

Robust Multi-Structure Segmentation of Magnetic Resonance Brain Images

Christian Ledig

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

Doctor of Philosophy

of

Imperial College London

November 2015

Imperial College London

Department of Computing

Abstract

Magnetic resonance (MR) imaging is a powerful technique for the non-invasive in-vivo imaging of the human brain. In this thesis several robust techniques are developed that allow the fully-automatic analysis of MR brain images. In particular an approach is presented that quantifies the volume of more than 100 individual structures within the whole brain. This methodology is extended to measure structural volume changes based on images acquired at multiple time points. The possibility to quantify volumetric change of numerous brain structures simultaneously sets the method apart from many established approaches that measure change of individual structures only. The presented methodology is robust in the presence of disease-related atrophy or severe brain deformation. The proposed algorithms are evaluated on widely used reference datasets. The obtained results compare favourable to state-of-the-art techniques. The methods are employed to perform a cross-sectional and a longitudinal analysis of a large cohort of Alzheimer’s disease (AD) patients. Employing the extracted information (volume, atrophy) as features it is shown that AD related disease states can be classified with high accuracy. In addition a prospective cohort of traumatic brain injury (TBI) patients is studied. Structural volumes are extracted at the acute disease stage and the extent of structural change occurring within the months following the head trauma is quantified. The results show that cerebral white matter atrophy is increased in TBI and that the involvement of individual brain structures, such as the hippocampus or the thalamus, is particularly predictive for the outcome of the head injury. These results confirm the methods’ potential to support large scale imaging studies by robustly extracting markers with clinical interpretation from MR brain images. In addition, a novel evaluation measure is proposed to quantitatively assess the accuracy of automatically calculated image segmentations with respect to a reference segmentation.

Declaration of Originality

I hereby declare that the work described in this thesis is my own, except where specifically acknowledged.

Christian Ledig

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Acknowledgements

I am grateful that I had the privilege of conducting my PhD studies in the outstanding environment within the BioMedIA group at Imperial College London. I feel very fortunate that I was supported by exceptional personalities who enabled my work and influenced me throughout this journey.

In particular I thank my supervisor, Daniel Rueckert, for the sensational supervision and guidance. During the last years he was always available to give invaluable advice. I am especially grateful for all the unique opportunities that I got to further develop myself. These went far beyond any expectations that I had when I started this program.

The work towards this thesis was challenging in many ways but always enjoyable. An important factor was the excellent atmosphere within the BioMedIA group. Conversations and discussions in this multi-cultural, diverse environment were a continuous source of experience. Of course, I also enjoyed the lunch and coffee routines, conference trips, sport sessions, and many other activities.

During the last years I had the pleasure to meet and collaborate with many great people within the PredictAD, TBIcare and PredictND projects. A substantial part of the work towards this thesis was funded by these projects. It was always exciting, and still is, to contribute to research objectives that require the close collaboration and synchronisation of researchers with highly distinct but complementary expertise.

I would also like to thank all my co-authors and collaborators, who enabled and supported the work presented in this thesis. Your expertise and enthusiasm was a seemingly endless resource and an important factor of my scientific education.

I also thank my outstanding friends who are great to spend time with but who also support me beyond the challenges related to my studies. Special thanks go to Debora for the never ending patience, tolerance and support. Above all, I thank my family and in particular Anneliese, Martin, Gudrun, Dieter and my brothers. It is you who always believed in me, supported me in so many stages of my life and laid the foundation of what I am. Thank you.

Contents

Abstract	2
List of Tables	11
List of Figures	18
Glossary	29
1 Introduction	36
1.1 Neuroimaging	36
1.1.1 Computed Tomography	37
1.1.2 Magnetic Resonance Imaging	37
1.1.3 Diffusion Weighted Imaging	38
1.1.4 Functional Magnetic Resonance Imaging	40
1.1.5 Positron Emission Tomography	40
1.2 Alzheimer's Disease	40
1.2.1 Prevalence	40
1.2.2 Diagnosis and disease progression	42
1.2.3 Diagnostic role of neuroimaging	43
1.2.4 Mild Cognitive Impairment	44
1.2.5 Potential of neuroimaging	46
1.3 Traumatic Brain Injury	47
1.3.1 Prevalence	47
1.3.2 Focal and diffuse injury	47
1.3.3 Clinical characterisation	48
1.3.4 Neuroimaging in TBI	50

1.3.5	Importance of multiple modalities	50
1.3.6	TBI: An understudied disease	52
1.4	Motivation	52
1.5	Contributions and organisation of the thesis	53
1.5.1	Contributions	53
1.5.2	Thesis overview	55
2	Background	56
2.1	Atlas-based segmentation methods	56
2.1.1	Brain extraction	57
2.1.2	Atlas alignment	59
2.1.3	Label fusion	65
2.1.4	Refinement of image segmentations	73
2.1.5	Remarks	75
2.2	Brain atlases from structural MRI	76
2.2.1	AAL atlas	78
2.2.2	IBSR18 atlas	80
2.2.3	Hammers atlas	81
2.2.4	LPBA40 atlas	82
2.2.5	NMM atlas	83
2.3	Biomarkers from structural MRI in AD	85
2.3.1	Motivation and potential	85
2.3.2	Literature overview	86
2.3.3	Limitations and challenges	90
2.4	Biomarkers from structural MRI in TBI	91
2.4.1	Motivation and potential	91
2.4.2	Literature overview	92
2.4.3	Limitations and challenges	94
3	Robust whole-brain segmentation: Multi-Atlas Label Propagation using Expectation- Maximisation (MALP-EM)	96
3.1	Introduction	97

3.2	Materials and methods	101
3.2.1	Materials	101
3.2.2	Multi-Atlas Label Propagation with Expectation-Maximisation based re- finement (MALP-EM)	103
3.3	Experiments and results	112
3.3.1	Quantitative evaluation on a benchmark dataset using manual labels	112
3.3.2	Segmentation and analysis of a TBI cohort	118
3.4	Discussion and future work	127
3.5	Conclusion	131
4	Longitudinal analysis of MR brain images	132
4.1	PBSI: A probabilistic extension of the Boundary Shift Integral	133
4.1.1	Introduction	133
4.1.2	Method	134
4.1.3	Experiments and results	139
4.1.4	Conclusion	144
4.2	Consistent and robust 4D whole-brain segmentation	145
4.2.1	Introduction	145
4.2.2	Method	147
4.2.3	Experiments and results	151
4.2.4	Conclusion	159
5	Automated brain morphometry in patients with Alzheimer’s disease	160
5.1	Materials	161
5.2	Methods	163
5.3	Features and classification	163
5.4	Correction of nuisance factors	165
5.5	Results	166
5.5.1	Cross-sectional analysis	166
5.5.2	Longitudinal analysis	171
5.6	Discussion	176
5.7	Conclusion	178

6	Automated brain morphometry in patients with traumatic brain injury	180
6.1	Materials	181
6.1.1	Clinical information	182
6.1.2	Age-matching of patient groups	183
6.2	Methods	184
6.3	Features and classification	185
6.4	Results	186
6.4.1	Cross-sectional analysis	186
6.4.2	Longitudinal analysis	189
6.5	Discussion	191
6.6	Conclusion	196
7	Patch-based Evaluation of Image Segmentation	197
7.1	Introduction	198
7.1.1	Contribution and overview	199
7.2	Method	201
7.2.1	Framework: The ratio model (RM)	201
7.2.2	Patch-based ratio model (PBRM)	202
7.2.3	Optimisation of local patch correspondences	203
7.2.4	Patch-based Evaluation of Image Segmentation (PEIS)	205
7.2.5	Quantification of segmentation bias	206
7.3	Experiments and results	208
7.3.1	Materials	208
7.3.2	Results on synthetic data	209
7.3.3	Results on medical imaging data	211
7.4	Discussion	213
7.5	Conclusion and future work	214
8	Conclusion	215
8.1	Limitations	216
8.2	Future work	218

Bibliography	220
Appendix	255
A Publications	255
A.1 Journal publications	255
A.2 Conference publications	256
A.3 Grand challenges	258
A.4 Workshop publications	258
A.5 Conference abstracts	259
B ROIs defined in the NMM atlas	260
C The Marshall Classification System	262
D The Glasgow Outcome Scale	262
E ADNI classification results	263
E.1 Cross-sectional analysis at baseline (AD vs. HC, uncorrected volumes) . .	263
E.2 Cross-sectional analysis at baseline (AD vs. HC, corrected volumes) . . .	264
E.3 Cross-sectional analysis at baseline (pMCI vs. sMCI, corrected volumes) .	265
E.4 Longitudinal analysis (baseline→m12, AD vs. HC)	266
E.5 Longitudinal analysis (baseline→m12, pMCI vs. sMCI)	267
E.6 Longitudinal analysis (baseline→m24, AD vs. HC)	268
E.7 Longitudinal analysis (baseline→m24, pMCI vs. sMCI)	269
F TBI classification results	270
F.1 Cross-sectional analysis at acute stage	270
F.2 Longitudinal analysis (baseline→follow-up)	271

List of Tables

1.1	Overview of selected 3D imaging modalities with a brief description of their respective mechanism, advantages and disadvantages. Table adapted from Shenton et al. (2012).	39
2.1	Overview over possible measures to quantify similarity between two images I and J : SSD, MI, NMI and CC. Note that SSD is a measure of dissimilarity. MI and NMI require the calculation of individual histograms (h_I, h_J) and the joint histogram ($h_{I,J}$) of the intensity images. The measures can then be calculated based on the joint, $H(\mathbf{I}, \mathbf{J})$, and the marginal entropies, $H(\mathbf{I})$ and $H(\mathbf{J})$. B denotes the number of bins of each histogram dimension which are indexed by b_i and b_j .	62
2.2	Overview of key characteristics of the five selected brain atlases.	79
2.3	Overview of relevant NMM subsets with respective age and gender information.	83
3.1	Overview of the available MR images with patient gender, patient age, scan time relative to injury, and GCS.	102
3.2	Clinical variables of the 61 acute-phase TBI images (MCS, GOS) and the 64 follow-up MRIs (GOS). n/a: not available.	102
3.3	Overview over all compared methods and their respective building blocks.	114
3.4	Similarity indices (SI) in % averaged (unweighted) over 20 subjects for all 36 non-cortical classes, all 98 cortical classes and all 134 classes. The methods that were compared are MAPER using majority vote, MALP-LWF using locally weighted fusion, MALP-JLF using joint label fusion, MALP-EM $_{\Gamma_i}^{\text{LWF}}$ /MALP-EM $_{\Gamma_i}$ with a local merging strategy and optional segmentation bias correction (BC), and PICSL-BC. *,** = significantly (*: $p < 10^{-2}$; **: $p < 10^{-4}$) different to the method in the column to the left. bold = significantly best results.	115

3.5	SIs in % averaged over 20 test images. ICC (two-way random single measures, absolute agreement, Shrout and Fleiss (1979)) and relative volume differences (Δ_{vol}) based on 5 subjects with available repeat scans for all 36 considered non-cortical structures. Significantly different SI compared to column to the left indicated at $p < 10^{-2}$ (*) and $p < 10^{-4}$ (**).	117
3.6	Mean (SD) of the expert scores for the assessed segmentation quality of hippocampus, thalamus, putamen and occipital pole. Significant improvement is indicated (*: $p < 0.05$; **: $p < 10^{-4}$). ICC (two-way random single measures, absolute agreement) between all available raters.	119
3.7	Classification results in % obtained separating $MCS < 4$ vs. $MCS \geq 4$ and $GOS < 4$ vs. $GOS \geq 4$ based on absolute asymmetry indices of either acute-phase (MCS and GOS) or follow-up (GOS) MR images only. Results shown are averaged over 1000 cross validation runs. bold = best.	124
3.8	Classification results in % (10-fold cross-validation, 1000 runs) obtained separating $MCS < 4$ vs. $MCS \geq 4$ based on AAI of acute-phase MR images. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.01$).	126
3.9	Classification results in % (10-fold cross-validation, 1000 runs) obtained separating $GOS < 4$ vs. $GOS \geq 4$ based on AAI of acute-phase MR images. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.01$).	126
3.10	Classification results in % (10-fold cross-validation, 1000 runs) obtained separating $GOS < 4$ vs. $GOS \geq 4$ based on AAI of follow-up MR images. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.01$).	127
4.1	Compared methods.	139
4.2	Summary of parameter setting.	141
4.3	Classification results for separating AD groups obtained with 4D-GC, DRC-HBSI, 4D-EM, $BSI^{\Omega_{BSI}}$, $PBSI_1^{\Omega_{PBSI}}$, $PBSI_\gamma^{\Omega_{BSI}}$ and $PBSI_\gamma^{\Omega_{PBSI}}$. 4D-EM was also evaluated without DBC. Results shown for the left hippocampus in % calculated by averaging 100 runs of a 6-fold cross validation.	142

4.4	Mean volume change of the left hippocampus in % (top) and sample sizes for atrophy measurement on the MCI and AD group (bottom). Compared methods are 4D-GC, DRC-HBSI, 4D-EM, $BSI^{\Omega_{BSI}}$, $PBSI_1^{\Omega_{PBSI}}$, $PBSI_{\gamma}^{\Omega_{BSI}}$ and $PBSI_{\gamma}^{\Omega_{PBSI}}$. 4D-EM was also evaluated without the DBC. Standard deviation in parentheses. Corrected sample sizes were computed on the excess change over normal aging.	143
4.5	Intraclass correlation coefficient (ICC) ($\text{mean} \pm \text{SD}$) based on label volumes averaged over 5 image pairs for all 36 subcortical and all 98 cortical regions obtained manually, with MAPER, and the proposed method without and with the temporal constraints as described in Section 4.2.2.3-4.2.2.4. Proposed method evaluated with unweighted, standard DBC and weighted DBC. Bold = significantly better than without coupling (paired, two-sided Student's t-test, $p < 0.05$).	153
4.6	Average Dice coefficient (Dice, 1945) (DC) values comparing the proposed approach without or with temporal constraints to the gold-standard segmentations of the 10 images of the 5 pairs of repeat scans. Proposed method evaluated with unweighted, standard DBC and weighted DBC. Bold significantly better than without coupling (paired, two-sided Student's t-test, $p < 0.001$).	154
4.7	Mean atrophy rates of the left hippocampus in % (top) and sample sizes for atrophy measurement on the MCI and AD group (bottom). Compared methods are the DRC-HBSI (<i>cf.</i> Section 4.1), MALP-EM and different instances of the proposed method. Standard deviation in parentheses. Corrected sample sizes were computed on the excess change over normal aging.	155
4.8	Mean atrophy rates of the ventricles in % (top) and sample sizes for atrophy measurement on the MCI and AD group (bottom). Compared methods are the VBSI, MALP-EM and different instances of the proposed method.	155
4.9	Mean atrophy rates of the whole brain in % (top) and sample sizes for atrophy measurement on the MCI and AD group (bottom). Compared methods are the KN-BBSI (Leung et al., 2010b), MALP-EM and different instances of the proposed method.	156
4.10	Mean atrophy rates of other structures in % (top) and sample sizes for atrophy measurement on the MCI and AD group (bottom) using MALP-EM4D with affine registration, unweighted DBC and the proposed spatially varying temporal coupling.	156

4.11	Mean atrophy rates of combined structures in % (top) and sample sizes for atrophy measurement on the MCI and AD group (bottom) using MALP-EM4D with affine registration, unweighted DBC and the proposed spatially varying temporal coupling. Structures were combined by averaging individual atrophy rates ($\dagger$) or calculating atrophy rates on merged structures ($\ddagger$). *: includes 3rd and 4th ventricle.	156
5.1	Overview of the processed subjects from the ADNI cohort, including age and clinical information at baseline.	162
5.2	Balanced classification accuracies in % for separating healthy control and AD subjects (effect sizes in parentheses) after correcting for different nuisance factors (100 runs, 6-fold cross-validation, LDA). Highest effect size in bold.	168
5.3	Classification results in % (6-fold cross-validation, LDA 100 runs, RF/SVM 20 runs) obtained separating AD from CN (top) and stable from progressive MCI (bottom). Individual structures are sorted by effect size. The 20 structures with highest effect size are listed explicitly. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses. Features were corrected for age, gender and brain size.	170
5.4	Mean volume change of selected structures in % with corresponding sample sizes for different disease groups. Standard deviation in parentheses. Measurements based on volume change from baseline to 12 months (top table) or 24 months (bottom table) follow-up visit. Corrected sample sizes were computed on the excess change over normal aging.	173
5.5	Classification results in % (6-fold cross-validation, LDA 100 runs, RF/SVM 20 runs) for separating AD from CN based on volume change from baseline to m12 (top) or m24 (bottom). Individual structures are sorted by effect size. The 20 structures with highest effect size are listed explicitly. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses.	174

5.6	Classification results in % (6-fold cross-validation, LDA 100 runs, RF/SVM 20 runs) for separating progressive from stable MCI based on volume change from baseline to m12 (top) or m24 (bottom). Individual structures are sorted by effect size. The 20 structures with highest effect size are listed explicitly. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses.	175
5.7	Overview over selected articles that use features from T1w MR images from the ADNI cohort. Table adapted from Falahati et al. (2014). CTH: cortical thickness, ENR: elastic net regression, HV: hippocampus volume, LLE: locally linear embedding, MBL: manifold-based learning, MIL: multiple instance learning, OPLS: orthogonal partial least square to latent structure, SR: sparse regression, TBM: tensor-based morphometry.	176
6.1	Overview of all processed MR images with patient gender, patient age, scan time relative to injury, GCS, MCS and injury severity.	182
6.2	Overview of the MR images used for the analysis with patient gender, patient age, scan time relative to injury, GCS, MCS and injury severity.	183
6.3	Classification results in % (6-fold cross-validation, LDA/RF/SVM 100 runs) obtained separating TBI patients with a severe disability from patients with low disability outcome based on structural volumes and asymmetry at the acute stage of the injury. Individual structures are sorted by effect size. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses.	188
6.4	Classification results in % (6-fold cross-validation, LDA/RF/SVM 100 runs) obtained separating TBI patients with a severe disability from patients with low disability outcome based on structural volume changes between the acute and chronic disease stage. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses.	192
7.1	Overview over all quantities available through PEIS and their description.	207

7.2	Similarity scores (Dice coefficient, PEIS) and mean segmentation bias (μ_b) and shape conservation (σ_b) averaged over 20 hippocampus segmentations for different methods.	212
B.1	Structure names of non-cortical structures in NMM brain atlas. Original ROI IDs were remapped to be consecutive.	260
B.2	Structure names of cortical structures in NMM brain atlas.	261
C.1	Marshall Classification system based on and modified from Marshall et al. (1991). .	262
D.1	Glasgow Outcome Scale based on and modified from Jennett and Bond (1975); Jennett et al. (1981).	262
E.1	Classification results in % (6-fold CV, LDA 100 runs, RF/SVM 20 runs) for separating AD from CN. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses. Sorted by effect size.	263
E.2	Classification results in % (6-fold CV, LDA 100 runs, RF/SVM 20 runs) for separating AD from CN. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses. Sorted by effect size.	264
E.3	Classification results in % (6-fold CV, LDA 100 runs, RF/SVM 20 runs) for separating progressive from stable MCI. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses. Sorted by effect size.	265
E.4	Classification results in % (6-fold CV, LDA 100 runs, RF/SVM 20 runs) for separating AD from CN. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses. Sorted by effect size.	266
E.5	Classification results in % (6-fold CV, LDA 100 runs, RF/SVM 20 runs) for separating progressive from stable MCI. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses. Sorted by effect size.	267
E.6	Classification results in % (6-fold CV, LDA 100 runs, RF/SVM 20 runs) for separating AD from CN. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses. Sorted by effect size.	268

