will develop AD in the future. However, the ADNI study has been extended in the form of the ADNI-GO and ADNI-2 studies, which are summarised in Table 8.2. These will provide additional clinical follow-up for the original ADNI participants. Further information is available *via* the ADNI website (<http://adni.loni.ucla.edu/about/about-the-study/>).

Study	Timeline	HC	eMCI	lMCI	AD	MRI	fMRI	FDG	AV45	Bio.
ADNI-1	2004–2010	200	–	400	200	✓	✗	✓	✗	✓
ADNI-GO	2009–2011	–	200	–	–	✓	✓	✓	✓	✓
ADNI-2	2011–2016	150	100	150	150	✓	✓	✓	✓	✓

Table 8.2: Summary of the ADNI studies, including the approximate timelines, number of participants in each diagnostic group, and availability of each modality. ADNI-GO and ADNI-2 continue to follow the ADNI-1 participants, as well as recruiting additional subjects as shown. MCI patients are divided into early (eMCI) and late (lMCI) groups. Available modalities include structural MRI, resting state functional MRI (fMRI), FDG-PET, AV45-PET, and biospecimens (bio.). [¹⁸F]-AV45 is an amyloid imaging PET tracer developed by Avid Radiopharmaceuticals. Biospecimens include DNA, CSF and blood samples.

As described in Section 1.5, changes in multiple biomarkers may provide complementary information for the diagnosis and prognosis of AD. This was demonstrated by the work on multi-modality classification presented in Chapter 6. It may be interesting to further investigate the

relationships among the various modalities, since this potentially enables decisions to be made on how to acquire the maximum amount of diagnostically relevant information for a patient using a minimum number of assessments. Additional information could also be incorporated into the described framework, including clinical and neuropsychological assessment scores, and longitudinal data. It would be particularly interesting to see if incorporating the additional clinical follow-up information, and longitudinal imaging and biological data enables improved discrimination between pMCI and sMCI patients. The random forest classifiers applied to derive the similarity measures extend naturally to multi-class problems. It could therefore be interesting to investigate the use of multi-modality classification for the differential diagnosis of AD. While it is possible that the ADNI dataset contains some patients with other dementias, such as frontotemporal dementia or dementia with Lewy bodies, these patients are not clinically labelled as such. To perform a thorough study on differential diagnosis, a large and varied cohort of dementia patients with autopsy-confirmed clinical diagnoses would be required, such as that described in [Silverman et al. \(2001\)](#).

The work presented in Chapter 7 is not based on classification, but instead presents novel findings of early signs of neurodegeneration in cognitively normal individuals at high risk of developing AD. There is much further research to be done based on these early results. For example, future work will aim to determine whether the observed group differences can be translated into markers for the prediction of future cognitive decline in individual patients. The ability to identify high-risk individuals based on structural MRI would be beneficial, because it is less invasive than either lumbar puncture or PET imaging. The continuation of the ADNI study will mean that further clinical follow-up information will be available, and it may then become possible to replicate the study of [Tondelli et al. \(2012\)](#). In addition, further FDG-PET data will be acquired, and it would be interesting to perform a similar study using multi-region FDG-PET signal intensities rather than MR volumes, since it is hypothesised that changes in metabolism can be detected on FDG-PET before corresponding structural changes are visible on MRI ([Aisen et al., 2010](#)).

Chapter 9

Publications

- K. R. Gray, R. Wolz, R. A. Heckemann, D. Rueckert, A. Hammers. Structural differences in cognitively normal elderly individuals with abnormal amyloid biomarkers: detection using volumetric MRI in ADNI and AIBL. (Abstract) Alzheimer's Association International Conference (AAIC'12), 2012 (Accepted).
- K. R. Gray, R. Wolz, R. A. Heckemann, P. Aljabar, A. Hammers, D. Rueckert, for the Alzheimer's Disease Neuroimaging Initiative. Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease. *Neuroimage*, 60(1):221–229, 2012.
<http://dx.doi.org/10.1016/j.neuroimage.2011.12.071>
- K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, D. Rueckert. Random forest-based manifold learning for classification of imaging data in dementia. (In Proceedings) Machine Learning in Medical Imaging Workshop at 14th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'11). *Lecture Notes in Computer Science*, 7009:159–166, 2011.
http://dx.doi.org/10.1007/978-3-642-24319-6_20
- K. R. Gray, R. A. Heckemann, A. Hammers, D. Rueckert. Multi-region analysis of longitudinal FDG-PET enables accurate AD classification. (Abstract) Alzheimer's Association International Conference (AAIC'11), 2011.

-
- R. A. Heckemann, S. Keihaninejad, K. R. Gray, D. Rueckert, J. V. Hajnal, A. Hammers. Visual presentation of regional volumetry data for diagnostic decision support in AD MR imaging. (Abstract) Alzheimer's Association International Conference (AAIC'11), 2011.
 - K. R. Gray, R. Wolz, S. Keihaninejad, R. A. Heckemann, P. Aljabar, A. Hammers, D. Rueckert. Regional analysis of FDG-PET for the classification of Alzheimer's disease. (In Proceedings) 15th Conference on Medical Image Understanding and Analysis (MIUA'11), 2011.
<http://www.biomedical-image-analysis.co.uk/images/stories/gray-posters2-77.pdf>
 - E. Janousova, M. Vounou, R. Wolz, K. R. Gray, D. Rueckert, G. Montana. Fast brain-wide search of highly discriminative regions in medical images: an application to Alzheimer's disease. (In Proceedings) 15th Conference on Medical Image Understanding and Analysis (MIUA'11), 2011.
<http://www.biomedical-image-analysis.co.uk/images/stories/janousova-mandsini-71.pdf>
 - R. A. Heckemann, S. Keihaninejad, P. Aljabar, K. R. Gray, C. Nielsen, D. Rueckert, J. V. Hajnal, A. Hammers, and the Alzheimer's Disease Neuroimaging Initiative. Automatic morphometry in Alzheimer's disease and mild cognitive impairment. *Neuroimage*, 56(4):2024-2037, 2011.
<http://dx.doi.org/10.1016/j.neuroimage.2011.03.014>
 - K. R. Gray, R. Wolz, S. Keihaninejad, R. A. Heckemann, P. Aljabar, A. Hammers, D. Rueckert. Regional analysis of FDG-PET for use in the classification of Alzheimer's disease. (In Proceedings) IEEE 8th International Symposium on Biomedical Imaging (ISBI'11), 1082–1085, 2011.
<http://dx.doi.org/10.1109/ISBI.2011.5872589>
 - R. A. Heckemann, S. Keihaninejad, P. Aljabar, K. R. Gray, C. Nielsen, D. Rueckert, J. V. Hajnal, A. Hammers. A repository of MR morphometry data in Alzheimer's disease

and mild cognitive impairment. (In Proceedings) IEEE 8th International Symposium on Biomedical Imaging (ISBI'11), 875–878, 2011.

<http://dx.doi.org/10.1109/ISBI.2011.5872543>

- K. R. Gray, R. Wolz, R. A. Heckemann, A. Hammers, D. Rueckert. Classification of ADNI subjects based on longitudinal analysis of the hippocampal FDG-PET signal. (Abstract) Alzheimer's Association International Conference on Alzheimer's Disease (ICAD'10). *Alzheimer's and Dementia*, 6(4, Supplement):S288–S289, 2010.
<http://dx.doi.org/10.1016/j.jalz.2010.05.949>
- K. R. Gray, K. B. Contractor, L. M. Kenny, A. Al-Nahhas, S. Shousha, J. Stebbing, H. S. Wasan, R. C. Coombes, E. O. Aboagye, F. E. Turkheimer, L. Rosso. Kinetic filtering of [18F]Fluorothymidine in positron emission tomography studies. *Physics in Medicine and Biology*, 55(3):695–709, 2010.
<http://dx.doi.org/10.1088/0031-9155/55/3/010>
- L. Rosso, K. R. Gray, K. B. Contractor, L. M. Kenny, A. Al-Nahhas, S. Shousha, J. Stebbing, H. S. Wasan, R. C. Coombes, E. O. Aboagye, F. E. Turkheimer. New method for segmentation of tumors from [18F]FLT-PET images using predefined kinetic classes. (Abstract) 22nd Annual Congress of the European Association of Nuclear Medicine (EANM'09), 2009.

Bibliography

ADNI Data Sharing and Publications Committee. *Alzheimer's Disease Neuroimaging Initiative (ADNI): data use agreement*, 2012.

http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Data_Use_Agreement.pdf.

ADNI MRI Core. *ADNI MRI technical procedures manual*, 2005.

http://www.adni-info.org/scientists/Pdfs/MRI_Tech_Manual.pdf.

ADNI PET Core. *ADNI PET technical procedures manual*, 2005.

http://www.adni-info.org/scientists/Pdfs/PIB_PET_Tech_Manual.pdf.

P. S. Aisen, R. C. Petersen, M. C. Donohue, A. Gamst, R. Raman, R. G. Thomas, S. Walter, J. Q. Trojanowski, L. M. Shaw, L. A. Beckett, C. R. Jack Jr., W. Jagust, A. W. Toga, A. J. Saykin, J. C. Morris, R. C. Green, M. W. Weiner, and the Alzheimer's Disease Neuroimaging Initiative. Clinical core of the Alzheimer's disease neuroimaging initiative: progress and plans. *Alzheimer's and Dementia*, 6(3):239–246, 2010.