E.7	Classification results in % (6-fold CV, LDA 100 runs, RF/SVM 20 runs) for separating progressive from stable MCI. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses. Sorted by effect size.	269
F.1	Classification results in % (6-fold cross-validation, LDA/RF/SVM 100 runs) obtained separating TBI patients with a severe disability from patients with moderate disability outcome based on structural volumes at the acute stage of the injury (top). Classification of patients with moderate and low disability outcome (bottom). Individual structures are sorted by effect size. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses.	270
F.2	Classification results in % (6-fold cross-validation, LDA/RF/SVM 100 runs) obtained separating TBI patients with a severe disability from patients with moderate disability outcome based on structural volume changes between the acute and chronic disease stage (top). Classification of patients with moderate disability and low disability outcome (bottom). The individual structures are sorted by effect size. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses.	271

List of Figures

- 1.1 Brain images acquired with different imaging sequences/modalities from a patient with traumatic brain injury. The good tissue contrast in T1-weighted MR images and the pronounced contusions in the FLAIR sequence are apparent. Furthermore, GRE and PD weighted images are shown. Diffusion derived FA and MD maps are also shown. CT is well suited to image bone injuries, oedema or intracranial bleeding. 38
- 1.2 Example of T1w MR image (left), FDG-PET image (middle) and AV45-PET image (right) in coronal view for a healthy control (top, male, 66 years) and patient with AD (bottom, male, 75 years). The PET images are superimposed on the structural MR image with red indicating high signal and blue low signal respectively. The PET images show reduced metabolic activity (FDG) and increased A β binding (AV45-PET) in patients with AD. 41
- 1.3 Hypothetical trajectories of biomarkers during the progression of AD. Figure adapted from Jack Jr et al. (2010a); Aisen et al. (2010). 44
- 1.4 Three examples of MR images (brain extracted) of subjects from the ADNI1 cohort visualised in a coronal view. A healthy control subject (male, 84.8 years at baseline) is shown in the top row. A MCI patient (female, 71.8 years at baseline) that converted to AD after three years is shown in the middle row. An AD patient (male, 77.5 years at baseline) is shown in the bottom row. Left: baseline MR image, Middle: 2 year follow-up MR image, Right: difference image of rigidly aligned images. The differences are visually very subtle but the increased atrophy in the medial temporal lobe (blue arrows) and the enlarged ventricles (red arrows) are apparent in the difference image. 45

- 1.5 Four examples of MR images (brain extracted) of subjects from a prospective TBI cohort visualised in coronal or axial view. Top row: patient with mild TBI (male, 72 years, GCS: 14, GOSe: 8, MCS: 1, fall accident). Second row: patient with moderate TBI (female, 55 years, GCS: 2, GOSe: 4, MCS: 4, fall accident). Third row: moderate TBI patient (male, 38 years, GCS: 11, GOSe: 5, MCS: n/a, car accident). Bottom row: patient with severe TBI (female, 33 years, GCS: 4, GOSe: 3, MCS: 2, car accident). Left: baseline MR image acquired in the acute phase (days after the injury), Middle: follow-up MR image acquired in the chronic phase (months after the injury), Right: difference image of rigidly aligned images. Enlarged ventricles (red arrows), a subdural haematoma (blue arrow) and deformed/compressed frontal region (yellow arrow) are indicated in the difference images. 51
- 1.6 Top row: Example of a healthy control subject (male, 65.1 years) from the ADNI1 cohort visualised in coronal, axial and sagittal view. Middle row: Three examples of MR images of subjects with AD from the ADNI1 cohort (from left to right: female, 84.5 years in coronal; male, 80.2 years in axial; male, 77.5 years in sagittal view). Bottom row: Three examples of MR images of subjects with TBI from a retrospective TBI cohort (from left to right: male, 23 years, GCS=5, GOS=1, MCS=5; male, 45 years, GCS=7, GOS=3, MCS=5; male 42 years, GCS=14, GOS=4, MCS=5). Indicated with arrows: enlarged ventricles (red), pronounced hippocampal/medial temporal lobe atrophy (blue), skull fracture/deformation (green), tissue resection (white), subdural haematoma (orange). 54
- 2.1 A brain atlas $\mathbf{A}_m$ consists of an intensity image $\mathbf{I}_m^A$ with corresponding segmentation $\mathbf{L}_m^A$. The reference segmentation is ideally created manually by an expert who follows a detailed segmentation protocol. This process is very time consuming and not scalable. Therefore it can only be performed on a small number of images. . . . 57
- 2.2 Schematic illustration of a typical segmentation pipeline with the common building blocks: brain extraction, atlas alignment, label fusion and segmentation refinement. The unsegmented target image and the estimated segmentation are outlined in blue, the atlas images are outlined in red and the propagated atlas labels in orange respectively. 58

2.3	Schematic process of registering a source image $\mathbf{I}_{\text{src}}$ to a target image $\mathbf{I}_{\text{tgt}}$ with calculated transformation ϕ . The difference image of $\mathbf{I}_{\text{src}}$, $\phi(\mathbf{I}_{\text{src}})$, and $\mathbf{I}_{\text{tgt}}$ is shown before and after registration respectively. The target image is outlined in blue, the source image is outlined in red (before) and orange (after registration).	60
2.4	Left: Schematic process of fusing K atlas labels that have been aligned to the target image into a consensus segmentation. Right: Illustrations of the increasing segmentation accuracy with increasing number of fused atlas labels. The model described in Equation 2.5 is shown for $\epsilon_{\text{sys}} = 0.144$, $\epsilon_{\text{rand}} = 0.1$ (red) with corresponding asymptotic limit (blue). A qualitative plot indicating the increased accuracy when atlas selection is used is also shown (green). The diagram is based on and adapted from Heckemann et al. (2006); Aljabar et al. (2009).	66
2.5	Illustration of the weight calculation when segmenting a voxel i . The weights $w_m^{\square}(i, j)$ are determined between the fixed reference patch $P(\mathbf{I}_{\text{tgt}}, i)$ (yellow) and the atlas patches $P(\mathbf{I}_{\mathbf{m}}^{\mathbf{A}}, j)$ (red). For each atlas image several atlas patches that are shifted within a local search neighbourhood (cyan) are considered. With a weighted voting approach a probabilistic segmentation (<i>cf.</i> Equation 2.8) is obtained. Illustration adapted and modified from Coupé et al. (2011); Rousseau et al. (2011). . . .	70
2.6	Schematic process of training classifiers for the detection and correction of systematic segmentation bias. This illustration corresponds to the variant named <i>explicit bias correction</i> in Wang et al. (2011a). Note that for binary segmentation tasks the error detection and error correction classifiers are identical. Illustration adapted and modified from Wang et al. (2011a).	76
2.7	The AAL atlas in axial (left), coronal (middle) and sagittal (right) view plane. T1w MR template with overlaid segmentation contours in a colour scheme that provides a good contrast between adjacent labels. The cross-hair in the axial view indicates the position of the illustrated coronal and sagittal slice.	80
2.8	An example of a subject (IBSR01, male, 37 years) of the IBSR atlas in axial (left), coronal (middle) and sagittal (right) view plane. T1w MR template with overlaid segmentation contours.	80

2.9	An example of a subject (a01) of the Hammers atlas in axial (left), coronal (middle) and sagittal (right) view plane. T1w MR template with overlaid segmentation contours.	82
2.10	An example of a subject (S01) of the LPBA atlas in axial (left), coronal (middle) and sagittal (right) view plane. T1w MR template with overlaid segmentation contours.	82
2.11	Two examples of subjects of the NMM atlas in axial (left), coronal (middle) and sagittal (right) view plane. T1w MR images of a male subject aged 20 years (OAS1_0285, top) and a female subject aged 90 years (OAS1_0083, bottom) with overlaid segmentation contours.	84
3.1	Example of segmentation results obtained on a subject with highly altered brain configuration. Segmentations calculated with MAPER using majority voting (left, Heckemann et al. (2010)) and SyN (Avants et al., 2008) from the ANTs package using either majority voting (middle) or the joint label fusion (right, Wang et al. (2013b)). Red arrows: substantial oversegmentation of the hippocampus; yellow arrows: inaccurate cortex segmentation due to gross brain deformation; blue arrows: ventricles incorrectly labelled as background; white arrows: region of missing tissue prohibits reasonable one-to-one mapping of the atlases.	98
3.2	Schematic process of the calculation of the subject specific spatial priors Π for an unsegmented target image $\mathbf{I}_{tgt}$. After brain extraction and bias correction, the available M atlases are registered to the space of $\mathbf{I}_{tgt}$. Using these transformations, label maps and corresponding T_1 -weighted MR images are mapped to the space of $\mathbf{I}_{tgt}$. The label maps are then averaged into probabilistic priors for the individual structures using the joint label fusion (Wang et al., 2013b). A subset of the 134 probabilistic labels is shown in green.	105

3.3	Schematic process of the segmentation refinement using prior relaxation, EM-optimisation and spatially weighted combination of probabilistic label maps on the example of the hippocampal region. If registration consistently fails joint label fusion tends to label a significant number of voxels belonging to the inferior lateral ventricle as hippocampus. These wrongly labelled low-intensity voxels lead to a high variance of the estimated intensity distribution within the hippocampus label (top left). The red interval (top left intensity distribution) indicates for which voxels prior relaxation will be carried out. EM-refinement then allows correction of the mislabeled CSF voxels leading to a sharper intensity distribution within the hippocampus (top right). The segmentations obtained using label fusion and EM-optimisation are finally merged into a consensus segmentation (bottom right). This combination is based on spatially varying weights that are calculated based on the overlap of intra-label intensity distributions (bottom left).	107
3.4	Illustration of the relaxation weights (left column) and the spatially varying combination weights as $1 - \Gamma_i$ (right column) for the two example subjects in Figure 3.8 (top row, subject with pathological changes) and Figure 3.10 (bottom row, subject with a close-to-normal brain configuration). It is noted that in cortical regions and regions where the label fusion fails, both increased prior relaxation (white arrows) and combination weights (red arrows) favouring EM-based segmentations are apparent.	113
3.5	MALP-EM $_{\Gamma}$ applied to the 15 training images with varying Γ . Here $\Gamma = 1.0$ is equivalent to exclusive joint label fusion and i denotes the spatially varying choice of Γ_i . Median similarity index with first and third quantiles of 36 non-cortical structures (left), 98 cortical structures (middle) and all 134 regions (right).	116
3.6	Distribution of expert scores for investigated structures. Note that structures in the left/right brain hemisphere were rated separately. This results in 78 ratings per structure ($3 \text{ raters} \times 13 \text{ subjects} \times 2 \text{ hemispheres}$). Occipital pole was not rated by rater C. Comparison of MALP-JLF (left bars) and MALP-EM (right bars).	120
3.7	Fraction of test images on which a method performs better than the other. Raters B and C were asked to rate which segmentation was better, even if scores were tied. .	120

3.8	Segmentations obtained with MALP-JLF (top) and MALP-EM (bottom) of a subject with a markedly changed brain configuration on MRI (male, age: 45 years, GOS = 3, follow-up MRI, taken one month after injury, axial(left)/coronal(right) view). While the label fusion clearly fails to match various intensity boundaries, for example in the hippocampal region, MALP-EM is able to correct this problem to a large extent due to the strong intensity contrast of CSF and grey matter structures. Red arrows highlight improvements obtained using MALP-EM over pure joint label fusion. White arrows highlight errors due to inconsistent brain extraction.	121
3.9	Axial slices of the two subjects for which the brain extraction was not successful. left: male, age: 56, GOS = 3, MCS = 5, follow-up MR image taken two months after injury, right: male, age: 16, GOS = 3, MCS = 5, follow-up MR image taken seven months after injury.	122
3.10	Typical segmentation result obtained using MALP-EM _{Γ_i} of a subject with a close-to-normal brain configuration in axial(left)/coronal(right) view (female, age: 17 years, GOS = 5, MCS = 2, follow-up MRI, taken nine months after injury).	123
3.11	Receiver operating characteristic (ROC) curves for classifying subjects according to MCS using the sum of cortical AAI (left), and according to GOS at baseline (middle) and follow-up time point (right) using the accumulated non-cortical AAI.	125
3.12	Relevance of individual brain structures for GOS group separation. Sorted histogram of how often a structure's absolute asymmetry index (AAI) was one of the 10 most significant indices in the 10000 (1000 rounds of 10-fold cross-validation) runs. The 10 structures that were picked most often based on acute-phase (left) or follow-up (right) MR images.	128
4.1	Construction of the ROI Ω_{PBSI} which is also shown in Figure 4.2 with parameters $\eta = 0.95$, $n_e = 0$ and $\zeta = 0.90$, $n_d = 1$. Probabilistic hippocampus segmentations (green) and binary masks (red) are overlaid over the respective subject MR image.	138
4.2	Binary ROIs (red) for the calculation of the respective integrals. ROI calculation based on hard labels (Ω_{BSI} , left) [$n_e^{\text{bin}} = 1$ and $n_d^{\text{bin}} = 1$] and based on the proposed method with probabilistic labels (Ω_{PBSI} , right) [$\eta = 0.95$, $n_e = 0$ and $\zeta = 0.90$, $n_d = 1$].	138

4.3	Example coupling weights γ_i dependent on local patch similarity w_i for different choices of parameter b . In this work, b is automatically determined based on the image intensities as $b = 5.0/\text{median}[w_i^{t,t+1}]$. The figure shows typical results for intrasubject repeat scans (red), follow-up images of healthy subjects, or of patients with AD or mild TBI (green, blue). For highly pathological cases such as the one illustrated in Figure 4.5 rather high values for b are determined (purple). Note: The low values of w_i are due to the fact that the weights are calculated on the log-transformed image intensities.	150
4.4	Simulated atrophy calculation for fixed baseline volume (v_{t^1}) and varying follow-up volume (v_{t^2}). Calculation based on $\Delta_v^{\text{frac}}(t^1, t^2) = 100\%(v_{t^2} - v_{t^1})/v_{t^1}$ (red) and the logarithmic transform $\Delta_v^{\text{log}}(t^1, t^2) = 100\% \ln(v_{t^2}/v_{t^1})$ (green) with corresponding approximation error (blue).	152
4.5	Visual results for a TBI subject (TBI066, GOSe = 3) with significant pathologies obtained with the proposed method. (a,d) T_1 -weighted baseline, (b,e) 7.9 month follow-up scan with overlaid segmentations and (c,f) corresponding differential bias fields. Results were calculated with unweighted (a,b) and with weighted (d,e) differential bias field correction. (c) differential bias field estimated as presented in Lewis and Fox (2004), (f) proposed polynomial fitted bias field. (g) corresponding temporal coupling weights according to Equation 4.10 (high weights in red). Boxplots of volume change in the thalamus and caudate calculated with (black) / without (blue) the proposed temporal coupling as described in Section 4.2.2.3-4.2.2.4 for GOSe groups 3-5 vs. 6-8.	158
5.1	Top: Dependence of hippocampal volume on age (left), gender (middle) and brain volume (right). Bottom: Corresponding volumes corrected for nuisance factors age, gender and brain size. Overlaid regression lines for distinct disease groups with corresponding regression lines.	165
5.2	Example cross-sectional segmentation results of baseline images of a healthy control subject (top, ADNI_005_S_0553, male, 85 years of age) and a patient diagnosed AD (bottom, ADNI_018_S_0286, male, 66 years of age) in axial (left), coronal (middle) and sagittal (right) view-plane.	166

5.3	Boxplots of structural volumes at baseline for six selected structures before (top two rows) and after (bottom two rows) correcting for nuisance factors for distinct disease groups. Note that the corrected volumes need to read with respect to a healthy population of matched age, gender and brain size. Structures were selected based on their performance in classifying the investigated disease groups (<i>cf.</i> Table 5.3).	167
5.4	Density plot of the hippocampal volumes before (left) and after (right) correction for age, gender and brain size. Note that the corrected volumes need to be interpreted with respect to the healthy reference group and have therefore zero-mean for HCs. .	169
5.5	Example longitudinal segmentation results of baseline (left) and month 24 (middle) follow-up images of a patient diagnosed AD (ADNI_018_S_0286) in axial view-plane. Substantial hippocampal atrophy (measured: -7.81%) and ventricular enlargement (16.5%) is apparent in the difference image after affine registration (right).	171
5.6	Boxplots of volume changes for selected brain structures (top) and surrogate structures (bottom) from baseline to month 24 follow-up image for different disease groups. Features selected based on their performance in classifying the investigated disease groups (<i>cf.</i> Table 5.5 and Table 5.6).	172
6.1	Boxplots of age, GCS, MCS and injury severity of patients with low, moderate or severe disability outcome. Shown are only subjects of the age-matched dataset. . .	183
6.2	Example cross-sectional segmentation results of images at the acute stage of a TBI. Top: TBI_010_2_8, male, 21 years of age, mild TBI, GCS: 15, GOSe: 8, MCS: 1, fall accident, image acquired 2 days after injury. Bottom: TBI038, female, 47 years of age, mild TBI, GCS: 15, GOSe: 4, MCS: 5, fall, image acquired 4 days after injury with clear sequelae of intra-cerebral haematoma.	184
6.3	Boxplots of selected structural volumes (first row), surrogate structures (second row) and asymmetry indices (third row) with respect to the investigated disease outcome groups. Features selected based on their performance in classifying severe disability vs. low disability outcome (<i>cf.</i> Table 6.3).	187

6.4	Example longitudinal segmentation results of images at the acute and chronic stage of a TBI. Top: TBI061, male, 69 years of age, GCS: 14, GOSe: 8, MCS: 1, fall accident, acute/chronic image acquired 8/265 days after injury. Bottom: TBI142, male, 51 years of age, GCS: 3, GOSe: 6, MCS: 5, transport accident, acute/chronic image acquired 2/383 days after injury, diffuse axonal injury. The difference image of subject TBI142 illustrates the clear ventricular enlargement (measured: 47%), hippocampal atrophy (-6.2%) and reduction of brain stem volume (-13%).	190
6.5	Boxplots of change rates of selected ROIs with respect to the investigated disease outcome groups. Features selected based on their performance in classifying severe disability vs. low disability outcome (<i>cf.</i> Table 6.4).	191
6.6	Example longitudinal segmentation results of images at the acute and chronic stage of a TBI. Top row, TBI150: male, 68 years of age, GCS: 15, GOSe: 5, MCS: 5, unclear accident, acute/chronic image acquired 14/264 days after injury, craniotomy and evacuation of an acute SDH, intracranial haematoma induced brain destruction. Middle row, TBI047: male, 62 years of age, GCS: 13, GOSe: 3, MCS: 5, fall accident, acute/chronic image acquired 29/306 days after injury, craniotomy and evacuation of an acute subdural haemorrhage, broad gliosis on left temporal lobe. Bottom row, TBI114, male, 69 years of age, GCS: 14, GOSe: 5, MCS: 5, cycling accident, acute/chronic image acquired 51/233 days after injury, no craniotomy, traumatic subarachnoid haemorrhage, intracranial haematoma with surrounding oedema in left temporal lobe, central and cortical atrophy.	193
7.1	The <i>spatial</i> overlaps $\mathcal{A}$ between the <i>supports</i> of a reference patch (grey) and a moving test patch (dark red) is shown in yellow. For increasing patch displacements $\Delta = (\Delta_x, \Delta_y)$ the spatial overlap decreases. If two displacements have the same L1-norm (both examples shown on the right with $\ (1, -1)\ _1 = \ (2, 0)\ _1 = 2$) the proposed method favours a diagonal displacement as $\mathcal{A}(1, -1) > \mathcal{A}(2, 0)$. The reference centre is indicated as circle. The centres of the test patches are indicated as crosses and correspond to the displacements shown in Figure 7.2.	203