<http://dx.doi.org/10.1016/j.jalz.2010.03.006>

M. Aizerman, E. Braverman, and L. Rozonoer. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25:821–837, 1964.

M. S. Albert, S. T. DeKosky, D. Dickson, B. Dubois, H. H. Feldman, N. C. Fox, A. Gamst, D. M. Holtzman, W. J. Jagust, R. C. Petersen, P. J. Snyder, M. C. Carrillo, B. Thies, and C. H. Phelps. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines

for Alzheimer's disease. *Alzheimer's and Dementia*, 7(3):270–279, 2011.

<http://dx.doi.org/10.1016/j.jalz.2011.03.008>

G. E. Alexander, K. Chen, P. Pietrini, S. I. Rapoport, and E. M. Reiman. Longitudinal PET evaluation of cerebral metabolic decline in dementia: a potential outcome measure in Alzheimer's disease treatment studies. *American Journal of Psychiatry*, 159(5):738–745, 2002.

<http://dx.doi.org/10.1176/appi.ajp.159.5.738>

P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage*, 46(3):726–738, 2009.

<http://dx.doi.org/10.1016/j.neuroimage.2009.02.018>

P. Aljabar, D. Rueckert, and W. R. Crum. Automated morphological analysis of magnetic resonance brain imaging using spectral analysis. *Neuroimage*, 43(2):225–235, 2008.

<http://dx.doi.org/10.1016/j.neuroimage.2008.07.055>

P. Aljabar, R. Wolz, and D. Rueckert. Manifold learning for medical image registration, segmentation, and classification. In K. Suzuki (editor), *Machine learning in computer-aided diagnosis: medical imaging intelligence and analysis*. IGI Global, 2012.

<http://dx.doi.org/10.4018/978-1-4666-0059-1.ch017>

P. Aljabar, R. Wolz, L. Srinivasan, S. Counsell, J. P. Boardman, M. Murgasova, V. Doria, M. Rutherford, A. D. Edwards, J. V. Hajnal, and D. Rueckert. Combining morphological information in a manifold learning framework: application to neonatal MRI. (In Proceedings) 13th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'10), *Lecture Notes in Computer Science*, 6363:1–8, 2010.

http://dx.doi.org/10.1007/978-3-642-15711-0_1

Y. Amit and D. Geman. Shape quantization and recognition with randomized trees. *Neural Computation*, 9(7):1545–1588, 1997.

<http://dx.doi.org/10.1162/neco.1997.9.7.1545>

D. Anchisi, B. Borroni, M. Franceschi, N. Kerrouche, E. Kalbe, B. Beuthien-Beumann, S. Cappa, O. Lenz, S. Ludecke, A. Marccone, R. Mielke, P. Ortelli, A. Padovani, O. Pelati, A. Pupi, E. Scarpini, S. Weisenbach, K. Herholz, E. Salmon, V. Holthoff, S. Sorbi, F. Fazio, and D. Perani. Heterogeneity of brain glucose metabolism in mild cognitive impairment and clinical progression to Alzheimer disease. *Archives of Neurology*, 62(11):1728–1733, 2005.

<http://dx.doi.org/10.1001/archneur.62.11.1728>

H. A. Archer, P. Edison, D. J. Brooks, J. Barnes, C. Frost, T. Yeatman, N. C. Fox, and M. N. Rossor. Amyloid load and cerebral atrophy in Alzheimer’s disease: an C-11-PiB positron emission tomography study. *Annals of Neurology*, 60(1):145–147, 2006.

<http://dx.doi.org/10.1002/ana.20889>

J. Ashburner and K. J. Friston. Unified segmentation. *Neuroimage*, 26(3):839–851, 2005.

<http://dx.doi.org/doi:10.1016/j.neuroimage.2005.02.018>

J. A. Aston, V. J. Cunningham, M. C. Asselin, A. Hammers, A. C. Evans, and R. N. Gunn. Positron emission tomography partial volume correction: estimation and algorithms. *Journal of Cerebral Blood Flow and Metabolism*, 22(8):1019–1034, 2002.

<http://dx.doi.org/10.1097/00004647-200208000-00014>

E. D. Becker. *High resolution NMR: theory and chemical applications*. Academic Press, New York, 3rd edition, 2000.

M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

<http://dx.doi.org/10.1162/089976603321780317>

L. Berg. Clinical Dementia Rating (CDR). *Psychopharmacology Bulletin*, 24(4):637–639, 1988.

C. E. Bonferroni. Il calcolo delle assicurazioni su gruppi di teste. (In Proceedings) Studi in Onore del Professore Salvatore Ortu Carboni, 13–60, 1935.

C. E. Bonferroni. Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8:3–62, 1936.

P. Borghammer, J. Aanerud, and A. Gjedde. Data-driven intensity normalization of PET group comparison studies is superior to global mean normalization. *Neuroimage*, 46(4):981–988, 2009.
<http://dx.doi.org/10.1016/j.neuroimage.2009.03.021>

B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. (In Proceedings) ACM 5th Annual Workshop on Computational Learning Theory (COLT'92), 144–152, 1992.
<http://dx.doi.org/10.1145/130385.130401>

H. Braak and E. Braak. Evolution of neuronal changes in the course of Alzheimer's disease. *Journal of Neural Transmission Supplement*, 53:127–140, 1998.

L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.
<http://dx.doi.org/10.1007/BF00058655>

L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001.
<http://dx.doi.org/10.1023/A:1010933404324>

R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi. Forecasting the global burden of Alzheimer's disease. *Alzheimer's and Dementia*, 3(3):186–191, 2007.
<http://dx.doi.org/10.1016/j.jalz.2007.04.381>

Central Intelligence Agency. *The world factbook*, 2012.
<https://www.cia.gov/library/publications/the-world-factbook/index.html>.

C. Chang and C. Lin. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27:1–27:27, 2011.
<http://doi.acm.org/10.1145/1961189.1961199>.

K. Chen, J. B. S. Langbaum, A. S. Fleisher, N. Ayutyanont, C. Reschke, W. Lee, X. Liu, D. Bandy, G. E. Alexander, P. M. Thompson, N. L. Foster, D. J. Harvey, M. J. de Leon, R. A. Koeppe, W. J. Jagust, M. W. Weiner, E. M. Reiman, and the Alzheimer's Disease Neuroimaging Initiative. Twelve-month metabolic declines in probable Alzheimer's disease and amnesic mild cognitive impairment assessed using an empirically pre-defined statistical

region-of-interest: findings from the Alzheimer's Disease Neuroimaging Initiative. *Neuroimage*, 51(2):654–664, 2010.

<http://dx.doi.org/10.1016/j.neuroimage.2010.02.064>

C. M. Clark, S. Xie, J. Chittams, D. Ewbank, E. Peskind, D. Galasko, J. C. Morris, D. W. McKeel, M. Farlow, S. L. Weitlauf, J. Quinn, J. Kaye, D. Knopman, H. Arai, R. S. Doody, C. DeCarli, S. Leight, V. M. Y. Lee, and J. Q. Trojanowski. Cerebrospinal fluid tau and beta-amyloid: how well do these biomarkers reflect autopsy-confirmed dementia diagnoses? *Archives of Neurology*, 60(12):1696–1702, 2003.

<http://dx.doi.org/10.1001/archneur.60.12.1696>

M. J. Clarkson, S. Ourselin, C. Nielsen, K. K. Leung, J. Barnes, J. L. Whitwell, J. L. Gunter, D. L. G. Hill, M. W. Weiner, C. R. Jack Jr., N. C. Fox, and the Alzheimer's Disease Neuroimaging Initiative. Comparison of phantom and registration scaling corrections using the ADNI cohort. *Neuroimage*, 47(4):1506–1513, 2009.

<http://dx.doi.org/10.1016/j.neuroimage.2009.05.045>

A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multimodality image registration using information theory. (In Proceedings) 14th International Conference on Information Processing in Medical Imaging (IPMI'95), 263–274, 1995.

O. Colliot, G. Chételat, M. Chupin, B. Desgranges, B. Magnin, H. Benali, B. Dubois, L. Garnero, F. Eustache, and S. Lehéricy. Discrimination between Alzheimer's disease, mild cognitive impairment and normal aging by using automated segmentation of the hippocampus. *Radiology*, 248(1):194–201, 2008.

<http://dx.doi.org/10.1148/radiol.2481070876>

E. H. Corder, A. M. Saunders, W. J. Strittmatter, D. E. Schmechel, P. C. Gaskell, G. W. Small, A. D. Roses, J. L. Haines, and M. A. Pericak-Vance. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*, 261(5123):921–923, 1993.

<http://dx.doi.org/10.1126/science.8346443>

C. Cortes and V. N. Vapnik. Support-Vector Networks. *Machine Learning*, 20(3):273–297, 1995.

<http://dx.doi.org/10.1007/BF00994018>

T. F. Cox and M. A. A. Cox. *Multidimensional scaling*. Chapman and Hall, 2001.

A. Criminisi, J. Shotton, and E. Konukoglu. Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends in Computer Graphics and Vision*, 7(2–3):81–227, 2012.

<http://dx.doi.org/10.1561/06000000035>

W. R. Crum, R. I. Scahill, and N. C. Fox. Automated hippocampal segmentation by regional fluid registration of serial MRI: validation and application in Alzheimer’s disease. *Neuroimage*, 13(5):847–855, 2001.

<http://dx.doi.org/10.1006/nimg.2001.0744>

R. Cuingnet, E. Gérardin, J. Tessieras, G. Auzias, S. Lehéricy, M. Habert, M. Chupin, H. Benali, O. Colliot, and the Alzheimer’s Disease Neuroimaging Initiative. Automatic classification of patients with Alzheimer’s disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage*, 56(2):766–781, 2011.

<http://dx.doi.org/10.1016/j.neuroimage.2010.06.013>

C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S. M. Resnick. Detection of prodromal Alzheimer’s disease via pattern classification of magnetic resonance imaging. *Neurobiology of Aging*, 29(4):514–523, 2008a.

<http://dx.doi.org/10.1016/j.neurobiolaging.2006.11.010>

C. Davatzikos, S. M. Resnick, X. Wu, P. Parmpi, and C. M. Clark. Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. *Neuroimage*, 41(4):1220–1227, 2008b.

<http://dx.doi.org/10.1016/j.neuroimage.2008.03.050>

D. Dawbarn and S. J. Allen (editors). *Neurobiology of Alzheimer’s disease*. Oxford University Press, New York, 3rd edition, 2007.

M. J. de Leon (editor). *An atlas of Alzheimer's disease*. Parthenon Publishing Group, London, 1999.

M. J. de Leon, A. Convit, O. T. Wolf, C. Y. Tarshish, S. DeSanti, H. Rusinek, W. Tsui, E. Kandil, A. J. Scherer, A. Roche, A. Imossi, E. Thorn, M. Bobinski, C. Caraos, P. Lesbre, D. Schlyer, J. Poirier, B. Reisberg, and J. Fowler. Prediction of cognitive decline in normal elderly subjects with 2-[F-18]fluoro-2-deoxy-D-glucose/positron-emission tomography (FDG/PET). *Proceedings of the National Academy of Sciences of the United States of America*, 98(19):10966–10971, 2001.

<http://dx.doi.org/10.1073/pnas.191044198>

M. J. de Leon, S. H. Ferris, A. E. George, B. Reisberg, D. R. Christman, I. I. Kricheff, and A. P. Wolf. Computed tomography and positron emission transaxial tomography evaluations of normal aging and Alzheimer's disease. *Journal of Cerebral Blood Flow and Metabolism*, 3(3):391–394, 1983.

<http://dx.doi.org/10.1038/jcbfm.1983.57>

B. C. Dickerson, I. Goncharova, M. P. Sullivan, C. Forchetti, R. S. Wilson, D. A. Bennett, L. A. Beckett, and L. deToledo–Morrell. MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease. *Neurobiology of Aging*, 22(5):747–754, 2001.

[http://dx.doi.org/10.1016/S0197-4580\(01\)00271-8](http://dx.doi.org/10.1016/S0197-4580(01)00271-8)

A. Drzezga, T. Grimmer, M. Riemenschneider, N. Lautenschlager, H. Siebner, P. Alexopoulos, S. Minoshima, M. Schwaiger, and A. Kurz. Prediction of individual clinical outcome in MCI by means of genetic assessment and F-18-FDG PET. *Journal of Nuclear Medicine*, 46(10):1625–1632, 2005.

<http://jnm.snmjournals.org/content/46/10/1625.abstract>.

A. Drzezga, M. Souvatzogloul, M. Eiber, A. J. Beer, S. Fürst, A. Martinez-Möller, S. G. Nekolla, S. Ziegler, C. Ganter, E. J. Rummeny, , and M. Schwaiger. First clinical experience with integrated whole-body PET/MR: comparison to PET/CT in patients with oncologic diagnoses. *Journal of Nuclear Medicine*, 53(6):845–855, 2012.

<http://dx.doi.org/10.2967/jnumed.111.098608>

S. Duchesne, A. Caroli, C. Geroldi, C. Barillot, G. B. Frisoni, and D. L. Collins. MRI-Based Automated Computer Classification of Probable AD Versus Normal Controls. *IEEE Transactions on Medical Imaging*, 27(4):509–520, 2008.

<http://dx.doi.org/10.1109/TMI.2007.908685>

K. A. Ellis, A. I. Bush, D. Darby, D. De Fazio, J. Foster, P. Hudson, N. T. Lautenschlager, N. Lenzo, R. N. Martins, P. Maruff, C. Masters, A. Milner, K. Pike, C. Rowe, G. Savage, C. Szoeki, K. Taddei, V. Villemagne, M. Woodward, D. Ames, and the AIBL research group. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer’s disease. *International Psychogeriatrics*, 21(4):672–687, 2009.

<http://dx.doi.org/10.1017/S1041610209009405>

K. A. Ellis, C. C. Rowe, V. L. Villemagne, R. N. Martins, C. L. Masters, O. Salvado, C. Szoeki, D. Ames, and the AIBL research group. Addressing population aging and Alzheimer’s disease through the Australian Imaging Biomarkers and Lifestyle study: collaboration with the Alzheimer’s Disease Neuroimaging Initiative. *Alzheimer’s and Dementia*, 6(3):291–296, 2010.

<http://dx.doi.org/10.1016/j.jalz.2010.03.009>

A. M. Fagan, M. A. Mintun, R. H. Mach, S. Y. Lee, C. S. Dence, A. R. Shah, G. N. LaRossa, M. L. Spinner, W. E. Klunk, C. A. Mathis, S. T. DeKosky, J. C. Morris, and D. M. Holtzman. Inverse relation between in vivo amyloid imaging load and cerebrospinal fluid Abeta42 in humans. *Annals of Neurology*, 59(3):512–519, 2006.

<http://dx.doi.org/10.1002/ana.20730>

R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Human Genetics*, 7(2):179–188, 1936.

<http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x>

P. Flach, J. Hernandez-Orallo, and C. Ferri. A coherent interpretation of AUC as a measure of aggregated classification performance. (In Proceedings) ACM 28th Annual International Conference on Machine Learning (ICML’11), 657–664, 2011.

R. Fletcher. *Practical methods of optimization*. Wiley, 1987.

M. F. Folstein, S. E. Folstein, and P. R. McHugh. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3):189–198, 1975.

[http://dx.doi.org/10.1016/0022-3956\(75\)90026-6](http://dx.doi.org/10.1016/0022-3956(75)90026-6)

A. Forsberg, H. Engler, O. Almkvist, G. Blomquist, G. Hagman, A. Wall, A. Ringheim, B. Långström, and A. Nordberg. PET imaging of amyloid deposition in patients with mild cognitive impairment. *Neurobiology of Aging*, 29(10):1456–1465, 2008.

<http://dx.doi.org/10.1016/j.neurobiolaging.2007.03.029>

D. R. Forsey and R. H. Bartels. Hierarchical B-spline refinement. (In Proceedings) ACM 15th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH’88), 205–212, 1988.

<http://dx.doi.org/10.1145/54852.378512>

R. A. Frank, D. Galasko, H. Hampel, J. Hardy, M. J. de Leon, P. D. Mehta, J. Rogers, E. Siemers, and J. Q. Trojanowski. Biological markers for therapeutic trials in Alzheimer’s disease proceedings of the Biological Measures Working Group: NIA Initiative on Neuroimaging in Alzheimer’s Disease. *Neurobiology of Aging*, 24(4):521–536, 2003.

[http://dx.doi.org/10.1016/S0197-4580\(03\)00002-2](http://dx.doi.org/10.1016/S0197-4580(03)00002-2)

P. A. Freeborough, N. C. Fox, and R. I. Kitney. Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. *Computer Methods and Programs in Biomedicine*, 53(1):15–25, 1997.

[http://dx.doi.org/10.1016/S0169-2607\(97\)01803-8](http://dx.doi.org/10.1016/S0169-2607(97)01803-8)

Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

<http://dx.doi.org/10.1006/jcss.1997.1504>

K. J. Friston, C. D. Frith, P. F. Liddle, R. J. Dolan, A. A. Lammertsma, and R. S. J. Frackowiak. The relationship between global and local changes in PET scans. *Journal of Cerebral Blood*

Flow and Metabolism, 10:458–466, 1990.

<http://dx.doi.org/10.1038/jcbfm.1990.88>

K. J. Friston, C. D. Frith, P. F. Liddle, and R. S. J. Frackowiak. Comparing functional (PET) images: the assessment of significant change. *Journal of Cerebral Blood Flow and Metabolism*, 11:690–699, 1991.

<http://dx.doi.org/10.1038/jcbfm.1991.122>

S. Gerber, T. Tasdizen, S. Joshi, and R. Whitaker. On the manifold structure of the space of brain images. (In Proceedings) 12th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'09), *Lecture Notes in Computer Science*, 5761:305–312, 2009.

http://dx.doi.org/10.1007/978-3-642-04268-3_38

I. S. Gousias, D. Rueckert, R. A. Heckemann, L. E. Dyet, J. P. Boardman, A. D. Edwards, and A. Hammers. Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. *Neuroimage*, 40(2):672–684, 2008.