7.2	Calculation of optimal patch centers i'_k at different distance-levels k . An example of a reference patch is shown in grey, while the best matching patches at the three corresponding distance-levels ($k = 0, 1, 2$) are shown in dark red.	204
7.3	Example reference patches of varying complexity. Foreground segmentation is shown in grey and corresponding facets in blue. The patch shown on the right has maximal complexity of $n_{\text{facets}}^{\max} = 4 \times (p_w - 1) = 16$ for a 2D patch with width $p_w = 5$	206
7.4	left: Example comparison of a polygon (test segmentation, outlined in dark red) to a circle (reference, outlined in black). Patch-correspondences of fixed reference patches (green) and moving test patches (orange). Vectors represent patch-normals $\mathbf{n}$ (dark blue), optimal patch displacements Δ (light blue) and the projections of Δ on $\mathbf{n}$ (red) which represents the estimated segmentation bias. The distance transform of the reference segmentation is shown in a rainbow colour scheme. right: Estimated voxel-wise segmentation bias $b(i) = \mathbf{n}(i) \cdot \Delta(i)$ within the domain Ω	206
7.5	Overview over employed datasets. Synthetic datasets with reference segmentation outline (black) and example test segmentation outlines (dark red). Medical imaging datasets with overlaid ground truth reference segmentation. Cardiac MR image with outlined right ventricle myocardium (yellow) and blood pool (red), and a MR brain scan with overlaid hippocampus segmentation (yellow).	208
7.6	<i>Circles</i> experiments for reference radiuses $r_R = 15$ (left) and $r_R = 80$ (right): Different similarity scores (colours) and segmentation bias μ_b (black, σ_b as error bars).	210
7.7	<i>Polygons</i> experiment: Different similarity scores (left) and estimated shape conservation, σ_b (right).	210
7.8	<i>Crosses</i> experiment: Estimated translational bias for all coordinate directions ($\sigma_{x,y,z}$ as error bars). Crosses were translated by $(c_x = 10\cos(t), c_y = 10\sin(t), c_z = 1)$ with $t \in [0, 2\pi]$	211
7.9	Similarity scores calculated on <i>cardiac data</i> segmented into myocardium and blood pool. Both the Dice coefficient and the ASD do not yield <i>comparable scores</i> for structures of different complexities.	212

7.10	Visualisation in coronal view of segmentation bias for the right hippocampus of three random subjects (rows) for five automatic segmentation methods (columns). Colours relate to positive bias or over-segmentation (red) and negative bias or under-segmentation (blue).	213
------	--	-----

Glossary

2D two-dimensional. 27, 78, 206

3D three-dimensional. 37, 69, 94, 211

4D four-dimensional. 132–136, 140, 146–148, 150, 154

A β amyloid beta. 18, 40–43, 46, 48

AAI absolute asymmetry index (Galaburda et al., 1987; Bonilha et al., 2013). 12, 23, 101, 124–127, 185, 188, 194

AAL Automated Anatomical Labeling. 78

ACC accuracy. 141, 142

AD Alzheimer’s disease. 14, 18, 24, 25, 36, 40–44, 46–48, 52, 53, 55, 56, 59, 63, 71, 85–90, 95, 97, 99, 132–134, 139, 142, 143, 145, 159–162, 166, 168, 169, 171–173, 176–180, 209, 215, 218

ADNI Alzheimer’s Disease Neuroimaging Initiative (<http://adni.loni.ucla.edu>, last accessed 29 June 2015). 15, 55, 87, 88, 132–134, 139, 145, 147, 151, 154, 160, 161, 163, 176–178, 184, 186, 195, 215–218

ADFDA Alzheimers Disease and Related Disorders Association. 42, 43

ANTs Advanced Normalization Tools. 21, 98, 114

ApoE apolipoprotein E. 41, 160

ASD average surface distance. 27, 200, 209, 211, 212, 214

bACC balanced accuracy (Brodersen et al., 2010). 124, 169, 171, 173, 176, 177, 189–191

BBSI Brain Boundary Shift Integral. 145

BET Brain Extraction Tool. 58

BOLD blood-oxygen-level-dependent. 40

BSE Brain Surface Extractor. 58

BSI Boundary Shift Integral. 132–136, 139–141, 144, 147, 151, 152, 154

CADDementia Computer-Aided Diagnosis of Dementia (<http://caddementia.grand-challenge.org>, last accessed 29 June 2015). 90

CC cross-correlation. 11, 61, 62, 67, 68

CDR Clinical Dementia Rating. 89

CSF cerebrospinal fluid. 43, 46, 106, 111, 128, 140, 160, 164, 185, 192

CT computed tomography. 18, 37, 38, 49, 50, 52, 67, 101, 186

CV cross-validation. 164, 186

DAI diffuse axonal injury. 48, 50, 52, 94, 195

DBC differential bias correction. 12–14, 132, 139, 142, 143, 145–147, 149–151, 153, 154, 156

DBM deformation-based morphometry. 61, 86

DC Dice coefficient (Dice, 1945). 13, 153, 154, 198

DOF degree of freedom. 62, 148

DTI diffusion tensor imaging. 38, 52, 93, 146

DWI diffusion weighted imaging. 38, 50, 91, 94, 97, 160, 218, 219

EM expectation-maximisation. 72, 74, 96, 99, 100, 103, 105, 107, 110, 111, 114, 116, 118, 129, 133–136, 139, 140, 145–150, 154

FA fractional anisotropy. 18, 38, 94

FDG 18-fluorodeoxyglucose. 18, 40, 41

FFD free form deformation. 103, 147

FLAIR fluid-attenuated inversion-recovery. 18, 37, 38, 50, 219

fMRI functional magnetic resonance imaging. 40, 50, 52

FreeSurfer FreeSurfer (<http://surfer.nmr.mgh.harvard.edu/>, last accessed 29 June 2015, Fischl and Dale (2000); Fischl et al. (2002, 2004)). 78, 87, 93, 146

FTD frontotemporal dementia. 85, 86, 90, 218

GC graph-cuts. 63, 74, 89, 99, 132, 134, 140, 143, 146

GCS Glasgow Coma Score. 11, 15, 25, 26, 48, 49, 101, 102, 182–185, 190, 193, 195

GM grey matter. 37, 78, 81, 83, 84, 86, 89, 92, 93, 128, 140, 155, 185, 189–192

GMM Gaussian mixture model. 99, 148

GOS Glasgow Outcome Score. 12, 49, 55, 97, 101, 123–127, 130, 182, 185

GOSe extended Glasgow Outcome Score. 25, 26, 157, 182–184, 186, 190, 193, 218

GRE gradient echo. 18, 38

GSAD Gaussian weighted sum of absolute differences. 149, 152

GSSD Gaussian weighted sum of squared differences. 104, 112, 115, 148, 152

GWF globally weighted fusion. 68

HC healthy control. 25, 71, 86–90, 93, 94, 139, 161, 162, 168, 169, 171–173, 176–178

IBSR Internet Brain Segmentation Repository. 80, 81

ICBM International Consortium for Brain Mapping. 77

ICC intraclass correlation coefficient. 12, 13, 116, 117, 119, 153

ICV intracranial volume. 166

IRTK Image Registration Toolkit (<https://github.com/BioMedIA/IRTK>, last accessed: 8 December 2014). 119

JLF joint label fusion. 71, 75, 96, 99, 100, 103, 104, 114, 119

LBD Lewy body dementia. 85, 90, 218

LDA linear discriminant analysis. 14, 89, 124, 141, 157, 164, 168, 176, 186

LOC loss of consciousness. 49

LPBA LONI Probabilistic Brain Atlas. 82, 83

LWF locally weighted fusion. 68, 69, 104, 112, 114, 115

m12 month-12. 14, 15, 160, 162, 163, 171–176, 178, 179, 218

m24 month-24. 14, 15, 160, 162, 163, 171–176, 179, 218

MALP-EM Multi-Atlas Label Propagation with Expectation-Maximisation based refinement. 53, 96, 97, 100, 103, 112, 119, 122–125, 127–132, 135, 148, 163, 176, 178–180, 184, 185, 215, 216, 218, 219

MALP-EM4D Multi-Atlas Label Propagation with Expectation-Maximisation based refinement in 4D. 132, 147, 154, 157, 159, 163, 171, 176–180, 184, 185, 215, 216, 218

MAP maximum a posteriori. 73

MAPER Multi-Atlas Propagation with Enhanced Registration (Heckemann et al., 2010, 2011). 63, 96, 97, 99, 100, 103, 112, 163

MCI mild cognitive impairment. 14, 15, 44, 46, 85–90, 133, 139, 143, 144, 160–162, 170, 172, 173, 175–177, 179

MCS Marshall Classification Score. 12, 15, 23, 25, 26, 49, 97, 101, 118, 119, 123–126, 182–186, 190, 193

MD mean diffusivity. 18, 38, 94

MI mutual information (Collignon et al., 1995; Viola and Wells III, 1997). 11, 61, 62, 67, 68

MIL multiple instance learning. 88, 90

MMSE mini-mental state examination. 42

MNI Montreal Neurological Institute (<http://www.mni.mcgill.ca/>, last accessed 29 June 2015). 77, 81, 82

MR magnetic resonance. 12, 15, 18, 36–38, 41, 44, 46, 50, 52, 53, 56, 63, 65, 77, 78, 80–84, 90–92, 94–98, 100–105, 111, 112, 116, 118, 121, 123, 126, 127, 130–134, 140, 145–147, 151, 154, 159–161, 176, 179–182, 186, 192, 194, 196, 197, 200, 215, 219

MRF Markov random field. 73, 74, 108, 109, 135, 140, 145–149, 154

MRI magnetic resonance imaging. 23, 37, 38, 43, 44, 46, 50, 52, 53, 55, 56, 59, 67, 68, 81, 85, 86, 88–90, 93, 101, 121, 124, 130, 160, 180, 191

mTBI mild TBI. 50

MTL medial temporal lobe. 43, 85, 86, 90, 169, 180, 215

MVF majority vote fusion. 65–68, 112

NINCDS National Institute of Neurological and Communicative Disorders. 42, 43

NMI normalised mutual information (Studholme et al., 1999). 11, 61, 62, 67–69

NMM Neuromorphometrics. 83, 102, 160, 163, 184, 209, 218

OASIS Open Access Series of Imaging Studies. 83

PBSI probabilistic Boundary Shift Integral. 133, 136, 140, 141, 144

PCA principal component analysis. 87

PD proton density. 18, 38

PEIS Patch-based Evaluation of Image Segmentation. 55, 197, 200, 201, 208–214, 216

PET positron emission tomography. 18, 40, 41, 43, 46, 91, 93, 160

pMCI progressive MCI. 46, 86, 88, 90, 139, 143, 162, 168, 169, 171–173, 176, 177, 217

PTA post-traumatic amnesia. 49, 182

RF random forest. 89, 164, 169, 171, 173, 176–178, 186, 191, 219

RM ratio model. 201

ROC receiver operating characteristic. 23, 125, 194

ROI region of interest. 26, 50, 52, 53, 56, 68, 71, 78, 81–84, 86, 88, 91, 93, 94, 97, 133, 136, 137, 139–141, 144, 166, 171, 176, 180, 186, 189–191, 194, 196, 215–218

SD standard deviation. 12, 119

SDH subdural haematoma. 26, 59, 193

SENS sensitivity. 87, 88, 141, 142, 159, 169, 171, 173, 176, 177, 189, 190, 194

SI similarity index. 12, 66, 114, 117, 153

SMC subjective memory complaints. 161

sMCI stable MCI. 46, 86, 88, 90, 139, 143, 161, 162, 168, 169, 171–173, 176, 177, 217

SNR signal to noise ratio. 78

SPEC specificity. 87, 88, 141, 142, 159, 169, 171, 173, 176, 177, 189, 190, 194

SPHARM spherical harmonics. 87

SPM Statistical Parametric Mapping (<http://www.fil.ion.ucl.ac.uk/spm/>, last accessed 29 June 2015, Ashburner and Friston (1997, 1999, 2005)). 81, 82

SSD sum of squared differences. 11, 61, 62, 67–69, 104

STAPLE Simultaneous Truth And Performance Level Evaluation (Warfield et al., 2004). 72, 73, 198

SVM support vector machine. 86, 89, 164, 169, 176, 178, 186, 189

SWI susceptibility weighted imaging. 37, 52, 219

T Tesla. 37, 81–83, 168, 177, 178, 181

T1w T1-weighted. 15, 18, 37, 41, 50, 56, 77, 78, 80–83, 90, 93, 134, 140, 160, 161, 176, 180, 181, 186, 215, 219

T2w T2-weighted. 37, 219

TBI traumatic brain injury. 15, 17, 25, 26, 36, 38, 47–50, 52, 53, 55, 56, 59, 64, 91–95, 97–101, 112, 116, 118, 119, 123–125, 127–131, 145–147, 151, 152, 157, 159, 180, 181, 184–186, 188, 190–196, 215–219, 270, 271

TBM tensor-based morphometry. 86, 88, 92, 94, 177

VBM voxel-based morphometry. 86, 92, 93

VBSI Ventricle Boundary Shift Integral. 145, 151

VD vascular dementia. 85, 218

WM white matter. 37, 38, 40, 48, 50, 81, 83, 84, 91, 93, 128, 140, 185, 189–192, 194, 196, 217, 218

1 Introduction

Over the last decades there have been substantial advances in neuroimaging which now allow the *in-vivo* imaging of both brain anatomy and brain function. Neuroimaging has opened exciting areas of research that aim to investigate the deeper cause and progression of many neurological diseases. Analysing medical images manually is a very time consuming endeavour that requires expert knowledge. It is expensive and not scalable to large studies or image databases. Furthermore, manual approaches are subject to inter-/intra-rater variability and can be of limited sensitivity towards subtle yet important information encoded in the images. To address these challenges, the focus of this thesis is the development of robust methods for the *automatic* analysis of magnetic resonance (MR) brain images. Based on two clinical applications the developed techniques will be employed to investigate the potential to extract imaging biomarkers, such as structural volume and atrophy, in the context of both Alzheimer’s disease (AD) and traumatic brain injury (TBI).

This introductory chapter gives a brief overview of neuroimaging techniques in Section 1.1. The need for advances in the understanding, diagnosis and treatment of both AD and TBI is further outlined in Section 1.2 and Section 1.3 respectively. The motivation for this work is summarised in Section 1.4. Finally, Section 1.5 describes the main contributions of this thesis and its organisation.

1.1 Neuroimaging

A multitude of neuroimaging techniques that exploit different physical concepts has emerged. There is, however, no single imaging technique that allows the comprehensive analysis of an individual brain. Rather, the optimal imaging modality and imaging sequence needs to be chosen based on the clinical question that is to be answered.

In the following a brief overview over widely used imaging techniques is provided. Table 1.1 summarises different imaging modalities with their respective characteristics. A more detailed overview can be found in the literature (Ryan and Alexander, 2007; Belanger et al., 2007; Shenton

et al., 2012; Nasrallah and Dubroff, 2013).

1.1.1 Computed Tomography

Computed tomography (CT) is an established technique that reconstructs a three-dimensional (3D) object based on multiple projection images (radiographs) obtained through ionising radiation (X-rays). To obtain the radiographs from varying viewing angles an X-ray source and the corresponding detector are rotated around a fixed axis of the object. The 3D volume can then be reconstructed using tomographic reconstruction. CT is fast, available in most hospitals and often the primary imaging modality for patients suffering an acute head injury as it is sensitive to injuries that require immediate action, such as skull fractures, brain swelling or intracranial haemorrhage (Shenton et al., 2012).

1.1.2 Magnetic Resonance Imaging

Another established imaging modality is magnetic resonance imaging (MRI). In MRI hydrogen nuclei (protons) are aligned in a strong uniform magnetic field. Common field strengths are 1.5 or 3 Tesla (T). Unique spatial positions can be encoded with magnetic gradient fields. The aligned protons are subsequently excited using radio-frequency pulses. While the protons return to their initial state they also emit a radio-frequency signal that is then measured by receiver coils in the scanner. Tissue contrast can be obtained by measuring this signal which decays exponentially dependent on the characteristics of the respective tissue. There are numerous MR sequences to obtain specific tissue contrasts. Commonly used sequences include: T1-weighted (T1w) MRI, which measures the *spin-lattice* relaxation time and provides good white matter (WM)/grey matter (GM) contrast; T2-weighted (T2w) MRI, which measures the *spin-spin* relaxation time and has advantages for the visualisation of for example oedema or haemorrhage; fluid-attenuated inversion-recovery (FLAIR), which is sensitive to *e.g.* lesions and contusions; susceptibility weighted imaging (SWI), which is another sequence and very sensitive to hemorrhagic lesions and microbleeds. To illustrate differences between protocols, several MR sequences of the same patient are shown in Figure 1.1. During image acquisition the patient can be subjected to substantial acoustic noise caused by rapid changes of the magnetic gradient fields. Furthermore, as a result of the required uniform and strong magnetic field the scanner gantry is usually designed to be quite narrow which can be problematic for claustrophobic patients. In contrast to CT, MRI does not rely on ionising

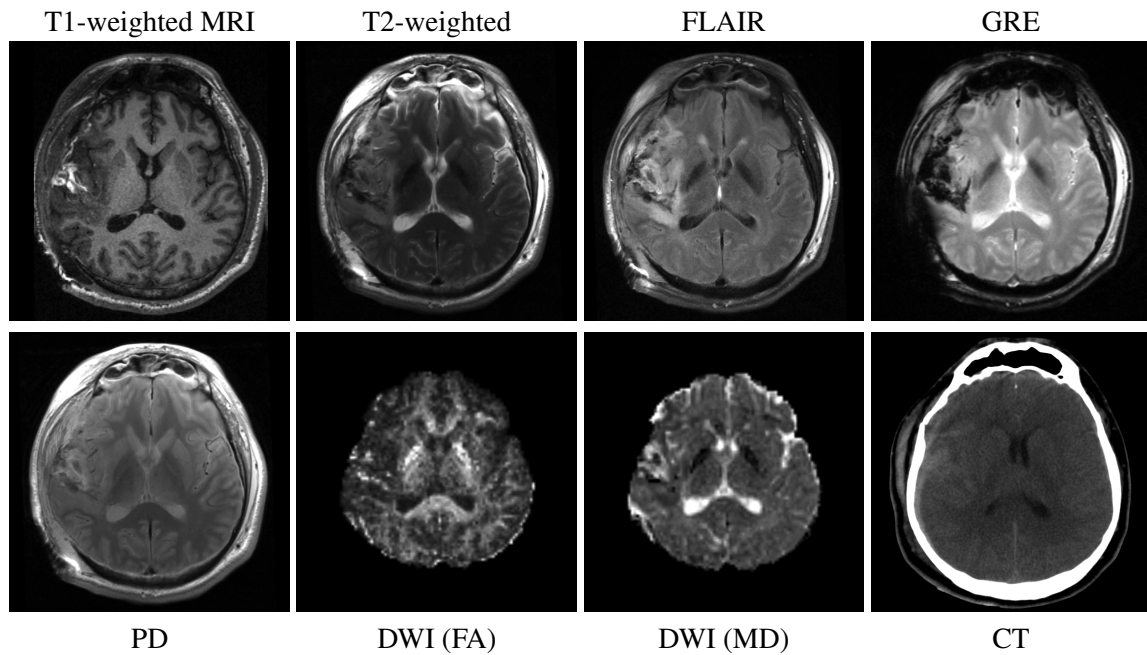


Figure 1.1: Brain images acquired with different imaging sequences/modalities from a patient with traumatic brain injury. The good tissue contrast in T1-weighted MR images and the pronounced contusions in the FLAIR sequence are apparent. Furthermore, gradient echo (GRE) and proton density (PD) weighted images are shown. Diffusion derived fractional anisotropy (FA) and mean diffusivity (MD) maps are also shown. CT is well suited to image bone injuries, oedema or intracranial bleeding.

radiation and is considered safe.