<http://dx.doi.org/10.1016/j.neuroimage.2007.11.034>

D. I. Graham and P. L. Lantos (editors). *Greenfield's neuropathology*. Arnold, London, 6th edition, 1997.

H. Gray. *Anatomy of the human body*. Lea and Febiger, Philadelphia, 20th edition, 1918.

<http://www.bartleby.com/107>.

C. Haense, K. Herholz, W. J. Jagust, and W. D. Heiss. Performance of FDG PET for detection of Alzheimer's disease in two independent multicentre samples (NEST-DD and ADNI). *Dementia and Geriatric Cognitive Disorders*, 28(3):259–266, 2009.

<http://dx.doi.org/10.1159/000241879>

J. V. Hajnal, D. J. Hawkes, and D. L. G. Hill (editors). *Medical image registration*. CRC Press, 2001.

A. Hammers, R. Allom, M. J. Koeppe, S. L. Free, R. Myers, L. Lemieux, T. N. Mitchell, D. J. Brooks, and J. S. Duncan. Three-dimensional maximum probability atlas of the human brain,

with particular reference to the temporal lobe. *Human Brain Mapping*, 19(4):224–247, 2003.

<http://dx.doi.org/10.1002/hbm.10123>

H. Hampel, K. Burger, S. J. Teipel, A. L. Bokde, H. Zetterberg, and K. Blennow. Core candidate neurochemical and imaging biomarkers of Alzheimer’s disease. *Alzheimer’s and Dementia*, 4(1):38–48, 2008.

<http://dx.doi.org/10.1016/j.jalz.2007.08.006>

O. Hansson, H. Zetterberg, P. Buchlave, E. Londos, K. Blennow, and L. Minthon. Association between CSF biomarkers and incipient Alzheimer’s disease in patients with mild cognitive impairment: a follow-up study. *Lancet Neurology*, 5(3):228–234, 2006.

[http://dx.doi.org/10.1016/S1474-4422\(06\)70355-6](http://dx.doi.org/10.1016/S1474-4422(06)70355-6)

T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer, New York, corrected 5th printing of 2nd edition, 2011.

R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *Neuroimage*, 33(1):115–126, 2006.

<http://dx.doi.org/10.1016/j.neuroimage.2006.05.061>

R. A. Heckemann, S. Keihaninejad, P. Aljabar, K. R. Gray, C. Nielsen, D. Rueckert, J. V. Hajnal, A. Hammers, and the Alzheimer’s Disease Neuroimaging Initiative. Automatic morphometry in Alzheimer’s disease and mild cognitive impairment. *Neuroimage*, 56(4):2024–2037, 2011.

<http://dx.doi.org/10.1016/j.neuroimage.2011.03.014>

R. A. Heckemann, S. Keihaninejad, P. Aljabar, D. Rueckert, J. V. Hajnal, A. Hammers, and the Alzheimer’s Disease Neuroimaging Initiative. Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. *Neuroimage*, 51(1):221–227, 2010.

<http://dx.doi.org/10.1016/j.neuroimage.2010.01.072>

K. Herholz, E. Salmon, D. Perani, J. C. Baron, V. Holthoff, L. Frolich, P. Schonknecht, K. Ito, R. Mielke, E. Kalbe, G. Zundorf, X. Delbeuck, O. Pelati, D. Anchisi, F. Fazio, N. Kerrouche, B. Desgranges, F. Eustache, B. Beuthien-Baumann, C. Menzel, J. Schroder, T. Kato, Y. Arahata, M. Henze, and W. D. Heiss. Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET. *Neuroimage*, 17(1):302–316, 2002.

<http://dx.doi.org/10.1006/nimg.2002.1208>

C. Hinrichs, V. Singh, L. Mukherjee, G. F. Xu, M. K. Chung, S. C. Johnson, and the Alzheimer’s Disease Neuroimaging Initiative. Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. *Neuroimage*, 48(1):138–149, 2009.

<http://dx.doi.org/10.1016/j.neuroimage.2009.05.056>

C. Hinrichs, V. Singh, G. Xu, S. C. Johnson, and the Alzheimer’s Disease Neuroimaging Initiative. Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. *Neuroimage*, 55(2):574–589, 2011.

<http://dx.doi.org/10.1016/j.neuroimage.2010.10.081>

T. Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998.

<http://dx.doi.org/10.1109/34.709601>

Y. Hochberg. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75(4):800–803, 1988.

<http://dx.doi.org/10.1093/biomet/75.4.800>

E. J. Hoffman, P. D. Cutler, W. M. Digby, and J. C. Mazziotta. 3-D phantom to simulate cerebral blood-flow and metabolic images for PET. *IEEE Transactions on Nuclear Science*, 37(2):616–620, 1990.

<http://dx.doi.org/10.1109/23.106686>

E. J. Hoffman, S.-C. Huang, and M. E. Phelps. Quantitation in positron emission computed tomography: 1. effect of object size. *Journal of Computer Assisted Tomography*, 3(3):299–308,

1979.

<http://dx.doi.org/10.1097/00004728-197906000-00001>

J. M. Hoffman, K. A. Welsh-Bohmer, M. Hanson, B. Crain, C. Hulette, N. Earl, and R. E. Coleman. FDG PET imaging in patients with pathologically verified dementia. *Journal Of Nuclear Medicine*, 41(11):1920–1928, 2000.

<http://jnm.snmjournals.org/content/41/11/1920.abstract>.

J. P. Hornak. *The basics of MRI*. Interactive Learning Software, New York, 2010.

<http://www.cis.rit.edu/htbooks/mri>.

C. Hsu, C. Chang, and C. Lin. *A practical guide to support vector classification*, 2010.

<http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.

X. Hua, A. D. Leow, N. Parikshak, S. Lee, M. C. Chiang, A. W. Toga, C. R. Jack Jr., M. W. Weiner, P. M. Thompson, and the Alzheimer’s Disease Neuroimaging Initiative. Tensor-based morphometry as a neuroimaging biomarker for Alzheimers disease: an MRI study of 676 AD, MCI, and normal subjects. *Neuroimage*, 43(3):458–469, 2008.

<http://dx.doi.org/10.1016/j.neuroimage.2008.07.013>

K. Ishiguro, H. Ohno, H. Arai, H. Yamaguchi, K. Urakami, J. M. Park, K. Sato, H. Kohno, and K. Imahori. Phosphorylated tau in human cerebrospinal fluid is a diagnostic marker for Alzheimer’s disease. *Neuroscience Letters*, 270(2):91–94, 1999.

[http://dx.doi.org/10.1016/S0304-3940\(99\)00476-0](http://dx.doi.org/10.1016/S0304-3940(99)00476-0)

C. R. Jack Jr., M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L. G. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, G. Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, M. W. Weiner, and the Alzheimer’s Disease Neuroimaging Initiative. The Alzheimer’s Disease Neuroimaging Initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*, 27(4):685–691, 2008a.

<http://dx.doi.org/10.1002/jmri.21049>

C. R. Jack Jr., V. J. Lowe, M. L. Senjem, S. D. Weigand, B. J. Kemp, M. M. Shiung, D. S. Knopman, B. F. Boeve, W. E. Klunk, C. A. Mathis, and R. C. Petersen. C-11-PiB and structural MRI provide complementary information in imaging of Alzheimers disease and amnestic mild cognitive impairment. *Brain*, 131(3):665–680, 2008b.

<http://dx.doi.org/10.1093/brain/awm336>

C. R. Jack Jr., V. J. Lowe, S. D. Weigand, H. J. Wiste, M. L. Senjem, D. S. Knopman, M. M. Shiung, J. L. Gunter, B. F. Boeve, B. J. Kemp, M. Weiner, R. C. Petersen, and the Alzheimer's Disease Neuroimaging Initiative. Serial PiB and MRI in normal, mild cognitive impairment and Alzheimers disease: implications for sequence of pathological events in Alzheimers disease. *Brain*, 132(5):1355–1365, 2009.

<http://dx.doi.org/10.1093/brain/awp062>

C. R. Jack Jr., S. D. Weigand, M. M. Shiung, S. A. Przybelski, P. C. O'Brien, J. L. Gunter, D. S. Knopman, B. F. Boeve, G. E. Smith, and R. C. Petersen. Atrophy rates accelerate in amnestic mild cognitive impairment. *Neurology*, 70(19):1740–1752, 2008c.

<http://dx.doi.org/10.1212/01.wnl.0000281688.77598.35>

W. J. Jagust, D. Bandy, K. Chen, N. L. Foster, S. M. Landau, C. A. Mathis, J. C. Price, E. M. Reiman, D. Skovronsky, R. A. Koeppe, and the Alzheimer's Disease Neuroimaging Initiative. The Alzheimer's Disease Neuroimaging Initiative positron emission tomography core. *Alzheimer's and Dementia*, 6(3):221–229, 2010.

<http://dx.doi.org/10.1016/j.jalz.2010.03.003>

A. K. Jain. *Fundamentals of digital image processing*. Prentice Hall, New York, 5th edition, 1989.

R. C. Jennison. *Fourier transforms and convolutions for the experimentalist*. Pergamon Press, New York, 1961.

A. Joshi, R. A. Koeppe, and J. A. Fessler. Reducing between scanner differences in multi-center PET studies. *Neuroimage*, 46(1):154–159, 2009.