1.1.3 Diffusion Weighted Imaging

Diffusion weighted imaging (DWI) or diffusion tensor imaging (DTI) is a variant of MRI that allows a more detailed analysis of WM anatomy or tissue microstructure. In DWI local diffusion of water molecules can be measured so that microscopic information about the diffusion properties can be inferred. The motion of these water molecules is influenced by the highly heterogeneous characteristics of biological tissues, including compartments and barriers of varying diffusivity (Hagmann et al., 2006). While CT or standard MR imaging can visualise macroscopic structures, DWI allows the assessment of changes at the microstructure level. DWI has been shown to be particularly valuable to examine WM pathology for example in the context of TBI (Kinnunen et al., 2011; Shenton et al., 2012). The measurement of diffusion with respect to multiple directions allows the calculation of a so-called diffusion tensor. Based on this diffusion tensor, further quantities such as voxel-wise scalar values describing the level of anisotropy, the FA index, or the mean diffusivity

Table 1.1: Overview of selected 3D imaging modalities with a brief description of their respective mechanism, advantages and disadvantages. Table adapted from Shenton et al. (2012).

Modality	Mechanism	Advantages	Disadvantages
CT	imaging of an object using X-rays; calculation of a volume based on tomographic reconstruction (Radon transform) of numerous projections acquired from different viewing angles	fast acquisition; comparably cheap; available in most hospitals; medical instruments can be close to/within the scanner (no magnetic field); good for the imaging of bones; can reveal gross injuries that require immediate (surgical) intervention; can detect subdural haematomas, foreign objects, brain swelling; no normalisation required (Hounsfield units)	ionising radiation; low soft tissue contrast
MRI	radio-frequency pulses excite hydrogen nuclei that are aligned in a strong magnetic field; the change in spin signal is measured	higher resolution than CT; good soft tissue contrast; allows distinction between white and grey matter tissue; can detect macroscopic areas of white matter damage and contusions	strong magnetic field: no metal objects close to scanner; can cause claustrophobia and discomfort due to noise; normalisation of measurements required
DWI, DTI	special type of MR sequence that measures the diffusion properties of water that correlate with micro-structural tissue architecture	ideal imaging sequence to examine white matter integrity; can reveal white matter damage and allows the inspection of fibre tracts within the brain	same as MRI but longer scanning times
fMRI	measures changes in blood flow due to increased neuronal activity	quantification of activity patterns in resting state and during specific tasks	same as MRI
PET	uses positron-emitting radionuclide (a tracer, most commonly FDG); the tracer is usually injected; areas of increased tracer uptake are visualised based on emitted gamma-rays	provides information on the activity, or density of a type of protein (<i>e.g.</i> amyloid beta)	ionising radiation; expensive as the tracers need to be produced in cyclotrons; the dose of the tracer must be carefully controlled as FDG has a very short half-life

can be derived. More advanced methods, such as tractography, can extract WM fibre bundles based on which global brain connectivity can be investigated.

1.1.4 Functional Magnetic Resonance Imaging

In functional magnetic resonance imaging (fMRI) brain activity is measured using the distinct magnetic attributes of oxygenated and non-oxygenated blood. The principle of blood-oxygen-level-dependent (BOLD) fMRI is that regional blood flow increases due to corresponding brain activity (Belanger et al., 2007). These changes in blood flow can be measured while the patient is asked to perform specific cognitive tasks (task fMRI) or during normal brain activity (resting-state fMRI).

1.1.5 Positron Emission Tomography

Positron emission tomography (PET) allows functional imaging of the brain by measuring the uptake of a radioactive tracer. The tracer is usually injected into the patient's blood stream and designed to accumulate at certain receptors in the brain. While the tracer decays (β -decay) positrons are emitted that annihilate with electrons after travelling very short distances in the surrounding tissue. During this process two gamma-rays are emitted into opposite directions to conserve momentum. The gamma-rays are then measured by circular gamma-detectors, which allows the reconstruction of the spatial activation pattern. Important in PET imaging is the choice of an appropriate tracer. Most widely used is 18-fluorodeoxyglucose (FDG) a tracer based on glucose that allows the quantification of brain metabolism. In the context of AD, AV45 is another tracer of interest as it binds to amyloid beta ($A\beta$) which is a relevant biomarker protein in dementia. Using PET imaging, the reduced brain metabolism and increased amyloid concentration in a patient with AD in comparison to a healthy person is illustrated in Figure 1.2.

1.2 Alzheimer's Disease

1.2.1 Prevalence

The first patient with a dementia type that is now referred to as AD was reported more than a century ago (Alzheimer, 1907). Meanwhile, AD is one of the major challenges in neurology (Citron, 2004; Klafki et al., 2006) and the main cause for people aged 60 or older to live with a disability (Ferri et al., 2006). It is expected that in 2013 around 35.6 million people worldwide (Prince et al.,

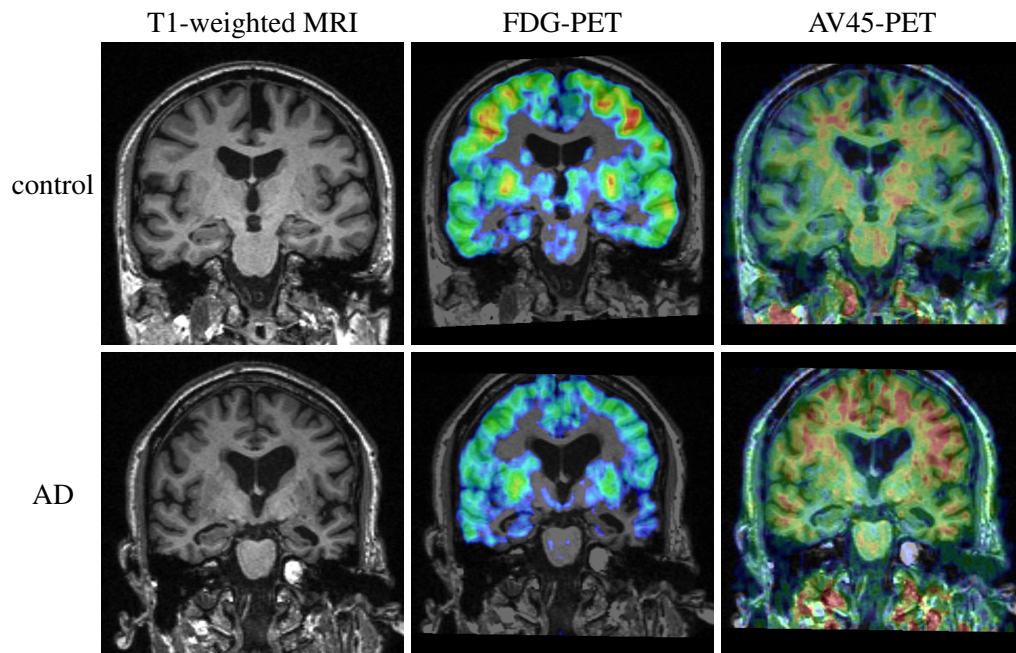


Figure 1.2: Example of T1w MR image (left), FDG-PET image (middle) and AV45-PET image (right) in coronal view for a healthy control (top, male, 66 years) and patient with AD (bottom, male, 75 years). The PET images are superimposed on the structural MR image with red indicating high signal and blue low signal respectively. The PET images show reduced metabolic activity (FDG) and increased amyloid beta ($A\beta$) binding (AV45-PET) in patients with AD.

2013) suffered from AD, a substantial increase from around 25 million in 2006 (Ferri et al., 2006; Brookmeyer et al., 2007). This number is expected to double every 20 years resulting in more than 100 million humans affected by AD in 2050 (Brookmeyer et al., 2007; Prince et al., 2013). With increasing age the prevalence of AD increases, for example from 0.9% at 65 years of age to 14.7% at 85 years of age respectively (Brookmeyer et al., 2007). At age 65 the lifetime risk of suffering an AD dementia is estimated at around 10.5% (Sperling et al., 2011). Genetic risk factors related to the apolipoprotein E (ApoE) genotype are well established (Lehtovirta et al., 1996; Roberson and Mucke, 2006). Carriers of the ApoE $\epsilon 4$ allele are at increased risk compared to non-carriers (Roberson and Mucke, 2006). The risk of developing AD could be reduced substantially if screening and disease delaying drugs were available (Sperling et al., 2011). Unfortunately, there is only symptomatic treatment available and effective drugs to delay or even cure AD remain unavailable (Citron, 2004; Klafki et al., 2006; Blennow et al., 2006; Roberson and Mucke, 2006; Davatzikos et al., 2011; Karran and Hardy, 2014).

More than a third of AD patients are expected to require care that is equivalent to a nursing-home

(Brookmeyer et al., 2007) and that is often provided by the patient's family (Ferri et al., 2006). This endeavour is, however, not only psychologically and practically demanding but also imposes a substantial financial burden to both caregivers and society (Ferri et al., 2006).

1.2.2 Diagnosis and disease progression

To obtain a diagnosis of *probable* AD, the National Institute of Neurological and Communicative Disorders (NINCDS)-Alzheimers Disease and Related Disorders Association (ADRDA) criterion was formulated in McKhann et al. (1984). It relies on clinical tests such as the mini-mental state examination (MMSE) (Folstein et al., 1975). A *definite* AD diagnosis can only be obtained based on histopathological evidence (McKhann et al., 1984). Here, the main pathological characteristics of AD are the selective loss of neurons as well as accumulated A β plaques and neurofibrillary tangles (McKhann et al., 1984; Braak and Braak, 1991; Wenk, 2003; Klafki et al., 2006; Heppner et al., 2015). It is assumed that abnormal processing of A β , even if it does not cause cognitive impairment directly (Jack Jr et al., 2010b), initiates a cascade of degenerative processes that eventually manifests as AD (Wenk, 2003; Karran et al., 2011; Falahehi et al., 2014; Heppner et al., 2015). Thus, many efforts to tackle AD focus on the blockage of A β peptides and the prevention of further neuronal degeneration (Citron, 2004; Roberson and Mucke, 2006). However, several clinical phase III trials targeting A β have been not successful (Karran et al., 2011; Golde et al., 2011; Karran and Hardy, 2014). A β targeted therapies might be effective in preventing AD, however they are potentially ineffective in symptomatic patients (Golde et al., 2011). There is increasing discussion related to the A β cascade hypothesis (Karran et al., 2011; Golde et al., 2011; Karran and Hardy, 2014) with proposed modifications being a more central role of tau and the inclusion of neuroinflammation (Heppner et al., 2015).

Next to A β , tau proteins, which stabilise microtubules, play an important role (Klafki et al., 2006; Karran et al., 2011). In a hypothetical model, Klafki et al. (2006) describe how pathological conditions, possibly triggered by A β , can lead to an abnormal hyperphosphorylation of tau. This affects the function of the tau proteins and potentially leads to a destabilisation of microtubules and structural alterations such as the aggregation of tau as neurofibrillary tangles (Klafki et al., 2006). Eventually tau pathology can cause synaptic dysfunction and neuronal loss (Klafki et al., 2006; Karran et al., 2011). Even though tau pathology is considered a downstream event of A β accumulation (Karran et al., 2011) it seems to correlate better with cognitive decline (Heneka and

Obanion, 2007; Karran and Hardy, 2014). There is increased interest in tau-based drug discovery despite tau targeting therapies being more difficult than those directed to A β (Brunden et al., 2010).

Heneka and Obanion (2007) state that A β plaques or tau pathology might not explain all AD related symptoms and that A β potentially activates microglial cells resulting in an inflammatory response. Neuroinflammation markers are typically increased in patients with AD related pathology (Szekely et al., 2004; Klafki et al., 2006; Heneka and Obanion, 2007). Insoluble A β peptides, neurofibrillary tangles but also degenerating tissue are stimuli for neuroinflammation (Akiyama et al., 2000). The inflammatory processes can impair both brain function, especially in the earlier disease stages, and brain structure (Heneka and Obanion, 2007). The molecules evoked by neuroinflammation may be toxic to cells ultimately resulting in cell death (Heneka and Obanion, 2007). It has also been shown that the long-term use of nonsteroidal anti-inflammatory drugs has the potential to reduce the risk of AD (Szekely et al., 2004; Klafki et al., 2006; Heneka and Obanion, 2007).

Treatment within the rather long pre-symptomatic phase of AD might offer a unique chance to delay the disease onset (Szekely et al., 2004; Golde et al., 2011; Sperling et al., 2011). The first biomarkers indicating AD induced brain damage relate to A β , which is usually measured via cerebrospinal fluid (CSF) directly or indirectly via tracer retention in PET brain images (Jack Jr et al., 2010b; Falahati et al., 2014). Further, it has been shown in many studies that with progression of the disease there is significant atrophy in medial temporal lobe (MTL) structures such as the hippocampus, amygdala, and entorhinal and parahippocampal cortices (Jack Jr. et al., 1997; Wenk, 2003; Jack Jr et al., 2010b; Klöppel et al., 2012). A hypothetical model for AD progression is illustrated in Figure 1.3. The pronounced atrophy in MTL regions is shown in Figure 1.4.

1.2.3 Diagnostic role of neuroimaging

Recently, the NINCDS-ADRDA criterion for AD has been revised (Dubois et al., 2007; Jack Jr. et al., 2011; McKhann et al., 2011; Albert et al., 2011; Sperling et al., 2011) to take novel insights into account. The updated criterion now better discriminates AD from other dementia types (McKhann et al., 2011) such as Lewy Body dementia (McKeith et al., 2007), vascular dementia (Roman et al., 1993) and frontotemporal dementia (McKhann et al., 2001). Biomarker evidence available through PET or structural MRI has been incorporated. Particular emphasis has been placed on atrophy in MTL regions (Klöppel et al., 2012).

The diagnostic potential of biomarkers based on structural imaging has been outlined in Frisoni

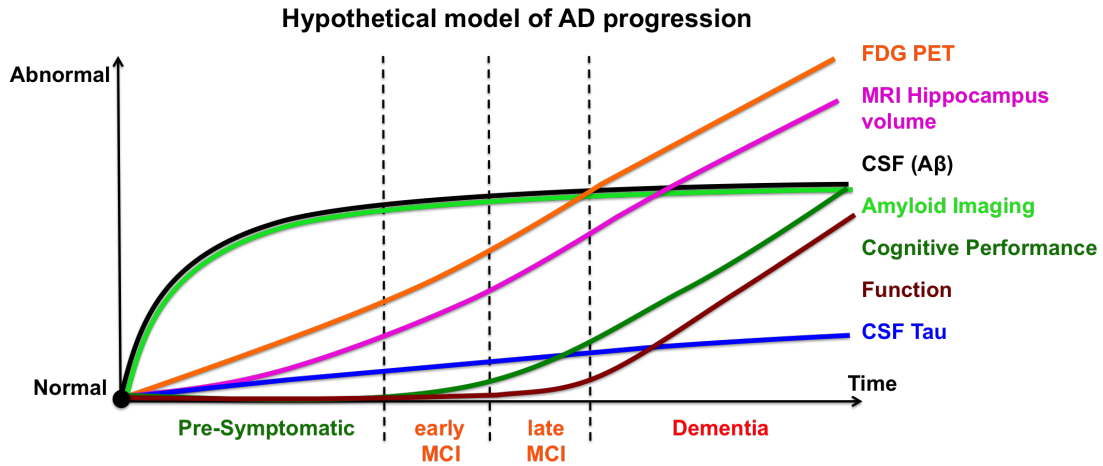


Figure 1.3: Hypothetical trajectories of biomarkers during the progression of AD. Figure adapted from Jack Jr et al. (2010a); Aisen et al. (2010).

et al. (2010) or Klöppel et al. (2012). These biomarkers can support or indicate a diagnosis whenever trained neuroradiologists are not available or accelerate expert diagnoses without sacrificing accuracy (Klöppel et al., 2008; Klöppel et al., 2012). Automatically extracted biomarkers could further increase objectivity in the disease assessment and could allow for improved differential diagnosis to rule out other forms of dementia (Frisoni et al., 2010; Klöppel et al., 2012). Another important avenue is the use of biomarkers for AD screening or for the recruitment for pharmaceutical trials (Klöppel et al., 2012; Falahati et al., 2014), where homogeneous patient cohorts might be required.

Subtle alterations in brain structures can be measured using MRI. Klöppel et al. (2008) showed that automatic methods can perform comparable to trained radiologists when classifying MR images of patients with AD. However, there is uncertainty that information based on *individual* brain structures is sufficient to fully characterise the complex progression of AD or even allow for a differential dementia diagnosis (Klöppel et al., 2008; Falahati et al., 2014). It might thus be desirable to analyse a number of structures of the whole brain rather than a limited number of selected brain structures only.

1.2.4 Mild Cognitive Impairment

Of particular interest are patients with mild cognitive impairment (MCI), who usually show slightly impaired memory but no other AD like symptoms that would impact their daily living (Dubois et al., 2007; Gerardin et al., 2009). MCI is not very well defined and can be considered as a

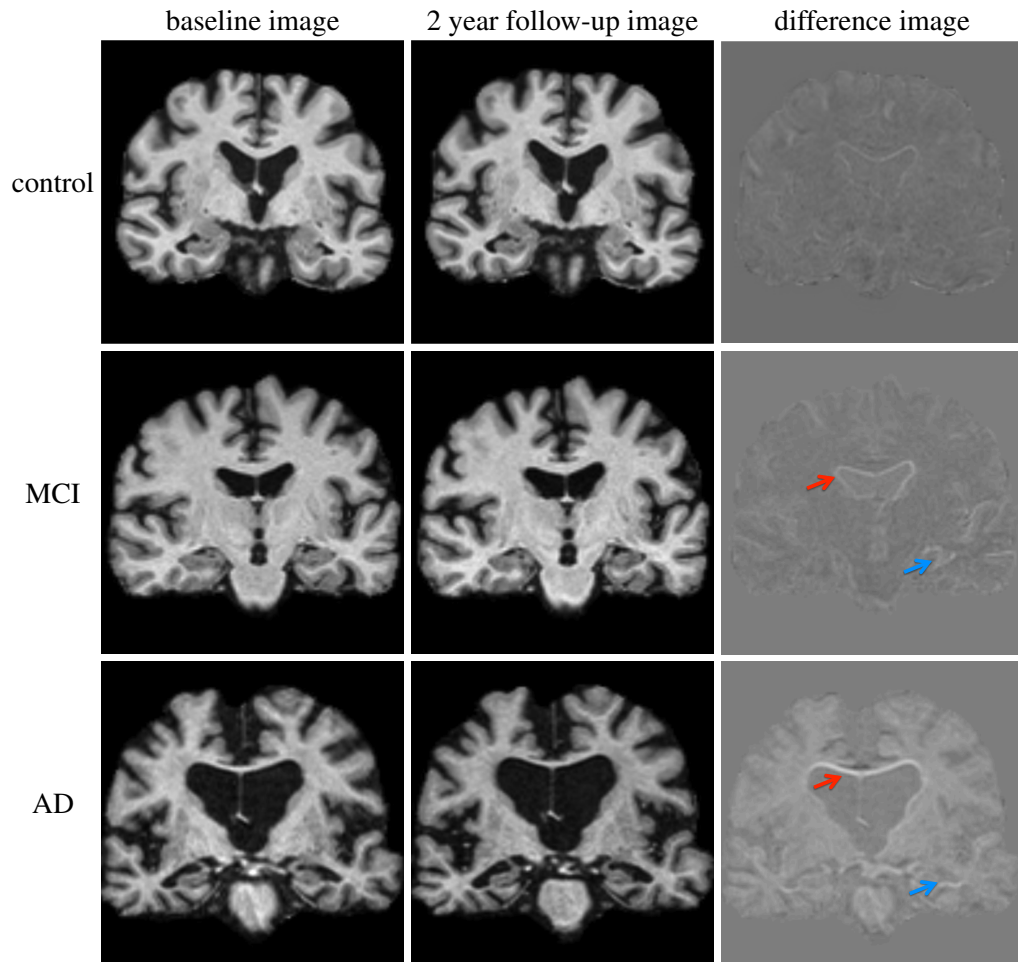


Figure 1.4: Three examples of MR images (brain extracted) of subjects from the ADNI1 cohort visualised in a coronal view. A healthy control subject (male, 84.8 years at baseline) is shown in the top row. A MCI patient (female, 71.8 years at baseline) that converted to AD after three years is shown in the middle row. An AD patient (male, 77.5 years at baseline) is shown in the bottom row. Left: baseline MR image, Middle: 2 year follow-up MR image, Right: difference image of rigidly aligned images. The differences are visually very subtle but the increased atrophy in the medial temporal lobe (blue arrows) and the enlarged ventricles (red arrows) are apparent in the difference image.

surrogate class for the early stages of multiple types of dementia. The unclear definition potentially causes high variability between clinical centres and should be considered when conducting clinical trials (Dubois et al., 2007). As an example subtle anatomical changes over a period of two years are illustrated in Figure 1.4 for a healthy control subject, a MCI subject converting to AD and a patient with AD.

Subjects with MCI do not fulfil the clinical AD criterion (Petersen et al., 1999; Dubois et al., 2007), but are at increased risk of conversion to AD (Dubois and Albert, 2004; Davatzikos et al., 2011). However, not all subjects with MCI will go on to develop AD (Dubois et al., 2007; Cuingnet et al., 2011). Discriminating these patients with stable MCI (sMCI) from patients with progressive MCI (pMCI), those converting to AD, is of particular relevance to clinicians, patients and caregivers alike. It is believed that identifying potential AD patients before any functional impairment has occurred offers a better opportunity for early intervention and to evaluate the disease progression (Dubois et al., 2007; Colliot et al., 2008). The targeted treatment of early disease mechanisms in the early phases of AD is likely to be more effective than later treatment when the disease is advanced (Dubois et al., 2007). Reliably identifying patients with stable or progressive MCI is thus of high value when treatment decisions have to be made (Davatzikos et al., 2011), but also when clinical trials are being conducted (Dubois et al., 2007; Gerardin et al., 2009; Klöppel et al., 2012).

Modern neuroimaging can help to improve the accuracy of MCI diagnosis by adding positive predictive value when combined with other clinical measures (Dubois et al., 2007; Klöppel et al., 2008; Colliot et al., 2008; Cuingnet et al., 2011).

1.2.5 Potential of neuroimaging

It is becoming clear that structural MRI in combination with other criteria, such as PET or CSF-based markers measuring A β , can offer a powerful approach to detect pathological AD-related change years before the onset of a fully developed and probably irreversible AD dementia (Frisoni et al., 2010; Sperling et al., 2011). Structural MRI will play an important role in both future diagnosis but also treatment of AD (Fennema-Notestine et al., 2009; Frisoni et al., 2010; Klöppel et al., 2012). The development of automatic and robust techniques to automatically and systematically assess whole-brain MR images is thus an important factor to further increase the usability of structural imaging in the context of AD (Frisoni et al., 2010).

1.3 Traumatic Brain Injury

1.3.1 Prevalence

With an estimated annual global incidence of 6.8 million cases, TBI imposes a significant burden on patients, their families, and health services (Maas et al., 2008; Irimia et al., 2012). TBI is often referred to as the “silent epidemic” as symptoms, such as memory loss or cognitive deficits, tend to be less apparent (Langlois et al., 2004). However, a report summarising TBI-related data in the United States for the years 1996 to 2001 (Langlois et al., 2004) underlines the relevance of this disease: there were annually at least 1.4 million new cases of TBI. Most of these patients were treated in an emergency department (79.6%), but a substantial number was hospitalised (16.8%) or even died (3.6%). While most children required treatment in the emergency department only (>90%), the report also suggests that the mortality increases with age to more than 7% for those aged 55 or older (Langlois et al., 2004). Next to assaults (11%) and motor-vehicle accidents (20%), falls (28%) are the most common cause of TBI (Langlois et al., 2004). The mortality for men is almost three times higher than for females (Langlois et al., 2004).

Research findings on TBI obtained while doing sports (De Beaumont et al., 2009) or in military conflicts (Taber et al., 2006) have increasingly brought the disease into the focus of the public (Irimia et al., 2011). Moderate and severe TBIs are further assumed to be an important risk-factor for dementias such as AD in late life (Plassman et al., 2000; Shively et al., 2012; Bigler, 2013; Smith et al., 2013; Sharp et al., 2014).

1.3.2 Focal and diffuse injury

TBI is typically caused by blunt force injury, penetrating injury, or blast injury and its pathology dependent on the forces associated with the event (Smith, 2011). The pathological processes following the injury are highly complex and the exact mechanism causing functional impairment not entirely understood (Warner et al., 2010a; Smith, 2011). A common categorisation of disease-related processes distinguishes primary and secondary injuries (Graham et al., 2000; Smith, 2011; Sharp et al., 2014).

Primary injuries have a focal component caused by events such as the direct impact of an object hitting the head. Common consequences that should be differentiated include skull fractures, contusions, haemorrhage and haematomas (Meythaler et al., 2001; Smith, 2011; Sharp et al., 2014). Intracranial haemorrhages are the most common cause for rapid clinical deterioration and the complications are generally dependent on location and size of the haematoma (Smith, 2011). Expanding haematomas and/or brain swelling can also lead to an increased intracranial pressure that potentially result in serious conditions such as brain ischemia (Smith, 2011). Another injury mechanism, diffuse injury, is initiated by the strong accelerating or decelerating forces during the injury event. This is referred to as diffuse axonal injury (DAI) and assumed to be the predominant mechanism of TBI (Meythaler et al., 2001; Warner et al., 2010b; Shenton et al., 2012; Sharp et al., 2014). TBI often results in both focal and diffuse injury which can evolve over time (Smith, 2011). DAI is usually a local injury of discrete regions and not necessarily an injury to the whole brain (Meythaler et al., 2001; Warner et al., 2010b). DAI is also a major determinant of disease outcome (Graham et al., 2000) and considered a continuous process that develops from focal axonal changes to ongoing axonal separation (Büki and Povlishock, 2006). Pathology related to DAI can be characterised by microscopic characteristics such as widespread damage of axons (Meythaler et al., 2001; Ding et al., 2008).

Next to the consequences of the primary head injury such as focal lesions, it is assumed that complex secondary pathophysiological processes continue damaging brain cells and thus contribute to TBI (Meythaler et al., 2001; Ding et al., 2008; Warner et al., 2010b; Sharp et al., 2014). Neurodegenerative diseases, inflammation, infections or increased intracranial pressure can potentially be initiated by TBI and the resulting damage of WM tracts (Ramlackhansingh et al., 2011; Smith et al., 2013; Sharp et al., 2014). It is hypothesised that neuroinflammatory processes following a head injury can lead to increased A β deposition and formation of tangles, both hall marks of AD (Griffin et al., 1998; Smith, 2011). A comprehensive description of TBI-related pathology and long-term secondary processes triggered by the injury event is given in Smith (2011).

1.3.3 Clinical characterisation

The *acute stage* of TBI is often clinically characterised with the Glasgow Coma Score (GCS). The GCS is a clinical score that quantifies a patient's level of consciousness at the acute stage of the injury (Teasdale and Jennett, 1974). The GCS is often assessed several times during the first hours

and days after the injury *e.g.* at the injury site and at hospital admission.

The GCS is the most common criteria to determine the severity (*e.g.* mild, moderate, severe) of a brain injury (Sherer et al., 2008; Schoenberger et al., 2009; Arlinghaus et al., 2011; Smith, 2011). Limitations that reduce the prognostic potential of the GCS include inconsistencies and controversy regarding the ideal time of assessment (Sherer et al., 2008). Next to the GCS other parameters such as loss of consciousness (LOC) duration and post-traumatic amnesia (PTA) duration have been shown to correlate with the outcome of TBI (Asikainen et al., 1998; Schoenberger et al., 2009; Arlinghaus et al., 2011). Intercorrelations among GCS, LOC and PTA have been investigated and confirmed in Sherer et al. (2008). Regardless of their predictive value, severity criteria defined based on these parameters have substantial limitations (Malec et al., 2007; Arlinghaus et al., 2011): First, many cases might remain unclassified if essential parameters are absent (Malec et al., 2007). Second, there are several confounding factors that can reduce the usefulness of GCS, LOC and PTA such as intoxication, intubation, sedation or facial injuries (Sherer et al., 2008; Arlinghaus et al., 2011).

The more recently proposed Mayo classification system (Malec et al., 2007) addresses many of these limitations by using positive evidence (Arlinghaus et al., 2011). The Mayo system distinguishes between moderate-severe (definite) TBI, mild (probable) TBI and symptomatic (possible) TBI. Patients are associated with the most severe group of which at least one predefined criterion is fulfilled. The system is thus not directly affected by missing data. The criteria include GCS, LOC and PTA but also other symptoms such as blurred vision, dizziness or headache. Further, radiological imaging findings such as fractures, contusions, haematomas or haemorrhage are incorporated (Malec et al., 2007).

Also assessed at the acute stage is the Marshall Classification Score (MCS) (Marshall et al. (1991), *cf.* Appendix C), a score based on the worst acute CT image within 24 hours of injury. MCS takes into account brain pathology such as lesion load or the presence of oedema and midline shift caused by the injury.

In contrast to this, the Glasgow Outcome Score (GOS) (Jennett and Bond (1975); Jennett et al. (1981), *cf.* Appendix D) is a clinical measure categorising the outcome of TBI at the *chronic stage* and is assessed around six months after injury or once the TBI outcome is considered stable.

1.3.4 Neuroimaging in TBI

Even though most of the cases of TBI are mild or moderate, a substantial fraction of up to 50% might suffer from a long-term disability (Thornhill et al., 2000; McAllister et al., 2006; Belanger et al., 2010; Shenton et al., 2012). This is particularly relevant as many of these mild TBI (mTBI) cases affect young adults with many more years to live (Shenton et al., 2012). Evidence from neuroimaging such as MRI or CT is often very subtle or completely absent, so that persistent symptoms tend to be explained through post-traumatic stress disorders or depression (Shenton et al., 2012). Four examples of subjects with TBI are visualised in Figure 1.5. These examples illustrate the heterogeneity of structural changes that can be subtle in both mild and severe TBI but also very apparent and variable.

Although patterns of brain alteration have been shown to be predictors of outcome, such use of imaging data is mainly based on expert interpretation of visually inspected X-ray CT images. In Maas et al. (2007) the authors confirmed in a multivariable analysis the predictive value of individual characteristics quantified from CT for 6-months outcome prediction. In Jacobs et al. (2010) it was shown for subjects with mTBI that CT-based criteria are a valuable indicator to identify patients at risk of deterioration. However, they are only of limited value to predict the eventual outcome when compared to criteria such as patient age, alcohol intake or extra-cranial injuries (Jacobs et al., 2010).

1.3.5 Importance of multiple modalities

TBI is a very heterogeneous disorder and thus images from multiple modalities (*cf.* Section 1.1) are required to characterise the disease (Sharp and Ham, 2011; Shenton et al., 2012). While contusions and lesions can be well assessed on MR FLAIR sequences, CT imaging is the modality of choice to rule out skull fractures or other gross injuries that require immediate action (Shenton et al., 2012). CT provides critical information when treatment decisions at the acute TBI stage need to be made (Maas et al., 2007; Jacobs et al., 2010). Next to this T1w MRI provides good tissue contrast allowing the accurate segmentation of distinct anatomical structures. The definition of these regions of interest (ROIs) is an important step towards a ROI-based analysis of information from DWI or functional imaging. As DWI is sensitive to micro-structural changes in WM it has great potential to allow a better understanding of DAIs (Kinnunen et al., 2011; Shenton et al., 2012; Sharp et al., 2014). In Bonnelle et al. (2012) the authors employed DWI and fMRI to investigate the

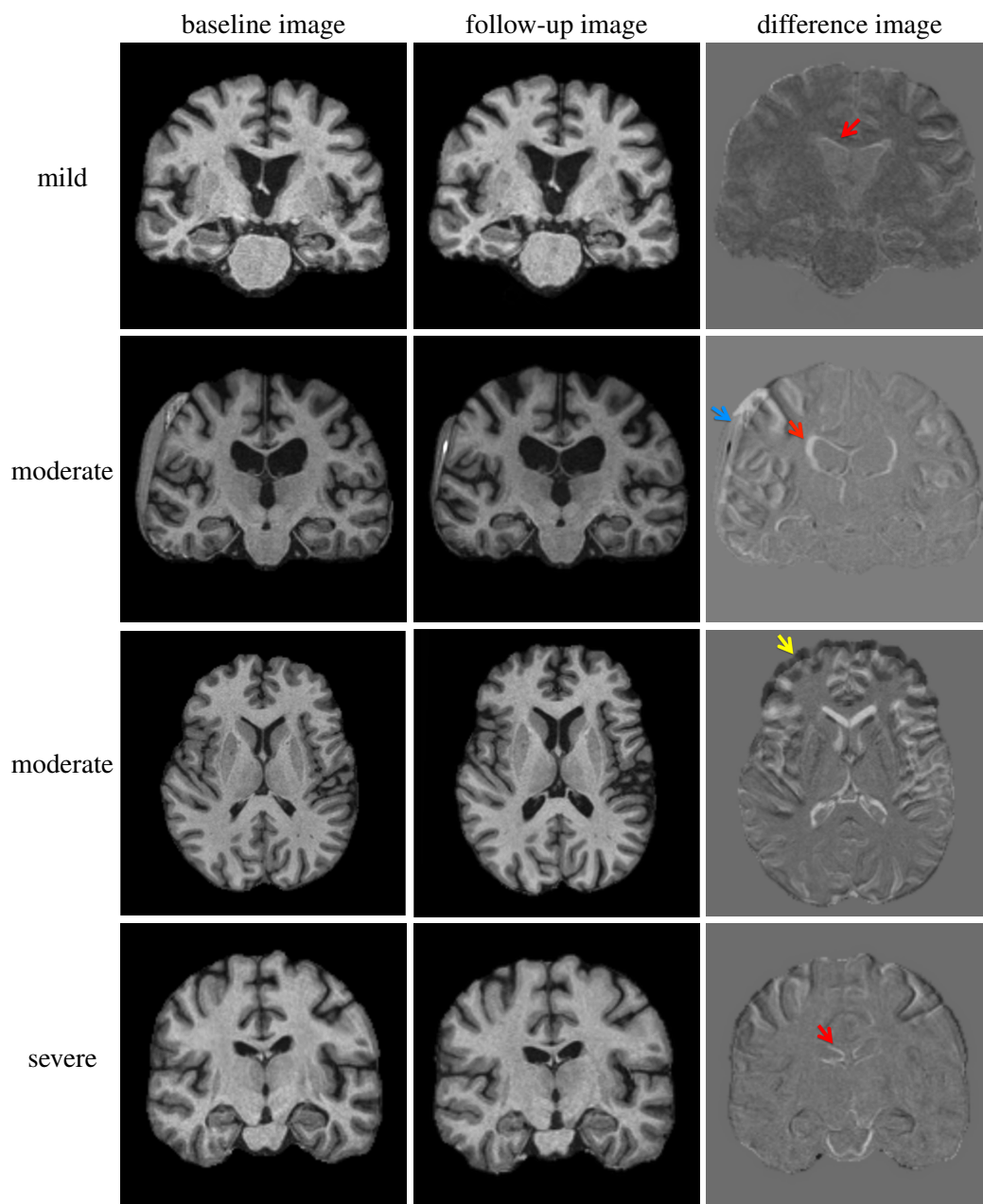


Figure 1.5: Four examples of MR images (brain extracted) of subjects from a prospective TBI cohort visualised in coronal or axial view. Top row: patient with mild TBI (male, 72 years, GCS: 14, GOSe: 8, MCS: 1, fall accident). Second row: patient with moderate TBI (female, 55 years, GCS: 2, GOSe: 4, MCS: 4, fall accident). Third row: moderate TBI patient (male, 38 years, GCS: 11, GOSe: 5, MCS: n/a, car accident). Bottom row: patient with severe TBI (female, 33 years, GCS: 4, GOSe: 3, MCS: 2, car accident). Left: baseline MR image acquired in the acute phase (days after the injury), Middle: follow-up MR image acquired in the chronic phase (months after the injury), Right: difference image of rigidly aligned images. Enlarged ventricles (red arrows), a subdural haematoma (blue arrow) and deformed/compressed frontal region (yellow arrow) are indicated in the difference images.

importance of structural integrity for brain network function following TBI. Examples of typical imaging sequences acquired in the context of TBI are illustrated in Figure 1.1.