<http://dx.doi.org/10.1016/j.neuroimage.2009.01.057>

J. Jovicich, S. Czanner, D. Greve, E. Haley, A. van der Kouwe, R. Gollub, D. Kennedy, F. Schmitt, G. Brown, J. MacFall, B. Fischl, and A. Dale. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. *Neuroimage*, 30(2):436–443, 2006.

<http://dx.doi.org/10.1016/j.neuroimage.2005.09.046>

P. Juszczak, D. M. J. Tax, and R. P. W. Dui. Feature scaling in support vector data descriptions. (In Proceedings) 8th Annual Conference for the Advanced School for Computing and Imaging (ASCI'02), 95–102, 2002.

P. J. Kahle, M. Jakowec, S. J. Teipel, H. Hampel, G. M. Petzinger, D. A. Di Monte, G. D. Silverberg, H. J. Möller, J. A. Yesavage, J. R. Tinklenberg, E. M. Shooter, and G. M. Murphy Jr. Combined assessment of tau and neuronal thread protein in Alzheimer's disease CSF. *Neurology*, 54(7):1498–1504, 2000.

<http://www.neurology.org/content/54/7/1498.abstract>.

B. Karaçali and C. Davatzikos. Estimating topology preserving and smooth displacement fields. *IEEE Transactions on Medical Imaging*, 23(7):868–880, 2004.

<http://dx.doi.org/10.1109/TMI.2004.827963>

J. S. Kippenhan, W. W. Barker, J. Nagel, C. Grady, and R. Duara. Neural-network classification of normal and Alzheimer's disease subjects using high-resolution and low-resolution PET cameras. *Journal of Nuclear Medicine*, 35(1):7–15, 1994.

<http://jnm.snmjournals.org/content/35/1/7.abstract>.

J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.

<http://dx.doi.org/10.1109/34.667881>

H.-W. Klafki, M. Staufienbiel, J. Kornhuber, and J. Wiltfang. Therapeutic approaches to Alzheimer's disease. *Brain*, 129(11):2840–2855, 2006.

<http://dx.doi.org/10.1093/brain/awl280>

S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack Jr., J. Ashburner, and R. S. J. Frackowiak. Automatic classification of MR scans in Alzheimer's disease. *Brain*, 131(3):681–689, 2008.

<http://dx.doi.org/10.1093/brain/awm319>

W. E. Klunk, H. Engler, A. Nordberg, Y. M. Wang, G. Blomqvist, D. P. Holt, M. Bergstrom, I. Savitcheva, G. F. Huang, S. Estrada, B. Ausén, M. L. Debnath, J. Barletta, J. C. Price, J. Sandell, B. J. Lopresti, A. Wall, P. Koivisto, G. Antoni, C. A. Mathis, and B. Långström. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Annals of Neurology*, 55(3):306–319, 2004.

<http://dx.doi.org/10.1002/ana.20009>

R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. (In Proceedings) 14th International Joint Conference on Artificial Intelligence (IJCAI'95), 2:1137–1143, 1995.

S. M. Landau, D. Harvey, C. M. Madison, E. M. Reiman, N. L. Foster, P. S. Aisen, R. C. Petersen, L. M. Shaw, J. Q. Trojanowski, C. R. Jack Jr., M. W. Weiner, and W. J. Jagust. Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology*, 75(3):230–238, 2010.

<http://dx.doi.org/10.1212/WNL.0b013e3181e8e8b8>

J. B. S. Langbaum, K. Chen, W. Lee, C. Reschke, D. Bandy, A. S. Fleisher, G. E. Alexander, N. L. Foster, M. W. Weiner, R. A. Koeppe, W. J. Jagust, E. M. Reiman, and the Alzheimer's Disease Neuroimaging Initiative. Categorical and correlational analyses of baseline fluorodeoxyglucose positron emission tomography images from the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Neuroimage*, 45(4):1107–1116, 2009.

<http://dx.doi.org/10.1016/j.neuroimage.2008.12.072>

G. Langs, B. Menze, D. Lashkari, and P. Golland. Detecting stable distributed patterns of brain activations using Gini contrast. *Neuroimage*, 56(2):497–507, 2011.

<http://dx.doi.org/10.1016/j.neuroimage.2010.07.074>

A. D. Leow, I. Yanovsky, N. Parikshak, X. Hua, S. Lee, A. W. Toga, C. R. Jack Jr., M. A. Bernstein, P. J. Britson, J. L. Gunter, C. P. Ward, B. Borowski, L. M. Shaw, J. Q. Trojanowski, A. S. Fleisher, D. Harvey, J. Kornak, N. Schuff, G. E. Alexander, M. W. Weiner, P. M. Thompson, and the Alzheimer's Disease Neuroimaging Initiative. Alzheimer's Disease Neuroimaging Initiative: a one-year follow up study using tensor-based morphometry correlating degenerative rates, biomarkers and cognition. *Neuroimage*, 45(3):645–655, 2009.

<http://dx.doi.org/10.1016/j.neuroimage.2009.01.004>

A. Liaw and M. Wiener. Classification and regression by randomForest. *R News*, 2:18–22, 2002.

B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Péligrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehericy, and H. Benali. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology*, 51(2):73–83, 2009.

<http://dx.doi.org/10.1007/s00234-008-0463-x>

K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate analysis*. Academic Press, London, 5th edition, 1979.

J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike, C. Holmes, L. Collins, P. Thompson, D. MacDonald, M. Iacoboni, T. Schormann, K. Amunts, N. Palomero-Gallagher, S. Geyer, L. Parsons, K. Narr, N. Kabani, G. L. Goualher, D. Boomsma, T. Cannon, R. Kawashima, and B. Mazoyer. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 356(1412):1293–1322, 2001.

<http://dx.doi.org/10.1098/rstb.2001.0915>

G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan. Clinical diagnosis of Alzheimer's disease - report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services task force on Alzheimer's disease. *Neurology*, 34(7):939–944, 1984.

<http://www.neurology.org/content/34/7/939.abstract>.

G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. J. Jr., C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, R. C. Mohs, J. C. Morris, M. N. Rossor, P. Scheltens, M. C. Carrillo, B. Thies, S. Weintraub, and C. H. Phelps. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia*, 7(3):263–269, 2011.

<http://dx.doi.org/10.1016/j.jalz.2011.03.005>

S. Minoshima, N. L. Foster, A. A. F. Sima, K. A. Frey, R. L. Albin, and D. E. Kuhl. Alzheimer's disease versus dementia with Lewy bodies: cerebral metabolic distinction with autopsy confirmation. *Annals of Neurology*, 50(3):358–365, 2001.

<http://dx.doi.org/10.1002/ana.1133>

S. Minoshima, K. A. Frey, N. L. Foster, and D. E. Kuhl. Preserved pontine glucose-metabolism in Alzheimer disease - a reference region for functional brain image (PET) analysis. *Journal of Computer Assisted Tomography*, 19(4):541–547, 1995.

<http://dx.doi.org/10.1097/00004728-199507000-00006>

M. A. Mintun, G. N. LaRossa, Y. I. Sheline, C. S. Dence, S. Y. Lee, R. H. Mach, W. E. Klunk, C. A. Mathis, S. T. DeKosky, and J. C. Morris. C-11-PiB in a nondemented population - Potential antecedent marker of Alzheimer disease. *Neurology*, 67(3):446–452, 2006.

<http://dx.doi.org/10.1212/01.wnl.0000228230.26044.a4>

L. Mosconi. Brain glucose metabolism in the early and specific diagnosis of Alzheimer's disease - FDG-PET studies in MCI and AD. *European Journal of Nuclear Medicine and Molecular Imaging*, 32(4):486–510, 2005.

<http://dx.doi.org/10.1007/s00259-005-1762-7>

L. Mosconi, D. Perani, S. Sorbi, K. Herholz, B. Nacmias, V. Holthoff, E. Salmon, J. C. Baron, M. T. R. De Cristofaro, A. Padovani, B. Borroni, M. Franceschi, L. Bracco, and A. Pupi. MCI conversion to dementia and the APOE genotype - a prediction study with FDG-PET. *Neurology*, 63(12):2332–2340, 2004.

<http://www.neurology.org/content/63/12/2332.abstract>.

L. Mosconi, W. H. Tsui, S. De Santi, J. Li, H. Rusinek, A. Convit, Y. Li, M. Boppana, and M. J. de Leon. Reduced hippocampal metabolism in MCI and AD - automated FDG-PET image analysis. *Neurology*, 64(11):1860–1867, 2005.

<http://dx.doi.org/10.1212/01.WNL.0000163856.13524.08>

L. Mosconi, W. H. Tsui, K. Herholz, A. Pupi, A. Drzezga, G. Lucignani, E. M. Reiman, V. Holthoff, E. Kalbe, S. Sorbi, J. Diehl-Schmid, R. Perneczky, F. Clerici, R. Caselli, B. Beuthien-Baurnann, A. Kurz, S. Minoshima, and M. J. de Leon. Multicenter standardized F-18-FDG PET diagnosis of mild cognitive impairment, Alzheimer’s disease, and other dementias. *Journal of Nuclear Medicine*, 49(3):390–398, 2008.

<http://dx.doi.org/10.2967/jnumed.107.045385>

L. Mosconi, W. H. Tsui, A. Pupi, S. De Santi, A. Drzezga, S. Minoshima, and M. J. de Leon. F-18-FDG PET database of longitudinally confirmed healthy elderly individuals improves detection of mild cognitive impairment and Alzheimer’s disease. *Journal of Nuclear Medicine*, 48(7):1129–1134, 2007.

<http://dx.doi.org/10.2967/jnumed.107.040675>

R. Motter, C. Vigopelfrey, D. Kholodenko, R. Barbour, K. Johnsonwood, D. Galasko, L. Chang, B. Miller, C. Clark, R. Green, D. Olson, P. Southwick, R. Wolfert, B. Munroe, I. Lieberburg, P. Seubert, and D. Schenk. Reduction of beta-amyloid peptide(42) in the cerebrospinal fluid of patients with Alzheimer’s disease. *Annals of Neurology*, 38(4):643–648, 1995.

<http://dx.doi.org/10.1002/ana.410380413>

National Institute on Aging. Alzheimer’s disease: unraveling the mystery. 2008.

<http://www.nia.nih.gov/alzheimers/publication/alzheimers-disease-unraveling-mystery>.

P. J. Nestor, T. D. Fryer, P. Smielewski, and J. R. Hodges. Limbic hypometabolism in Alzheimer’s disease and mild cognitive impairment. *Annals of Neurology*, 54(3):343–351, 2003.

<http://dx.doi.org/10.1002/ana.10669>

J. M. Ollinger and J. A. Fessler. Positron-emission tomography. *IEEE Signal Processing*

Magazine, 14(1):43–55, 1997.

<http://dx.doi.org/10.1109/79.560323>

M. B. Patwardhan, D. C. McCrory, D. B. Matchar, G. P. Samsa, and O. T. Rutschmann. Alzheimer disease: operating characteristics of PET – a meta-analysis. *Radiology*, 231(1):73–80, 2004.

<http://dx.doi.org/10.1148/radiol.2311021620>

R. C. Petersen. Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256(3):183–194, 2004.

<http://dx.doi.org/10.1111/j.1365-2796.2004.01388.x>

R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen. Mild cognitive impairment: clinical characterization and outcome. *Archives of Neurology*, 56(3):303–308, 1999.

<http://dx.doi.org/10.1001/archneur.56.3.303>

K. E. Pike, G. Savage, V. L. Villemagne, S. Ng, S. A. Moss, P. Maruff, C. A. Mathis, W. E. Klunk, C. L. Masters, and C. C. Rowe. Beta-amyloid imaging and memory in non-demented individuals: evidence for preclinical Alzheimer’s disease. *Brain*, 130(11):2837–2844, 2007.

<http://dx.doi.org/10.1093/brain/awm238>

C. Plant, S. J. Teipel, A. Oswald, C. Bhm, T. Meindl, J. Mourao-Miranda, A. W. Bokde, H. Hampel, and M. Ewers. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer’s disease. *Neuroimage*, 50(1):162–174, 2010.

<http://dx.doi.org/10.1016/j.neuroimage.2009.11.046>

J. Qi and R. M. Leahy. Iterative reconstruction techniques in emission computed tomography. *Physics in Medicine and Biology*, 51(15):R541–R578, 2006.

<http://dx.doi.org/10.1088/0031-9155/51/15/R01>

N. A. Ranginwala, L. S. Hynan, M. F. Weiner, and C. L. White III. Clinical criteria for the diagnosis of Alzheimer disease: still good after all these years. *American Journal Of Geriatric*

Psychiatry, 16(5):384–388, 2008.

<http://dx.doi.org/10.1097/JGP.0b013e3181629971>

D. X. Rasmusson, J. Brandt, C. Steele, J. C. Hedreen, J. C. Troncoso, and M. F. Folstein. Accuracy of clinical diagnosis of Alzheimer disease and clinical features of patients with non-Alzheimer disease neuropathology. *Alzheimer Disease and Associated Disorders*, 10(4):180–188, 1996.

<http://dx.doi.org/10.1097/00002093-199601040-00002>

E. M. Reiman, K. W. Chen, G. E. Alexander, R. J. Caselli, D. Bandy, D. Osborne, A. M. Saunders, and J. Hardy. Correlations between apolipoprotein E epsilon 4 gene dose and brain-imaging measurements of regional hypometabolism. *Proceedings of the National Academy of Sciences of the United States of America*, 102(23):8299–8302, 2005.

<http://dx.doi.org/10.1073/pnas.0500579102>

B. H. Ridha, J. Barnes, J. W. Bartlett, A. Godbolt, T. Pepple, M. N. Rossor, and N. C. Fox. Tracking atrophy progression in familial Alzheimer’s disease: a serial MRI study. *Lancet Neurology*, 5(10):828–834, 2006.

[http://dx.doi.org/10.1016/S1474-4422\(06\)70550-6](http://dx.doi.org/10.1016/S1474-4422(06)70550-6)

W. A. Rocca, A. Hofman, C. Brayne, M. M. B. Breteler, M. Clarke, J. R. M. Copeland, J.-F. Dartigues, K. Engedal, O. Hagnell, T. J. Heeren, C. Jonker, J. Lindesay, A. Lobo, A. H. Mann, P. K. Mölsä, K. Morgan, D. W. O’Connor, A. d. S. Droux, R. Sulkava, D. W. K. Kay, and L. Amaducci. Frequency and distribution of Alzheimer’s disease in Europe: a collaborative study of 1980–1990 prevalence findings. *Annals of Neurology*, 30(3):381–390, 1991.

<http://dx.doi.org/10.1002/ana.410300310>

T. Rohlfing, R. Brandt, R. Menzel, and C. Maurer Jr. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *Neuroimage*, 21(4):1428–1442, 2004.

<http://dx.doi.org/10.1016/j.neuroimage.2003.11.010>

C. Rorden. MRICro: neuroanatomy atlas tutorial. 2002.

<http://www.cabiatl.com/mricro/anatomy/home.html>.

A. D. Roses and A. M. Saunders. ApoE, Alzheimer's disease, and recovery from brain stress. *Cerebrovascular Pathology in Alzheimer's Disease*, 826:200–212, 1997.

<http://dx.doi.org/10.1111/j.1749-6632.1997.tb48471.x>

O. Rousset, A. Rahmim, A. Alavi, and H. Zaidi. Partial volume correction strategies in PET. *PET Clinics*, 2(2):235–249, 2007.

<http://dx.doi.org/10.1016/j.cpet.2007.10.005>

C. C. Rowe, K. A. Ellis, M. Rimajova, P. Bourgeat, K. E. Pike, G. Jones, J. Fripp, H. Tochon-Danguy, L. Morandau, G. O'Keefe, R. Price, P. Raniga, P. Robins, O. Acosta, N. Lenzo, C. Szoeki, O. Salvado, R. Head, R. Martins, C. L. Masters, D. Ames, and V. L. Villemagne. Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging. *Neurobiology of Aging*, 31(8):1275–1283, 2010.

<http://dx.doi.org/10.1016/j.neurobiolaging.2010.04.007>

M. Rudin. *Molecular imaging: basic principles and applications in biomedical research*. Imperial College Press, London, 2005.

D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.

<http://dx.doi.org/10.1109/42.796284>

D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, I. A. Illán, M. López, F. Segovia, R. Chaves, P. Padilla, and C. G. Puntonet. Feature selection using factor analysis for Alzheimer's diagnosis using 18F-FDG PET images. *Medical Physics*, 37(11):6084–6095, 2010.

<http://dx.doi.org/10.1118/1.3488894>

D. Scheuner, C. Eckman, M. Jensen, X. Song, M. Citron, N. Suzuki, T. D. Bird, J. Hardy, M. Hutton, W. Kukull, E. Larson, E. LevyLahad, M. Viitanen, E. Peskind, P. Poorkaj, G. Schellenberg, R. Tanzi, W. Wasco, L. Lannfelt, D. Selkoe, and S. Younkin. Secreted amyloid beta-

protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nature Medicine*, 2(8):864–870, 1996.

<http://dx.doi.org/10.1038/nm0896-864>

J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, G. P. Hartkens, T Penney, W. A. Hall, H. Liu, C. L. Truweit, F. A. Gerritsen, D. L. G. Hill, and D. J. Hawkes. A generic framework for nonrigid registration based on non-uniform multi-level free-form deformations. (In Proceedings) 4th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'01), *Lecture Notes in Computer Science*, 2208:573–581, 2001.

http://dx.doi.org/10.1007/3-540-45468-3_69

J. M. Schott, J. W. Bartlett, N. C. Fox, J. Barnes, and for the Alzheimer's Disease Neuroimaging Initiative investigators. Increased brain atrophy rates in cognitively normal older adults with low cerebrospinal fluid $A\beta_{1-42}$. *Annals of Neurology*, 68(6):825–834, 2010.

<http://dx.doi.org/10.1002/ana.22315>

N. Schuff, N. Woerner, L. Boreta, T. Kornfield, L. M. Shaw, J. Q. Trojanowski, P. M. Thompson, C. R. Jack Jr., M. W. Weiner, and the Alzheimer's Disease Neuroimaging Initiative. MRI of hippocampal volume loss in early Alzheimers disease in relation to ApoE genotype and biomarkers. *Brain*, 132(4):1067–1077, 2009.