In this thesis particular focus is placed on structural MR images, in which subtle volumetric changes can be assessed. There is consensus that there is ongoing structural atrophy following TBI (Bigler, 2001). However, as these structural changes might be difficult to assess visually on MRI or CT there is a high need for advanced methods that allow the quantification of atrophy (Bigler, 2001). The advancement of sensitive neuroimaging techniques is thus critical as it offers the potential to better understand, diagnose and treat TBI (Bigler, 2001; Shenton et al., 2012; Irimia et al., 2012).

1.3.6 TBI: An understudied disease

In contrast to other diseases of the human brain, such as AD, there is only a limited number of studies investigating the spatial distribution of structural changes (Bendlin et al., 2008; Warner et al., 2010b). Recently, studies have found that next to a reduction in total brain volume several distinct structures such as amygdala, hippocampus or thalamus are involved in TBI (Bigler, 2001; Strangman et al., 2010; Warner et al., 2010b,a; Ramlackhansingh et al., 2011; Irimia et al., 2012). The need for accurate prognostic assessments was formulated already by Jennett et al. (1981), however, prediction of TBI outcome remains a challenge. World-wide, TBI and the related processes are an understudied research area (Meythaler et al., 2001) and standard models to predict the outcome of TBI patients remain unavailable (Irimia et al., 2012). It is stated in Irimia et al. (2012) that the combination of volumetric measures with brain connectivity/integrity measures from DTI (*e.g.* Bendlin et al. (2008); Kinnunen et al. (2011)) or with functional measures obtained through fMRI (*e.g.* McAllister et al. (2006); Sharp et al. (2011); McDonald et al. (2012)) might be key for future TBI research. Further, more recent MR imaging protocols such as SWI have great potential in the context of DAI due to their sensitivity to microbleeds (Sharp and Ham, 2011).

1.4 Motivation

Falahati et al. (2014) concluded that ROI based biomarkers offer advantages over, often very high dimensional, voxel-based features. Integrating information that can be readily interpreted by a clinician is easier than the integration of abstract features calculated by complex machine learning

techniques (Falahati et al., 2014).

One of the main goals of this thesis is to develop methodology that automatically extracts structural ROIs of the whole brain. Based on these ROIs both structural volume and structural atrophy can be readily quantified. Particular focus is placed on the methods' robustness when applied to diseased brain anatomy. In AD substantial brain atrophy poses challenges to automatic methods. In TBI, lesions, contusions but also resections of brain tissue are common obstacles that render the automatic extraction of biomarkers difficult. Visual examples of MR brain images of both patients with AD and TBI are shown in Figure 1.6. The focus of this work lies on the volumetric analysis of single time point images but also on the consistent estimation of structural atrophy from image sequences. The developed methods should support and be applicable in both cross-sectional and longitudinal studies. Of particular interest is the evaluation of biomarkers extracted with the proposed techniques in clinically relevant scenarios, specifically in the context of both AD and TBI.

1.5 Contributions and organisation of the thesis

1.5.1 Contributions

The following gives an itemised overview over the main contributions of this thesis:

- **Multi-Atlas Label Propagation with Expectation-Maximisation based refinement**

A fully-automatic segmentation framework called Multi-Atlas Label Propagation with Expectation-Maximisation based refinement (MALP-EM) is developed. MALP-EM allows the segmentation of a whole brain MR image into multiple distinct ROIs. The approach combines established segmentation techniques and adapts them to increase robustness. A thorough validation on a benchmark dataset and visual expert quality assessment of the segmentations based on selected subjects from a TBI cohort is conducted. The experiments confirm both the method's high accuracy and its robustness.

- **Consistent quantification of structural atrophy**

Two methods are proposed that measure structural atrophy from longitudinal MRIs acquired at multiple time points. While one method relies on intensity differences to estimate the relative volume change of individual structures, the other approach simultaneously segments the image sequences into numerous ROIs.

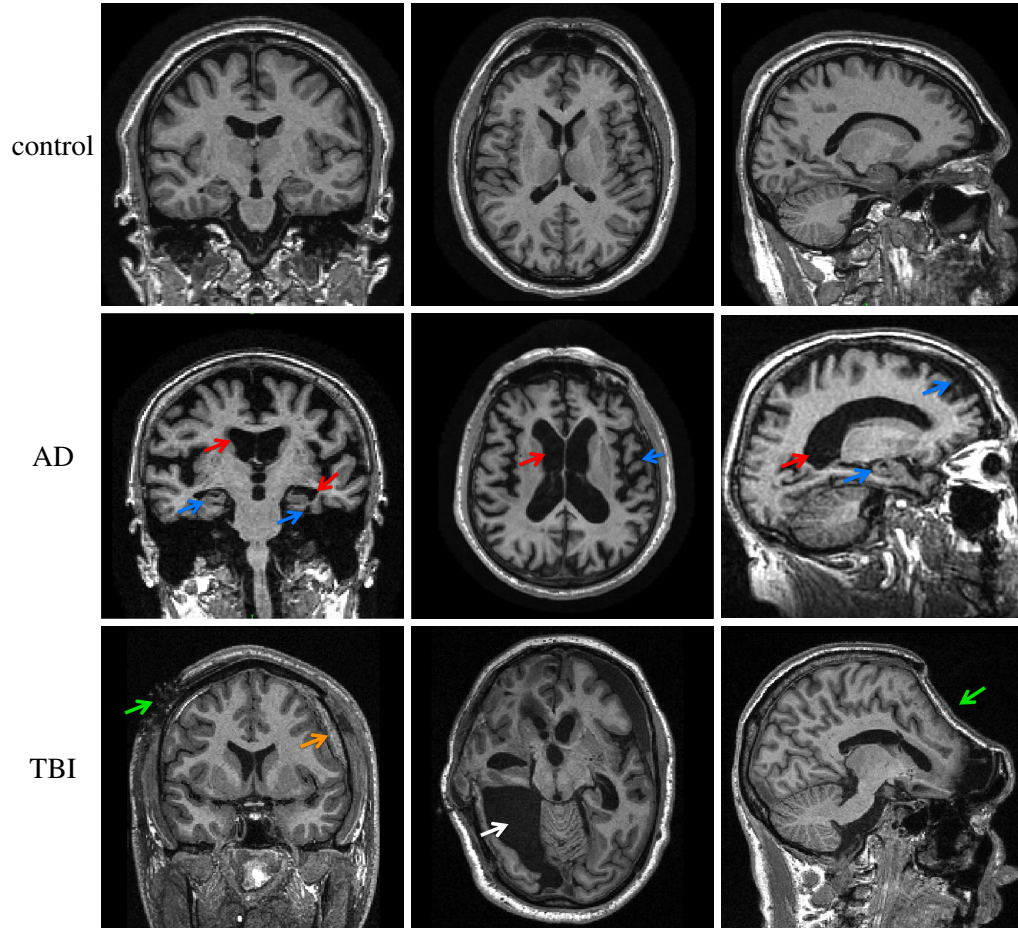


Figure 1.6: Top row: Example of a healthy control subject (male, 65.1 years) from the ADNI1 cohort visualised in coronal, axial and sagittal view. Middle row: Three examples of MR images of subjects with AD from the ADNI1 cohort (from left to right: female, 84.5 years in coronal; male, 80.2 years in axial; male, 77.5 years in sagittal view). Bottom row: Three examples of MR images of subjects with TBI from a retrospective TBI cohort (from left to right: male, 23 years, GCS=5, GOS=1, MCS=5; male, 45 years, GCS=7, GOS=3, MCS=5; male 42 years, GCS=14, GOS=4, MCS=5). Indicated with arrows: enlarged ventricles (red), pronounced hippocampal/medial temporal lobe atrophy (blue), skull fracture/deformation (green), tissue resection (white), subdural haematoma (orange).

- **Analysis of Alzheimer’s disease and traumatic brain injury cohorts using the developed methods**

The developed methods are applied to a large AD cohort from the Alzheimer’s Disease Neuroimaging Initiative (<http://adni.loni.ucla.edu>, last accessed 29 June 2015) (ADNI) and a prospective TBI cohort, acquired in the course of the TBicare project (<http://www.tbicare.eu>, last accessed: 24 June 2015). Both cross-sectional biomarkers (*i.e.* structural volumes) and longitudinal biomarkers (*i.e.* structural atrophy) are extracted. The biomarkers’ potential to support AD disease diagnosis and predict its progression is investigated. Further, correlations with clinical variables such as the GOS are identified in the context of TBI.

- **Patch-based Evaluation of Image Segmentation**

A novel similarity measure called Patch-based Evaluation of Image Segmentation (PEIS) is presented. PEIS allows to visualise and quantitatively assess the accuracy of a given image segmentation with respect to a ‘ground truth’.

1.5.2 Thesis overview

In Chapter 2 a detailed overview over existing, atlas-based automatic brain segmentation methods is provided. Furthermore, the role of quantitative biomarkers extracted from MRI is reviewed in the context of both AD and TBI. A robust whole-brain segmentation framework is described and evaluated in Chapter 3. In Chapter 4 two alternative approaches to accurately measure longitudinal volume change in image series are described. The developed methods are employed to extract biomarkers based on structural volume and atrophy to analyse an AD cohort in Chapter 5 and a TBI cohort in Chapter 6. A novel approach to quantify and visualise the similarity of image segmentations is presented in Chapter 7. The thesis concludes with a discussion of the presented results, their limitations and an outline of future research directions in Chapter 8. A full list of articles published during the work towards this thesis is provided in Appendix A.

2 Background

The extraction of a single or multiple ROIs from a subject's brain image is essential for the calculation of many biomarkers which are clinically useful. The process of subdividing an image into distinct regions is referred to as *segmentation*. Once a brain scan is segmented into its individual anatomical components, volumes or shape related measures can be readily quantified.

In this chapter, an overview of existing atlas-based segmentation methods is given in Section 2.1. Even though most techniques are applicable to the segmentation of other organs and imaging modalities, particular focus is placed on methods for the segmentation of T1w MR brain images into anatomical structures. In Section 2.2 several publicly available brain atlases are described that are essential for automatic methods as they allow the incorporation of expert knowledge into the segmentation process. In Section 2.3 and Section 2.4 the literature on biomarkers from structural MRI is reviewed with respect to both AD and TBI.

2.1 Atlas-based segmentation methods

In atlas-based methods the segmentation of a brain image is inferred by aligning M brain atlases denoted by $\mathbf{A}_m$ with $m = 1, \dots, M$ with the image. A brain atlas is usually created by an expert by manually or semi-automatically annotating a given volumetric brain scan. The term atlas comprises the annotated atlas intensity image $\mathbf{I}_m^A$ and the corresponding reference segmentation $\mathbf{L}_m^A$, thus $\mathbf{A}_m = \{\mathbf{I}_m^A, \mathbf{L}_m^A\}$. A single brain atlas is illustrated in Figure 2.1. A detailed overview over a selection of commonly used brain atlases is provided in Section 2.2.

The aim of whole-brain segmentation is to segment a target image $\mathbf{I}_{\text{tgt}}$ into K distinct structural ROIs. The unsegmented image $\mathbf{I}_{\text{tgt}} \in \Omega$ is indexed as $\mathbf{I}_{\text{tgt}} = \{y_1, y_2, \dots, y_N\}$ where $y_i \in \mathbb{R}^+$. Here y_i ($i = 1, \dots, N$) denotes the intensity value of the i -th voxel. The corresponding, inferred label map $\mathbf{L}_{\text{tgt}}$ is indexed accordingly as $\mathbf{L}_{\text{tgt}} = \{l_1, l_2, \dots, l_N\}$. The probabilistic (or soft) segmentation is denoted by $\mathbf{P}_{\text{tgt}} = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N\}$, where $\mathbf{p}_i$ is a K -valued vector with $\|\mathbf{p}_i\|_1 = 1$ and $\mathbf{p}_i(k) \geq 0$

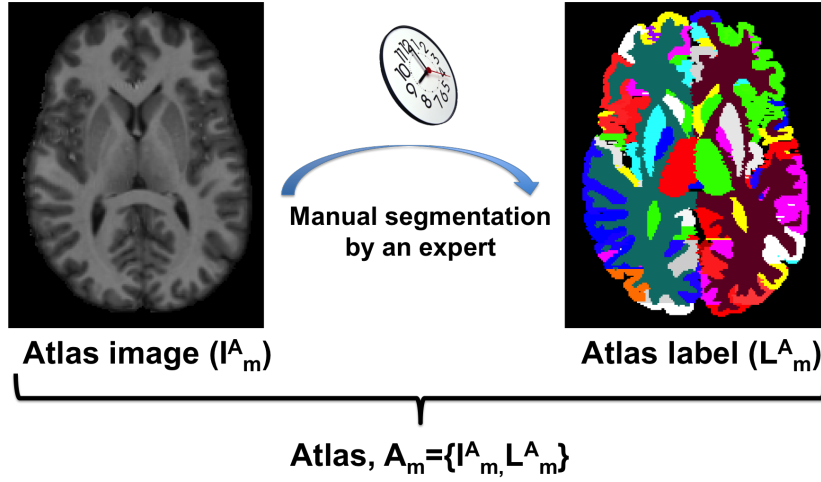


Figure 2.1: A brain atlas $\mathbf{A}_m$ consists of an intensity image $\mathbf{I}_m^A$ with corresponding segmentation $\mathbf{L}_m^A$. The reference segmentation is ideally created manually by an expert who follows a detailed segmentation protocol. This process is very time consuming and not scalable. Therefore it can only be performed on a small number of images.

describes the probability of voxel i belonging to structure $k \in \{1, \dots, K\}$. It can be observed that the label map $\mathbf{L}_{\text{tgt}}$ is defined as:

$$l_i = \arg \max_k \mathbf{p}_i(k) \quad (2.1)$$

If $\mathbf{p}_i$ is multi-modal l_i can be chosen as any of the modes, for example at random (Heckemann et al., 2006).

In the following an overview over widely used atlas-based segmentation approaches and their essential building blocks, *brain extraction*, *atlas alignment*, *label fusion* and *segmentation refinement* is provided. A schematic illustration of this procedure is provided in Figure 2.2.

2.1.1 Brain extraction

A prerequisite of most anatomical segmentation approaches is the availability of a binary brain mask discriminating the actual brain from non-brain tissue such as the skull, the neck or the eyes. The calculation of this mask is, however, not trivial. Many brain extraction methods tend to produce either too restrictive or too generous masks and there is no consensus of what constitutes an ideal brain mask (Eskildsen et al., 2012).

A rough categorisation can be made by distinguishing manual approaches (Eritaia et al., 2000), semi-automatic approaches (Freeborough et al., 1997) and automatic approaches (Sandor and Leahy, 1997; Smith, 2002; Ségonne et al., 2004; Leung et al., 2011; Eskildsen et al., 2012; Manjón et al.,

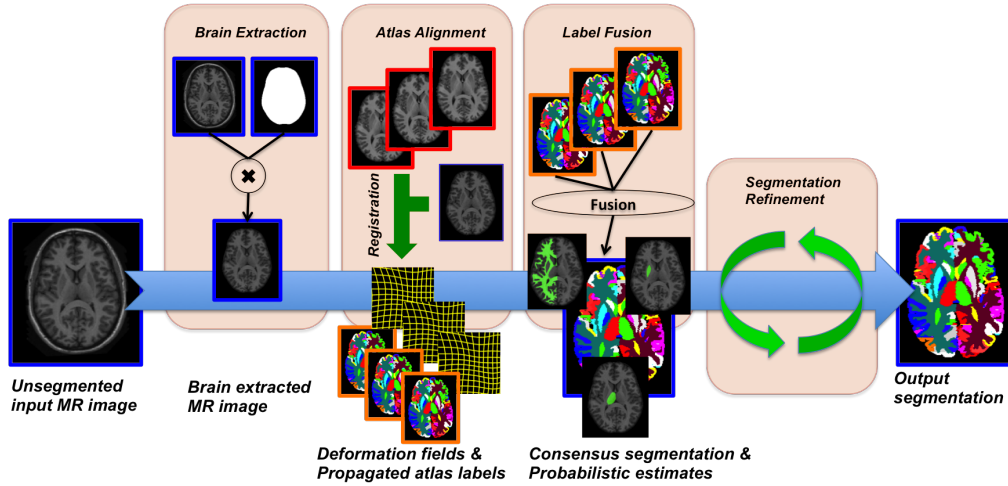


Figure 2.2: Schematic illustration of a typical segmentation pipeline with the common building blocks: brain extraction, atlas alignment, label fusion and segmentation refinement. The unsegmented target image and the estimated segmentation are outlined in blue, the atlas images are outlined in red and the propagated atlas labels in orange respectively.

2014; Heckemann et al., 2015). Approaches that require user interaction usually require substantial expertise, have poor inter- and intrarater reliability and are very time consuming, which can be prohibitive when large cohorts are to be analysed (Freeborough et al., 1997; Warfield et al., 2004; Eskildsen et al., 2012).

Skull-stripping is not the focus of this work, but a brief overview over existing automatic approaches is provided in the following.

2.1.1.1 Deformable models

Sandor and Leahy (1997) (Brain Surface Extractor (BSE)) employ an edge-detector and morphological operations to calculate a brain mask. In Smith (2002); Ségonne et al. (2004) a deformable model is automatically fit to the brain surface. Smith (2002) (Brain Extraction Tool (BET)) calculate the center of gravity and an approximate radius of the head by thresholding the image using robust minimum/maximum estimates of the image intensities. Based on this information a brain surface tessellation is initialised that iteratively evolves. The process is constrained by both intensity and smoothness terms. Ségonne et al. (2004) obtain an initial estimate of the brain mask using a watershed algorithm. The surface of this approximate but robust estimate is then employed as initialisation of an active contour model (Kass et al., 1988) that integrates both geometric and atlas-based information (Ségonne et al., 2004). A quantitative comparison of these methods (Sandor and

Leahy, 1997; Smith, 2002; Ségonne et al., 2004) can be found in Fennema-Notestine et al. (2006).

2.1.1.2 Database of extracted reference images

More recent and potentially more accurate methods for brain extraction rely on a database of brain extracted reference MRIs (Leung et al., 2011; Eskildsen et al., 2012; Manjón et al., 2014; Heckemann et al., 2015). In Leung et al. (2011) the authors nonrigidly align the reference images to the subject that is to be extracted (*cf.* Section 2.1.2) and perform label fusion (*cf.* Section 2.1.3.1). In Heckemann et al. (2015) an atlas-based approach was developed with particular focus on robustness that relies on iterative refinement of the brain mask. Other approaches such as Eskildsen et al. (2012); Manjón et al. (2014) follow a patch-based label fusion approach using linear registration only (*cf.* Section 2.1.3.4). Approaches for brain extraction tend to employ similar methodology to the techniques for the segmentation of anatomical structures. A comprehensive summary of relevant literature is provided in Leung et al. (2011) or Eskildsen et al. (2012).