<http://dx.doi.org/10.1093/brain/awp007>

D. J. Selkoe. The molecular pathology of Alzheimer's disease. *Neuron*, 6(4):487–498, 1991.

[http://dx.doi.org/10.1016/0896-6273\(91\)90052-2](http://dx.doi.org/10.1016/0896-6273(91)90052-2)

D. J. Selkoe. Cell biology of protein misfolding: the examples of Alzheimer's and Parkinson's diseases. *Nature Cell Biology*, 6(11):1054–1061, 2004.

<http://dx.doi.org/10.1038/ncb1104-1054>

C. E. Shannon. Communication in the presence of noise. (In Proceedings) Institute of Radio

Engineers, 37(1):10–21, 1949.

<http://dx.doi.org/10.1109/JRPROC.1949.232969>

L. M. Shaw, H. Vanderstichele, M. Knapik-Czajka, C. M. Clark, P. S. Aisen, R. C. Petersen, K. Blennow, H. Soares, A. Simon, P. Lewczuk, R. Dean, E. Siemers, W. Potter, V. M. Y. Lee, J. Q. Trojanowski, and the Alzheimer's Disease Neuroimaging Initiative. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Annals of Neurology*, 65(4):403–413, 2009.

<http://dx.doi.org/10.1002/ana.21610>

T. Shi and S. Horvath. Unsupervised learning with random forest predictors. *Journal of Computational and Graphical Statistics*, 15(1):118–138, 2006.

<http://dx.doi.org/10.1198/106186006X94072>

T. Shi, D. Seligson, A. S. Beldegrun, A. Palotie, and S. Horvath. Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma. *Modern Pathology*, 18(4):547–557, 2005.

<http://dx.doi.org/10.1038/modpathol.3800322>

D. H. S. Silverman, G. W. Small, C. Y. Chang, C. S. Lu, M. A. K. de Aburto, W. Chen, J. Czernin, S. I. Rapoport, P. Pietrini, G. E. Alexander, M. B. Schapiro, W. J. Jagust, J. M. Hoffman, K. A. Welsh-Bohmer, A. Alavi, C. M. Clark, E. Salmon, M. J. de Leon, R. Mielke, J. L. Cummings, A. P. Kowell, S. S. Gambhir, C. K. Hoh, and M. E. Phelps. Positron emission tomography in evaluation of dementia - regional brain metabolism and long-term outcome. *JAMA - Journal Of The American Medical Association*, 286(17):2120–2127, 2001.

<http://dx.doi.org/10.1001/jama.286.17.2120>

J. G. Sled, A. P. Zijdenbos, and A. C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, 17(1):87–97, 1998.

<http://dx.doi.org/10.1109/42.668698>

S. M. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–155,

2002.

<http://dx.doi.org/10.1002/hbm.10062>

G. Soza, R. Grosso, C. Nimsy, G. Greiner, and P. Hastreiter. Estimating mechanical brain tissue properties with simulation and registration. (In Proceedings) 7th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI '04), *Lecture Notes in Computer Science*, 3217:276–283, 2004.

http://dx.doi.org/10.1007/978-3-540-30136-3_35

R. A. Sperling, P. S. Aisen, L. A. Beckett, D. A. Bennett, S. Craft, A. M. Fagan, T. Iwatsubo, C. R. Jack Jr., J. Kaye, T. J. Montine, D. C. Park, E. M. Reiman, C. C. Rowe, E. Siemers, Y. Stern, K. Yaffe, M. C. Carrillo, B. Thies, M. Morrison-Bogorad, M. V. Wagster, and C. H. Phelps. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia*, 7(3):280–292, 2011.

<http://dx.doi.org/10.1016/j.jalz.2011.03.003>

Student. The probable error of a mean. *Biometrika*, 6(1):1–25, 1908.

<http://dx.doi.org/10.1093/biomet/6.1.1>

C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, 1999.

[http://dx.doi.org/10.1016/S0031-3203\(98\)00091-0](http://dx.doi.org/10.1016/S0031-3203(98)00091-0)

T. Sunderland, G. Linker, N. Mirza, K. T. Putnam, D. L. Friedman, L. H. Kimmel, J. Bergeson, G. J. Manetti, M. Zimmermann, B. Tang, J. J. Bartko, and R. M. Cohen. Decreased beta-amyloid 1-42 and increased tau levels in cerebrospinal-fluid of patients with Alzheimer disease. *JAMA - Journal of the American Medical Association*, 289(16):2094–2103, 2003.

<http://dx.doi.org/10.1001/jama.289.16.2094>

J. Talairach and P. Tournoux. *Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system - an approach to cerebral imaging*. Thieme Medical Publishers, New York, 1988.

T. Tapiola, T. Pirttila, P. D. Mehta, I. Alafuzoff, M. Lehtovirta, and H. Soininen. Relationship between apoE genotype and CSF beta-amyloid (1-42) and tau in patients with probably and definite Alzheimer's disease. *Neurobiology of Aging*, 21(5):735–740, 2000.

[http://dx.doi.org/10.1016/S0197-4580\(00\)00164-0](http://dx.doi.org/10.1016/S0197-4580(00)00164-0)

J. Tenenbaum, V. de Silva, and J. Langford. A global geometric framework for non-linear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

<http://dx.doi.org/10.1126/science.290.5500.2319>

The Ronald and Nancy Reagan Research Institute of the Alzheimer's Association and the National Institute on Aging Working Group. Consensus report of the Working Group on: biological markers of Alzheimer's Disease. *Neurobiology of Aging*, 19(2):109–116, 1998.

[http://dx.doi.org/10.1016/S0197-4580\(98\)00022-0](http://dx.doi.org/10.1016/S0197-4580(98)00022-0)

F. Thiele, F. Wenzel, S. Young, D. Cross, R. Buchert, and S. Minoshima. Metabolic heterogeneity in ADNI PET subjects. *Journal of Nuclear Medicine Meetings Abstracts*, 50:429, 2009.

M. Tondelli, G. K. Wilcock, P. Nichelli, C. A. D. Jager, M. Jenkinson, and G. Zamboni. Structural MRI changes detectable up to ten years before clinical Alzheimer's disease. *Neurobiology of Aging*, 33(4):825.e25–825.e36, 2012.

<http://dx.doi.org/10.1016/j.neurobiolaging.2011.05.018>

W. S. Torgerson. Multidimensional scaling: I. Theory and method. *Psychometrika*, 17(4):401–419, 1952.

<http://dx.doi.org/10.1007/BF02288916>

J. Q. Trojanowski, H. Vandeerstichele, M. Korecka, C. M. Clark, P. S. Aisen, R. C. Petersen, K. Blennow, H. Soares, A. Simon, P. Lewczuk, R. Dean, E. Siemers, W. Z. Potter, M. W. Weiner, C. R. Jack Jr., W. Jagust, A. W. Toga, V. M.-Y. Lee, and L. M. Shaw. Update on the biomarker core of the Alzheimer's Disease Neuroimaging Initiative subjects. *Alzheimer's and Dementia*, 6(3):230–238, 2010.

<http://dx.doi.org/10.1016/j.jalz.2010.03.008>

F. E. Turkheimer, C. B. Smith, and K. Schmidt. Estimation of the number of “true” null hypotheses in multivariate analysis of neuroimaging data. *Neuroimage*, 13(5):920–930, 2001.

<http://dx.doi.org/10.1006/nimg.2001.0764>

L. A. van de Pol, W. M. van der Flier, E. S. C. Korf, N. C. Fox, F. Barkhof, and P. Scheltens. Baseline predictors of rates of hippocampal atrophy in mild cognitive impairment. *Neurology*, 69(15):1491–1497, 2007.

<http://dx.doi.org/10.1212/01.wnl.0000277458.26846.96>

M. Vandermeeren, M. Mercken, E. Vanmechelen, J. Six, A. Vandevoorde, J. J. Martin, and P. Cras. Detection of tau proteins in normal and Alzheimer’s disease cerebrospinal fluid with a sensitive sandwich enzyme-linked immunosorbent assay. *Journal of Neurochemistry*, 61(5):1828–1834, 1993.

<http://dx.doi.org/10.1111/j.1471-4159.1993.tb09823.x>

V. N. Vapnik and A. Lerner. Pattern recognition using generalized portrait method. *Automation and Remote Control*, 24(6):774–780, 1963.

P. Vemuri, J. L. Gunter, M. L. Senjem, J. L. Whitwell, K. Kantarci, D. S. Knopman, B. F. Boeve, R. C. Petersen, and C. R. Jack Jr. Alzheimer’s disease diagnosis in individual subjects using structural MR images: validation studies. *Neuroimage*, 39(3):1186–1197, 2008.

<http://dx.doi.org/10.1016/j.neuroimage.2007.09.073>

V. L. Villemagne, K. E. Pike, D. Darby, P. Maruff, G. Savage, S. Ng, U. Ackermann, T. F. Cowie, J. Currie, S. G. Chan, G. Jones, H. Tochon-Danguy, G. O’Keefe, C. L. Masters, and C. C. Rowe. Abeta deposits in older non-demented individuals with cognitive decline are indicative of preclinical Alzheimer’s disease. *Neuropsychologia*, 46(6):1688–1697, 2008.

<http://dx.doi.org/10.1016/j.neuropsychologia.2008.02.008>

P. Viola and W. M. Wells III. Alignment by maximization of mutual information. (In Proceedings) 5th International Conference on Computer Vision (ICCV’95), 16–23, 1995.