In the remainder of this review it is assumed that $\mathbf{I}_{\text{tgt}}$ and the atlas images $\mathbf{I}_{\text{m}}^{\text{A}}$ are brain extracted. Skull-stripping the atlas images is often trivial as the corresponding expert label maps $\mathbf{L}_{\text{m}}^{\text{A}}$ can be binarised and employed as mask. The calculation of accurate brain masks for both AD or patients with TBI is, however, more complicated. In these cases the presence of severe atrophy (*e.g.* as observed in AD) or brain injury (*e.g.* subdural haematoma (SDH) as observed in TBI) might pose substantial challenges for an automatic brain extraction technique.

2.1.2 Atlas alignment

To exploit the labelling information encoded in the atlases most segmentation methods rely on aligning each individual atlas $\mathbf{A}_{\text{m}}$ with $\mathbf{I}_{\text{tgt}}$. This process is commonly referred to as *image registration* and usually driven by the image intensities of the corresponding images. Image registration is highly complex and an active area of research that received continuous attention over the last decades. Therefore, it is only briefly discussed here. A comprehensive overview and evaluation of established registration approaches can be found in Zitová and Flusser (2003); Klein et al. (2009); Sotiras et al. (2013).

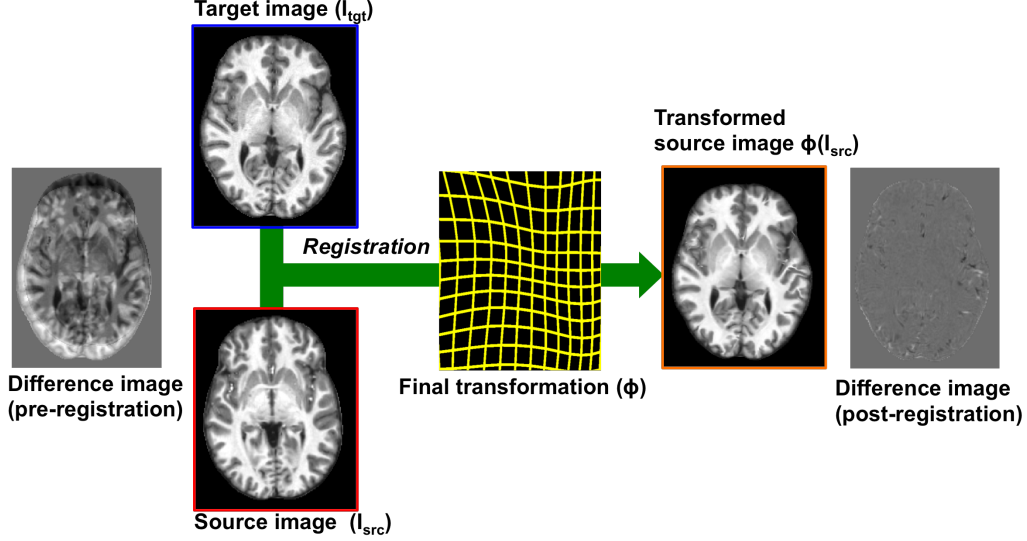


Figure 2.3: Schematic process of registering a source image $\mathbf{I}_{\text{src}}$ to a target image $\mathbf{I}_{\text{tgt}}$ with calculated transformation ϕ . The difference image of $\mathbf{I}_{\text{src}}$, $\phi(\mathbf{I}_{\text{src}})$, and $\mathbf{I}_{\text{tgt}}$ is shown before and after registration respectively. The target image is outlined in blue, the source image is outlined in red (before) and orange (after registration).

2.1.2.1 Formulation of the registration problem

In general, image registration seeks to calculate a transformation $\phi : \Omega \rightarrow \Omega$ that transforms a source (moving) image $\mathbf{I}_{\text{src}}$ such that its similarity with a target (fixed) image $\mathbf{I}_{\text{tgt}}$ is maximised. The process of image registration is illustrated in Figure 2.3. For a given similarity measure $\mathcal{S}(\cdot, \cdot) : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$ the optimisation problem can be formulated as:

$$\hat{\phi} = \arg \min_{\phi} [-\mathcal{S}(\mathbf{I}_{\text{tgt}}, \phi(\mathbf{I}_{\text{src}})) + \mathcal{R}(\phi)] \quad (2.2)$$

Here $\mathcal{R}$ is a regularisation term. In nonrigid registration, regularisation of the usually ill-posed optimisation problem is an important factor to encourage smooth and ideally even diffeomorphic transformations (Rueckert et al., 1999; Vercauteren et al., 2009). Sotiras et al. (2013) summarise important characteristics for deformable image registration algorithms that can be particularly relevant in the context of biomedical applications:

- *Inverse consistency*: The forward and backward transformations are calculated simultaneously and penalties are introduced that ensure/encourage that the forward and backward transformations are the inverse of the other. The backward transformation is calculated by exchanging the roles of $\mathbf{I}_{\text{tgt}}$ and $\mathbf{I}_{\text{src}}$.

- *Symmetry*: A symmetric registration approach can be obtained by using either a completely symmetric similarity measure S or by aligning both images simultaneously to a common target. The final transformation is then obtained by composition. Symmetry is a stronger attribute than inverse consistency.
- *Topology Preservation*: The calculated transformation is continuous, at least locally one-to-one and has a continuous inverse.
- *Diffeomorphism*: This attribute includes topology preservation but further guarantees that both the transformation and its inverse are differentiable.

Many of the widely used registration algorithms, however, fulfil none of these attributes (Joshi et al., 2004; Sotiras et al., 2013). This can be seen by the definition of the similarity measure in Equation 2.2 where only the source image is interpolated yielding an asymmetric formulation. Approaches that are both inverse consistent/symmetric and diffeomorphic were proposed for example in Joshi et al. (2004); Avants et al. (2008); Vercauteren et al. (2009). While the mentioned characteristics are desirable, especially for morphometric studies that derive further information from the transformations (*e.g.* deformation-based morphometry (DBM)), they are not a necessary requirement for accurate label propagation as evaluated in Klein et al. (2009); Landman and Warfield (2012). However, for other applications such as atlas construction they can be key to obtain unbiased estimates (Joshi et al., 2004).

2.1.2.2 Similarity measures

To quantify image similarity several measures have been proposed. Common choices for S include sum of squared differences (SSD), mutual information (Collignon et al., 1995; Viola and Wells III, 1997) (MI), normalised mutual information (Studholme et al., 1999) (NMI) or cross-correlation (CC). A definition of these similarity measures is provided in Table 2.1. The basic assumption of SSD is that intensities of perfectly matched images are identical (Roche et al., 1999). This assumption can be relaxed when using CC (affine relationship of intensities) or MI/NMI (statistical relationship) (Roche et al., 1999; Artachevarria et al., 2009). SSD and CC are often used to register images from the same modality/sequence, while the statistical measures have been successfully applied to multi-modal registration.

Table 2.1: Overview over possible measures to quantify similarity between two images $\mathbf{I}$ and $\mathbf{J}$: SSD, MI, NMI and CC. Note that SSD is a measure of dissimilarity. MI and NMI require the calculation of individual histograms (h_I, h_J) and the joint histogram ($h_{I,J}$) of the intensity images. The measures can then be calculated based on the joint, $H(\mathbf{I}, \mathbf{J})$, and the marginal entropies, $H(\mathbf{I})$ and $H(\mathbf{J})$. B denotes the number of bins of each histogram dimension which are indexed by b_i and b_j .

Measure	Definition	Value range	Similarity/ Dissimilarity
SSD($\mathbf{I}, \mathbf{J}$)	$\sum_{i=1}^N (I(i) - J(i))^2$	$\text{SSD} \geq 0$	Dissimilarity
MI($\mathbf{I}, \mathbf{J}$)	$H(\mathbf{I}) + H(\mathbf{J}) - H(\mathbf{I}, \mathbf{J})$ with: $H(\mathbf{I}) = -\sum_{b_i=1}^B h_I(b_i) \log(h_I(b_i))$ $H(\mathbf{J}) = -\sum_{b_j=1}^B h_J(b_j) \log(h_J(b_j))$ $H(\mathbf{I}, \mathbf{J}) = -\sum_{b_i=1}^B \sum_{b_j=1}^B h_{I,J}(b_i, b_j) \log(h_{I,J}(b_i, b_j))$	$\text{MI} \geq 0$	Similarity
NMI($\mathbf{I}, \mathbf{J}$)	$\frac{H(\mathbf{I}) + H(\mathbf{J})}{H(\mathbf{I}, \mathbf{J})}$	$1 \leq \text{NMI} \leq 2$	Similarity
CC($\mathbf{I}, \mathbf{J}$)	$\frac{\text{Cov}(\mathbf{I}, \mathbf{J})}{\sqrt{\text{Var}(\mathbf{I})} \sqrt{\text{Var}(\mathbf{J})}}$	$-1 \leq \text{CC} \leq 1$	Similarity*

* negative CC values indicate negative correlation which might be considered similar in some applications

2.1.2.3 Transformation models

In image registration the degrees of freedom (DOFs) of ϕ are essential as they determine the level of detail of the computed transformation. Generally, it can be distinguished between *rigid*, *affine* and *non-rigid* image alignment.

Rigid registration establishes the optimal alignment of two images by allowing translations and rotations only. Affine registration additionally compensates for transformations that include shear and scale. The recovery of both rigid and affine transformations requires the solution of a low-dimensional optimisation problem that can be solved efficiently. An accurate affine transformation is usually a crucial prerequisite for a subsequent non-rigid registration (Rueckert et al., 1999).

The aim of non-rigid registration is the recovery of a deformation field that brings both images into optimal alignment. Here, the deformation field can be described by a transformation model which is often built on cosine or B-spline basis functions (Ashburner and Friston, 1999; Rueckert et al., 1999; Andersson et al., 2007; Klein et al., 2009; Modat et al., 2010). Next to these approaches, non-parametric methods that do not model the transformation explicitly have been successfully applied. Popular examples include the so-called Demons algorithm (Thirion, 1998; Vercauteren et al., 2009) or the symmetric registration method proposed in Avants et al. (2008). The parameter space

of the resulting optimisation problem is large and can exceed millions of parameters, especially if no transformation model is used (Klein et al., 2009; Vercauteren et al., 2009).

In the context of atlas alignment for image segmentation it should be noted that non-rigid transformations can be calculated at different levels of detail. For example, when an explicit transformation model based on B-spline basis functions is used, the control point spacing and thus the number of parameters can be varied. Usually a finer control point spacing yields a more accurate registration result, however, it comes at a substantially increased computational cost.

2.1.2.4 Large deformation registration problem

The accurate registration of an atlas and unsegmented MR image can be difficult if the target image differs from the available atlases due to general anatomical variability or pathological changes. Also intensity inhomogeneities or noise can facilitate an accurate registration difficult (Vovk et al., 2007).

There have been several methods proposed to address the large deformation registration problem in adult brains in the context of AD. In Heckemann et al. (2010) the authors proposed Multi-Atlas Propagation with Enhanced Registration (Heckemann et al., 2010, 2011) (MAPER). MAPER employs automatically calculated brain tissue segmentations to guide the registration process. This allows a robust image alignment, even if the target image shows severe brain atrophy. In another approach, Wolz et al. (2010a) described an iterative approach to improve segmentation accuracy by propagating atlas labels over a learned manifold while refining intermediate segmentations based on image intensities using graph-cuts (GC) (*cf.* Section 2.1.4). It was shown by Gerber et al. (2010) that coordinates within a low-dimensional space, a nonlinear manifold, allow for a meaningful image comparison and statistical tests.

A manifold of the anatomical variation in a given dataset is also learned in Hamm et al. (2010). In this work it was shown that the problem of recovering a large deformation between two images can be simplified by solving a series of small deformation registration problems (Hamm et al., 2010). The small deformations are calculated along the shortest path between the images on the learned manifold (Hamm et al., 2010).

2.1.2.5 Registration in the presence of pathology

A particularly challenging endeavour is the alignment of the atlas images to a target image with severe pathology such as focal lesions, tumours or surgical resections. The major problem here is that the assumption of an one-to-one mapping is clearly violated. Correspondences of pathological regions are missing in the atlas images which are usually acquired from healthy subjects (Bauer et al., 2013).

Several early approaches address this problem by excluding image regions with pathology when evaluating the similarity measure (Brett et al., 2001; Stefanescu et al., 2004; Angelini et al., 2007; Andersen et al., 2010; Bauer et al., 2013). Brett et al. (2001) proposed cost-function masking based on manually obtained masks of lesions or other image artefacts. In Stefanescu et al. (2004) the authors integrate a confidence map into the registration algorithm giving zero confidence to voxels within an automatically obtained outlier region.

Alternative approaches are based on models that simulate growth, for example of a tumour, within the healthy atlas or template image (Bach Cuadra et al., 2004; Angelini et al., 2007; Zacharaki et al., 2009; Niethammer et al., 2011; Bauer et al., 2013). Bach Cuadra et al. (2004) model radial tumour growth starting from a manually placed seed voxel. Also based on a seed, Zacharaki et al. (2009) simulate tumour growth using a biomechanical model in an atlas image before registering it to an image of a tumour patient. In Niethammer et al. (2011), the authors formulated a geometric metamorphosis model to address the challenges arising in the registration of images from TBI, tumour or stroke patients.

Other approaches register and segment the images simultaneously to identify missing correspondences (Periaswamy and Farid, 2006; Chitphakdithai and Duncan, 2010; Parisot et al., 2012). Recently, Liu et al. (2014) presented a promising approach based on low-rank matrix decomposition to register multiple images of patients with TBI simultaneously to a reference image. In another recent approach a similarity measure is introduced that is based on the angle between gradient orientations (Pszczolkowski et al., 2014). This reduces bias in the similarity measure caused by image regions with intensity outliers such as pathologies.

Methods that rely on strong prior knowledge such as masks or tumour growth models are not always applicable. The manual annotation of masks is very time consuming (Andersen et al., 2010). Further, it is not obvious whether growth models can be beneficial in the presence of heterogeneous pathologies as they are often observed in patients with TBI. Pathology specific adaptations

(*e.g.* confidence maps) might be more challenging but are likely superior to simplistic masking approaches (Bauer et al., 2013). Assessing the quality or validity of a calculated transformation is difficult, especially in the presence of pathology (Angelini et al., 2007). A more comprehensive overview on image analysis in the context of brain tumours can be found in Angelini et al. (2007); Bauer et al. (2013).

In the following, ϕ_m denotes the calculated transformation from the atlas space of $\mathbf{A}_m$ to the coordinate system of $\mathbf{I}_{\text{tgt}}$ and $\mathbf{A}_m^\phi$ denotes the propagated atlas respectively. In the following, the superscript ϕ is neglected for readability. Once the atlas label maps, $\mathbf{L}_m^A$, reside in the same coordinate system as the unsegmented target image, $\mathbf{I}_{\text{tgt}}$, a consensus segmentation can be inferred using a variety of label fusion techniques (*cf.* Section 2.1.3).

2.1.3 Label fusion

2.1.3.1 Majority vote fusion

In majority vote fusion (MVF), also called “vote-rule” or “decision” fusion, each of the M propagated atlas labels $\mathbf{L}_m^A$ contributes equally to the final segmentation. A certain voxel is thus labelled according to the opinion on which the majority of the propagated atlas segmentations agree. Formally the probability of voxel i being labelled as structure k can be calculated as:

$$\mathbf{p}_i(k) = \frac{1}{M} \sum_{m=1}^M \delta(L_m^A(i), k) \quad (2.3)$$

Here, $\delta(\cdot, \cdot)$ is the Kronecker delta defined as:

$$\delta(v, w) = \begin{cases} 1 & \text{if } v = w \\ 0 & \text{otherwise} \end{cases} \quad (2.4)$$

The actual segmentation estimate $\mathbf{L}_{\text{tgt}}$ is then readily inferred through Equation 2.1.

MVF was first described by Rohlfing et al. (2004) where it was applied to the segmentation of bee brains. Heckemann et al. (2006) successfully employed MVF to segment MR brain images and presented a model that describes the basic assumption of label fusion. In this model it is

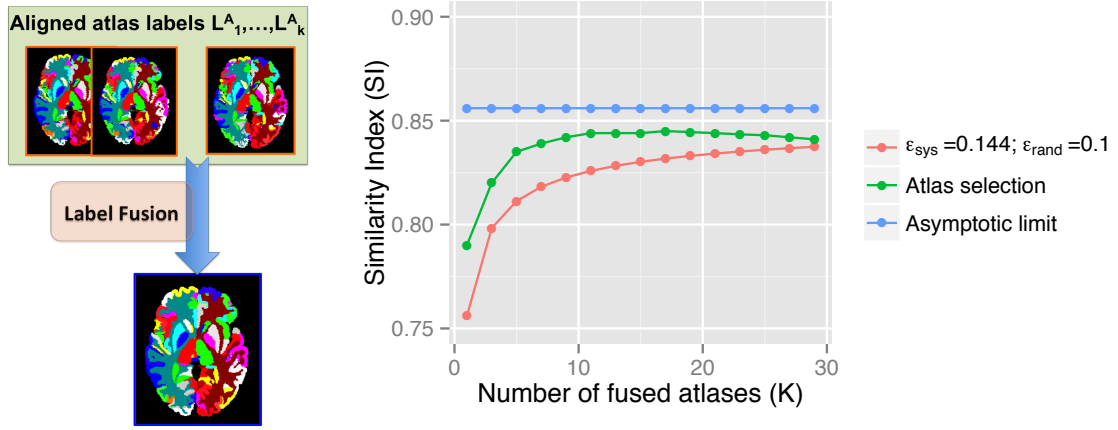


Figure 2.4: Left: Schematic process of fusing K atlas labels that have been aligned to the target image into a consensus segmentation. Right: Illustrations of the increasing segmentation accuracy with increasing number of fused atlas labels. The model described in Equation 2.5 is shown for $\epsilon_{\text{sys}} = 0.144$, $\epsilon_{\text{rand}} = 0.1$ (red) with corresponding asymptotic limit (blue). A qualitative plot indicating the increased accuracy when atlas selection is used is also shown (green). The diagram is based on and adapted from Heckemann et al. (2006); Aljabar et al. (2009).

assumed that there are two major sources of segmentation errors: 1) *Systematic errors* (ϵ_{sys}) due to deviations of the (manual) reference segmentation from the true segmentation or due to consistent registration bias. 2) *Random errors* (ϵ_{rand}) introduced due to inaccuracies in individual reference labels or image registrations (Heckemann et al., 2006). With this assumption random errors can be corrected by fusing several label estimates. However, the systematic errors asymptotically limit the segmentation accuracy as the number of fused atlases increases (Heckemann et al., 2006). In Heckemann et al. (2006) the following relation is suggested to model the segmentation accuracy SI_{model} , measured as Dice coefficient (Dice, 1945) or so-called similarity index (SI), dependent on the number of atlases K :

$$SI_{\text{model}} = 1 - \epsilon_{\text{sys}} - \frac{\epsilon_{\text{rand}}}{\sqrt{K}} \quad (2.5)$$

Figure 2.4 illustrates the schematic process and the qualitative behaviour of label fusion approaches using the model from Equation 2.5.

According to Artaechevarria et al. (2009) MVF is the most simple and most popular label fusion strategy, and does not require any a priori knowledge except for the actual segmentations that are to be fused. MVF has its origins in the more general combination of classifiers as for example Xu et al. (1992), Kittler et al. (1998) or Kuncheva (2004).