<http://dx.doi.org/10.1109/ICCV.1995.466930>

K. B. Walhovd, A. M. Fjell, A. M. Dale, L. K. McEvoy, J. Brewer, D. S. Karow, D. P. Salmon, and C. Fennema-Notestine. Multi-modal imaging predicts memory performance in normal aging and cognitive decline. *Neurobiology of Aging*, 31(7):1107–1121, 2010.

<http://dx.doi.org/10.1016/j.neurobiolaging.2008.08.013>

G. J. Wang, N. D. Volkow, A. P. Wolf, J. D. Brodie, and R. J. Hitzemann. Intersubject variability of brain glucose metabolic measurements in young normal males. *Journal of Nuclear Medicine*, 35(9):1457–1466, 1994.

<http://jnm.snmjournals.org/content/35/9/1457.abstract>.

S. K. Warfield, K. H. Zou, and W. M. Wells III. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921, 2004.

<http://dx.doi.org/10.1109/TMI.2004.828354>

D. Wechsler. *WMS-R Wechsler Memory Scale – revised manual*. The Psychological Corporation, New York, 1987.

S. D. Weigand, P. Vemuri, H. J. Wiste, M. L. Senjem, V. S. Pankratz, P. S. Aisen, M. W. Weiner, R. C. Petersen, L. M. Shaw, J. Q. Trojanowski, D. S. Knopman, C. R. Jack Jr., and the Alzheimer’s Disease Neuroimaging Initiative. Transforming CSF A β 42 measures into calculated PiB units of brain A β amyloid. *Alzheimer’s and Dementia*, 7(2):133–141, 2011.

<http://dx.doi.org/10.1016/j.jalz.2010.08.230>

J. L. Whitwell, W. R. Crum, H. C. Watt, and N. C. Fox. Normalization of cerebral volumes by use of intracranial volume: implications for longitudinal quantitative MR imaging. *American Journal of Neuroradiology*, 22(8):1483–1489, 2001.

<http://www.ajnr.org/content/22/8/1483.abstract>.

R. Wolz, R. A. Heckemann, P. Aljabar, J. V. Hajnal, A. Hammers, J. Lötjönen, D. Rueckert, and the Alzheimer’s Disease Neuroimaging Initiative. Measurement of hippocampal atrophy using 4D graph-cut segmentation: application to ADNI. *Neuroimage*, 52(1):109–118, 2010.

<http://dx.doi.org/10.1016/j.neuroimage.2010.04.006>

K. J. Worsley, A. C. Evans, S. Marrett, and P. Neelin. A three dimensional statistical analysis for CBF activation studies in the human brain. *Journal of Cerebral Blood Flow and Metabolism*, 12:900–918, 1992.

<http://dx.doi.org/10.1038/jcbfm.1992.127>

K. J. Worsley, S. Marrett, P. Neelin, A. C. Vandal, K. J. Friston, and A. C. Evans. A unified statistical approach for determining significant voxels in images of cerebral activation. *Human Brain Mapping*, 4(1):58–73, 1996.

[http://dx.doi.org/10.1002/\(SICI\)1097-0193\(1996\)4:1<58::AID-HBM4>3.0.CO;2-0](http://dx.doi.org/10.1002/(SICI)1097-0193(1996)4:1<58::AID-HBM4>3.0.CO;2-0)

Y. Xia, L. Wen, S. Eberl, M. Fulham, and D. Feng. Genetic algorithm-based PCA eigenvector selection and weighting for automated identification of dementia using FDG-PET imaging. (In Proceedings) 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS'08), 4812–4815, 2008.

<http://dx.doi.org/10.1109/IEMBS.2008.4650290>

I. Yakushev, A. Hammers, A. Fellgiebel, I. Schmidtman, A. Scheurich, H. G. Buchholz, J. Peters, P. Bartenstein, K. Lieb, and M. Schreckenberger. SPM-based count normalization provides excellent discrimination of mild Alzheimer's disease and amnesic mild cognitive impairment from healthy aging. *Neuroimage*, 44(1):43–50, 2009.

<http://dx.doi.org/10.1016/j.neuroimage.2008.07.015>

I. Yakushev, C. Landvogt, H. G. Buchholz, A. Fellgiebel, A. Hammers, A. Scheurich, I. Schmidtman, A. Gerhard, M. Schreckenberger, and P. Bartenstein. Choice of reference area in studies of Alzheimer's disease using positron emission tomography with fluorodeoxyglucose-F18. *Psychiatry Research: Neuroimaging*, 164(2):143–153, 2008.

<http://dx.doi.org/10.1016/j.psychresns.2007.11.004>

V. I. Zannis, P. W. Just, and J. L. Breslow. Human apolipoprotein E isoprotein subclasses are genetically determined. *American Journal of Human Genetics*, 33(1):11–24, 1981.

D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, and the Alzheimer's Disease Neuroimaging Initiative. Multimodal classification of Alzheimer's disease and mild cognitive impairment.

Neuroimage, 55(3):856–867, 2011.

<http://dx.doi.org/10.1016/j.neuroimage.2011.01.008>

Appendix A

Hammers brain atlases

The Hammers brain atlases comprise 30 T1-weighted MR images which have been manually segmented into the 83 anatomically defined structures listed in Table A.1. Details of the subject demographics and MR imaging protocol are provided in [Hammers et al. \(2003\)](#). Protocols for the manual delineation are described in [Hammers et al. \(2003\)](#) and [Gousias et al. \(2008\)](#).

No. right	No. left	Structure
1	2	Hippocampus
3	4	Amygdala
5	6	Anterior temporal lobe, medial part
7	8	Anterior temporal lobe, lateral part
9	10	Gyri parahippocampalis et ambiens
11	12	Superior temporal gyrus, posterior part
13	14	Middle and inferior temporal gyri
15	16	Lateral occipitotemporal gyrus, gyrus fusiformis
17	18	Cerebellum
	19	Brainstem, spans the midline
21	20	Insula
23	22	Occipital lobe
25	24	Cingulate gyrus, anterior part

No. right	No. left	Structure
27	26	Cingulate gyrus, posterior part
29	28	Frontal lobe, becomes middle frontal gyrus after subdivision
31	30	Posterior temporal lobe
33	32	Parietal lobe
35	34	Caudate nucleus
37	36	Nucleus accumbens
39	38	Putamen
41	40	Thalamus
43	42	Pallidum, globus pallidus
44		Corpus callosum, spans the midline
45	46	Lateral ventricle, frontal horn, central part and occipital horn
47	48	Lateral ventricle, temporal horn
49		Third ventricle, spans the midline
51	50	Precentral gyrus
53	52	Straight gyrus, gyrus rectus
55	54	Anterior orbital gyrus
57	56	Inferior frontal gyrus
59	58	Superior frontal gyrus
61	60	Postcentral gyrus
63	62	Superior parietal gyrus
65	64	Lingual gyrus
67	66	Cuneus left
69	68	Medial orbital gyrus
71	70	Lateral orbital gyrus
73	72	Posterior orbital gyrus
75	74	Substantia nigra
77	76	Subgenual frontal cortex
79	78	Subcallosal area

No. right	No. left	Structure
81	80	Pre-subgenual frontal cortex
83	82	Superior temporal gyrus, anterior part

Table A.1: Anatomically defined regions manually delineated in the Hammers atlases.

Appendix B

Excluded images

ADNI subject identifiers for all participants whose images were excluded from the analyses described in Chapter 4 are provided in Table B.1. These participants were additionally excluded from the analyses described in Chapter 5, either for the same reasons, or because they did not have longitudinal FDG-PET or MR imaging data available.

Subject ID	Reason for Exclusion
005.S_0223	progressed from HC to MCI
006.S_0484 *	failed segmentation due to brain mask
010.S_0422	reverted from MCI to HC
011.S_0002 *	timeframe information missing in header
011.S_0003	timeframe information missing in header
021.S_0178	cerebellum partially outside field of view
022.S_0096	cerebellum partially outside field of view
024.S_0985	cerebellum partially outside field of view
024.S_1063	progressed from HC to MCI
024.S_1393	frontal cortex partially outside field of view
027.S_0120	transformation to MNI space failed
041.S_0898	progressed from HC to MCI
057.S_0779	progressed from HC to MCI

Subject ID	Reason for Exclusion
094.S_0498 *	transformation to MNI space failed
094.S_1188	transformation to MNI space failed
099.S_0551	reverted from MCI to HC
100.S_0743 *	cerebellum partially outside field of view
109.S_1343 *	scan time under 30 minutes
126.S_1221	timeframe information missing in header
127.S_0112	oscillates between HC and MCI
127.S_0754	cerebellum partially outside field of view
132.S_0987 *	scan time under 30 minutes
137.S_0443	reverted from MCI to HC
137.S_0669	oscillates between HC and MCI
137.S_0722	reverted from MCI to HC
141.S_1245	reverted from MCI to HC
941.S_1194	scan time under 30 minutes
941.S_1202	progressed from HC to MCI

Table B.1: Participants excluded from the work described in Chapter 4. These participants were additionally excluded from the analyses described in Chapter 5, either for the same reasons, or because they did not have longitudinal FDG-PET or MR imaging data available (indicated with asterisks *).