2.1.3.2 Atlas selection

If only a single atlas from a given atlas database is to be used, a carefully selected atlas that is similar to $\mathbf{I}_{\text{tgt}}$ allows a higher segmentation accuracy than choosing an atlas at random. This was confirmed in a variety of applications, such as the segmentation of CT images of the head (Han et al., 2008) or the segmentation of bee brains (Rohlfing et al., 2004) and human brains in MRI (Wu et al., 2007).

However, the fusion of multiple atlases (*e.g.* using MVF) substantially outperforms the segmentation based on a single atlas only (Rohlfing et al., 2004; Heckemann et al., 2006; Aljabar et al., 2009). Klein et al. (2007) and Aljabar et al. (2009) thus suggested to combine these approaches and perform label fusion on a *selection* of atlases from a larger database.

As selection criterion, the image similarity between the aligned atlas images and $\mathbf{I}_{\text{tgt}}$ can be calculated using a variety of similarity measures such as SSD, MI (Collignon et al., 1995; Viola and Wells III, 1997; Klein et al., 2007), NMI (Studholme et al., 1999; Wu et al., 2007), or CC (Aljabar et al., 2009) (*cf.* Table 2.1). Further, image similarity can be assessed on different alignment levels, *e.g.* after affine or nonlinear registration (Rohlfing et al., 2004; Klein et al., 2007; Wu et al., 2007), but also in a template space (Aljabar et al., 2009). When the selection is performed in a template space all atlas images can be pre-registered to this template. This reduces the computational burden dramatically as during atlas selection a single registration, aligning $\mathbf{I}_{\text{tgt}}$ to the template, suffices. Only the selected atlases are then aligned with $\mathbf{I}_{\text{tgt}}$ in a non-rigid fashion (Aljabar et al., 2009). In addition to image similarity, other criteria such as characteristics of the required deformation to align the images (Rohlfing et al., 2004) or meta-information (age, sex, etc.) (Aljabar et al., 2009) can be used.

In summary, atlas selection addresses two limitations of standard multi-atlas label fusion: First, the accurate nonlinear alignment (*cf.* Section: 2.1.2) of numerous atlases to the unsegmented image $\mathbf{I}_{\text{tgt}}$ is computationally intensive. Second, it was shown that there is an asymptotic limit of the segmentation accuracy caused by systematic and anatomical variations (Heckemann et al., 2006; Aljabar et al., 2009). Segmentation results can substantially improve up to around 20 selected atlases (Heckemann et al., 2006; Aljabar et al., 2009) (*cf.* Figure 2.4). However, using more atlases with an anatomy that increasingly differs from $\mathbf{I}_{\text{tgt}}$ might even deteriorate segmentation accuracy (Aljabar et al., 2009). Nevertheless it should be noted that the ideal number of selected atlases might be strongly dependent on the quality of the atlases, the anatomical variability of the subjects

and the ROI that is to be segmented.

2.1.3.3 Weighted vote fusion

Instead of selecting atlases, and thus potentially neglecting relevant information, another approach is to weight the contribution of each individual atlas. A detailed description and comparison of this category of label fusion can be found in Artaechevarria et al. (2009).

In globally weighted fusion (GWF) the probabilistic label estimates are calculated based on *global*, atlas-dependent voting weights w_m^g as:

$$\mathbf{p}_i(k) = \frac{\sum_{m=1}^M w_m^g \delta(L_m^A(i), k)}{\sum_{k'=1}^K \sum_{m=1}^M w_m^g \delta(L_m^A(i), k')} \quad (2.6)$$

Here, the voting weight w_m^g of each atlas is generally determined by the image similarity of the transformed atlas MRIs $\mathbf{I}_m^A$ and the subject image $\mathbf{I}_{\text{tgt}}$. Potential measures to quantify this similarity of intensity images include SSD, MI, NMI or CC (Artaechevarria et al., 2009). The calculation of global voting weights based on these measures, w_m^{gSSD} , w_m^{gCC} , w_m^{gMI} , w_m^{gNMI} , is described in Table 2.1. As the SSD increases with increasing image *dissimilarity* the voting weight is defined as the inverse of the actual SSD, $w_m^{\text{gSSD}} = \text{SSD}(\mathbf{I}_m^A, \mathbf{I}_{\text{tgt}})^{-1}$. Artaechevarria et al. (2009) further introduced a gain parameter p that potentiates the respective weights, *e.g.* $(w_m^{\text{gNMI}})^p$. However it was found that the segmentation result is usually not very sensitive towards this parameter and the optimal choice depends on the dataset (Artaechevarria et al., 2009).

Segmentation accuracy can be further increased by calculating *local* voting weights $w_m^l(i)$ within a location-specific region (Artaechevarria et al., 2009). These regions can be for example a spherical or cubical neighbourhood of the voxel under consideration. The size of the neighbourhood is a tuneable parameter, however, a neighbourhood radius of $r = 5$ was shown to yield good results (Artaechevarria et al., 2009). In locally weighted fusion (LWF) the probabilistic segmentation estimates are calculated as:

$$\mathbf{p}_i(k) = \frac{\sum_{m=1}^M w_m^l(i) \delta(L_m^A(i), k)}{\sum_{k'=1}^K \sum_{m=1}^M w_m^l(i) \delta(L_m^A(i), k')} \quad (2.7)$$

It was shown that weighted fusion overall outperforms MVF (Artaechevarria et al., 2009; Sabuncu et al., 2010). LWF is usually superior to GWF and particularly useful in image regions with high

contrast (Artaechevarria et al., 2009; Sabuncu et al., 2010). The ideal fusion strategy is thus dependent on the brain anatomy that is to be segmented. LWF with SSD often performs well (Artaechevarria et al., 2009; Wang et al., 2013b). However, when compared to NMI, SSD is more sensitive to noise in low-contrast regions and relies on the assumption of similar intensity profiles of the images (Roche et al., 1999; Artaechevarria et al., 2009).

2.1.3.4 Patch-based label fusion

Inspired by the work on non-local means filtering for image denoising (Buades et al., 2005; Coupé et al., 2008), Coupé et al. (2010) proposed a *patch-based* approach to address the problem of label fusion. Patch-based label fusion was subsequently described thoroughly in Rousseau et al. (2011) and Coupé et al. (2011). In patch-based label fusion techniques, image patches of the unsegmented image are locally compared to image patches in the atlas images. Usually patches are defined as 3D volumes of a given diameter. For a fixed patch $P(\mathbf{I}_{\text{tgt}}, i)$ in $\mathbf{I}_{\text{tgt}}$ with centre voxel i numerous patches in $\mathbf{I}_{\text{m}}^{\text{A}}$ are taken into account. Specifically, all patches $P(\mathbf{I}_{\text{m}}^{\text{A}}, j)$ with centre j with the spatial constraint that j lies within a defined neighbourhood $\mathcal{N}_i$ of i are considered. Both the size of the 3D patches (patch size) and the size of the neighbourhood (search window size) are parameters that need to be chosen. An illustration of this process is provided in Figure 2.5. The exact label fusion procedure can be formalised as:

$$\mathbf{p}_i(k) = \frac{\sum_{m=1}^M \sum_{j \in \mathcal{N}_i} w_m^{\square}(i, j) \delta(L_m^{\text{A}}(j), k)}{\sum_{k'=1}^K \sum_{m=1}^M \sum_{j \in \mathcal{N}_i} w_m^{\square}(i, j) \delta(L_m^{\text{A}}(j), k')} \quad (2.8)$$

Here $w_m^{\square}(i, j)$ is a weight that is calculated based on the similarity of the patch $P(\mathbf{I}_{\text{tgt}}, i)$, centred at voxel i in $\mathbf{I}_{\text{tgt}}$, and the patch $P(\mathbf{I}_{\text{m}}^{\text{A}}, j)$, centred at voxel j in atlas $\mathbf{I}_{\text{m}}^{\text{A}}$. A common definition (Buades et al., 2005; Coupé et al., 2011; Rousseau et al., 2011) of this similarity measure is:

$$w_m^{\square}(i, j) = \exp \frac{-\|P(\mathbf{I}_{\text{tgt}}, i) - P(\mathbf{I}_{\text{m}}^{\text{A}}, j)\|^2}{Nh} \quad (2.9)$$

Here $\|\cdot\|$ is the L_2 -norm and N is the number of voxels in the patches. Further, h is a decay parameter that requires tuning (Coupé et al., 2011). If $h \rightarrow \infty$ all patches are weighted equally, if $h \rightarrow 0$ only the best matching patch is selected. Potentially, $w_m^{\square}(i, j)$ might be set to zero below a given threshold (Coupé et al., 2011).

The main motivation for patch-based label fusion is to remove the computationally expensive

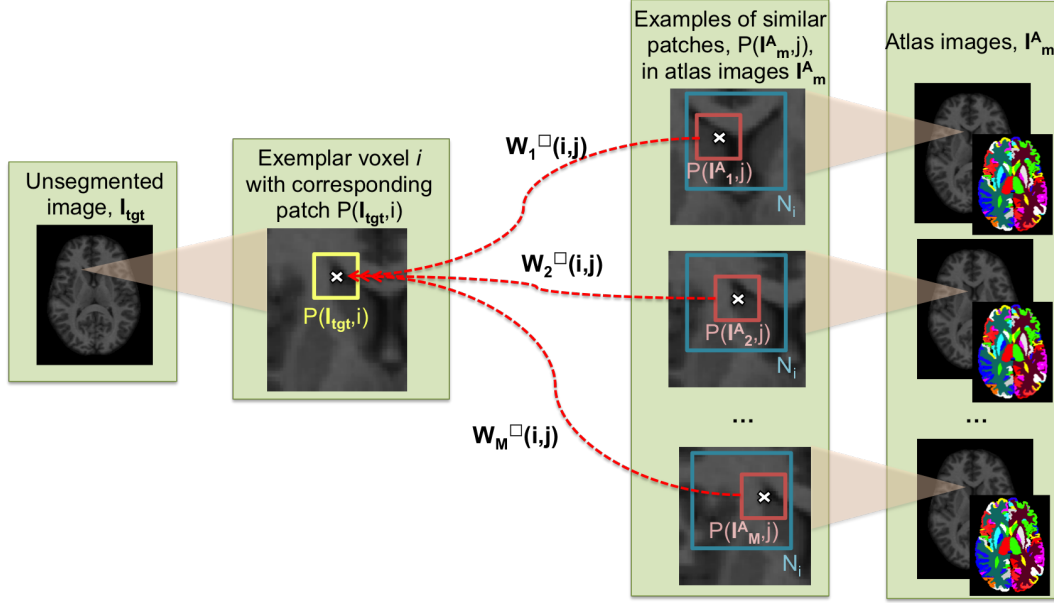


Figure 2.5: Illustration of the weight calculation when segmenting a voxel i . The weights $w_m^\square(i, j)$ are determined between the fixed reference patch $P(I_{tgt}, i)$ (yellow) and the atlas patches $P(I_m^A, j)$ (red). For each atlas image several atlas patches that are shifted within a local search neighbourhood (cyan) are considered. With a weighted voting approach a probabilistic segmentation (cf. Equation 2.8) is obtained. Illustration adapted and modified from Coupé et al. (2011); Rousseau et al. (2011).

requirement of nonrigidly aligning the reference atlases (cf. Section 2.1.2). The patch-based formulation further relaxes the assumption of an one-to-one mapping between the atlas images I_m^A and the unsegmented image I_{tgt} (Rousseau et al., 2011). An approximate alignment, using for example affine registration (Coupé et al., 2011), is sufficient to establish reasonable patch correspondences. Nevertheless, it must be noted that nonrigid alignment can further improve patch-based segmentation as it allows the more rigorous incorporation of topological constraints (Coupé et al., 2011; Rousseau et al., 2011).

In patch-based label fusion one can discriminate *point-wise* and *multipoint* estimators (Rousseau et al., 2011). In Equation 2.8 a point-wise estimator is described, in which the label probability of each individual voxel (or point) i is estimated individually. In contrast to this, multipoint estimators calculate at each individual voxel a label patch estimate (Rousseau et al., 2011). Specifically, instead of considering only the centre label $L_m^A(j)$ (cf. Equation 2.8) label information from the whole patch centred at j is fused into a patch estimate. While point-wise estimators are commonly used and computationally more efficient, it was shown that multipoint estimators can potentially increase segmentation accuracy (Rousseau et al., 2011). More details on multipoint estimators can be found

in Katkovnik et al. (2010); Rousseau et al. (2011).

Recent advances in patch-based segmentation include the application of sparsity techniques (Tong et al., 2013) or the efficient search of patch correspondences using the PatchMatch algorithm (Barnes et al., 2009a; Shi et al., 2014; Ta et al., 2014). Next to image segmentation, it was proposed to employ patch-based methods to grade subjects according to clinical variables (Coupé et al., 2012a,b). In this patch-based grading approach a meaningful clinical label, *e.g.* AD or healthy control (HC), is associated with each patch of a reference database. Instead of fusing anatomical labels $L_m^A(j)$ these clinical labels are fused in Equation 2.8 within a region. It was shown that the resulting grading value allows the accurate discrimination of different AD disease states (Coupé et al., 2012a,b).

2.1.3.5 Joint label fusion

The label fusion techniques that were presented so far treat each atlas independent from the others for the calculation of voting weights. Thus, none of these approaches takes into account that similar or, in general, correlated atlases might produce similar segmentation errors (Wang et al., 2013b). To address this limitation Wang et al. (2013b) introduced an approach called joint label fusion (JLF) in which the joint probability distribution of two atlases producing the same labelling error is explicitly modelled.

In particular the authors aim to minimise the expected error between the true, yet unknown segmentation of $\mathbf{I}_{\text{tgt}}$ and the consensus segmentation estimate $\mathbf{L}_{\text{tgt}}$. This optimisation problem can be solved once dependency matrices $\mathbf{M}_i$ of size $M \times M$ are established that model the likelihood that a pair of the M available atlases make an identical error at voxel i . In Wang et al. (2013b), the authors calculate the correlation of the intensity similarity of two atlases with respect to $\mathbf{I}_{\text{tgt}}$. A local estimation of $\mathbf{M}_i$ is obtained by performing the similarity computation in a local neighbourhood of the corresponding voxel i . More details on JLF can be found in Wang et al. (2013b).

Many approaches ranking among the top performing methods in the literature are based on JLF strategies that incorporate the non-local patch-based concept. This was confirmed for the segmentation of various anatomies in recent Segmentation Challenges (Landman and Warfield, 2012; Asman et al., 2013a). For example, JLF performs very well for the segmentation of the brain into distinct ROIs (Landman and Warfield, 2012; Asman et al., 2013a), but also for the segmentation of the

myocardium (Asman et al., 2013a; Bai et al., 2015).

2.1.3.6 Statistical label fusion

With origins in the popular Simultaneous Truth And Performance Level Evaluation (Warfield et al., 2004) (STAPLE) algorithm there is another category of statistical label fusion approaches. Given several *manual* label sets, STAPLE was proposed to simultaneously estimate the performance level parameters of each expert rater and the most probable ground truth. The STAPLE algorithm computes a probabilistic estimate of the true segmentation (Warfield et al., 2004). For the sake of consistency with existing literature the commonly used notation is employed to outline the STAPLE algorithm. This notation is slightly different in comparison to the rest of this section. This also accommodates the fact that STAPLE was originally described for the estimation of a ground truth from several expert reference segmentations rather than for the fusion of automatic label estimates. The following description is adapted from Warfield et al. (2004), where more details about the exact implementation of STAPLE can be found.

Let $\mathbf{D}$ be a $N \times M$ matrix that describes the decisions of M raters (or atlases) at N voxels. Let $\mathbf{T}$ be a vector of size N indicating the true but unknown segmentation. For K possible labels $\mathbf{D}$ and $\mathbf{T}$ are indexed as $D_{im} \in \{1, \dots, K\}$ and $T_i \in \{1, \dots, K\}$. The goal of the STAPLE algorithm is to estimate a performance tensor $\hat{\theta}$ of size $M \times K \times K$ that maximises the log-likelihood of observing the complete data $(\mathbf{D}, \mathbf{T})$. Here, element $\theta_{ms's}$ quantifies the probability that rater m decides on label s' given that the true label is s . For a probability mass function $f(\mathbf{D}, \mathbf{T}|\theta)$ that describes the probability of observing the complete data, $\hat{\theta}$ can be calculated by solving:

$$\hat{\theta} = \arg \max_{\theta} \ln f(\mathbf{D}, \mathbf{T}|\theta) \quad (2.10)$$

The solution of this problem would be trivial if the true segmentation $\mathbf{T}$ was known. As $\mathbf{T}$ is however unknown Warfield et al. (2004) suggested to employ the expectation-maximisation (EM) algorithm to solve Equation 2.10. In this approach, a ground truth $\mathbf{T}$ is estimated in the Expectation-step and the ideal performance parameters $\hat{\theta}$ found in the Maximisation-step. This is repeated iteratively until the model converges. Convergence to a local maximum of Equation 2.10 is guaranteed (Warfield et al., 2004). There are different strategies to initialise θ in the first iteration (Warfield et al., 2004). Furthermore, spatial prior information can be incorporated through probabilistic priors or smooth-

ness constraints based on Markov random fields (MRFs) (Warfield et al., 2004).

Next to the basic STAPLE algorithm outlined above, there have been several successful attempts to incorporate both image intensity information (Cardoso et al., 2013a; Asman and Landman, 2013) and the concept of non-local patch-based approaches into the STAPLE framework (Asman and Landman, 2013). Many other extensions include formulations that allow the application to data with missing labels (Landman et al., 2012), local estimation of rater performances (Commowick et al., 2012; Asman and Landman, 2012) or the fusion of probabilistic decisions (Akhondi-Asl and Warfield, 2013).

2.1.4 Refinement of image segmentations

The segmentation estimates obtained by multi-atlas label fusion, as described in the previous Section 2.1.3, can be further improved. In the following two successful strategies to the refinement of image segmentations are presented: The refinement of segmentations based on image intensities and the refinement based on classifiers that were trained to correct systematic bias of a segmentation method.

2.1.4.1 Intensity-based refinement

The probabilistic segmentation estimates obtained by automatic label fusion techniques (*cf.* Section 2.1.3) can be further refined based on the image intensities of $\mathbf{I}_{\text{tgt}}$. It was shown that modifying image segmentations based on the actual image intensities can substantially improve segmentation results (van der Lijn et al., 2008; Wolz et al., 2009; Lötjönen et al., 2010; Ledig et al., 2015).

Using Bayes' theorem the refinement task is often modelled as calculating the maximum a posteriori (MAP) estimate as:

$$\mathbf{L}_{\text{tgt}} = \arg \max_{\mathbf{L}} p(\mathbf{L} | \mathbf{I}_{\text{tgt}}) = \arg \max_{\mathbf{L}} \frac{p(\mathbf{I}_{\text{tgt}} | \mathbf{L}) p(\mathbf{L})}{p(\mathbf{I}_{\text{tgt}})} = \arg \max_{\mathbf{L}} p(\mathbf{I}_{\text{tgt}} | \mathbf{L}) p(\mathbf{L}) \quad (2.11)$$

Assuming voxel-wise independence the image likelihood, $p(\mathbf{I}_{\text{tgt}} | \mathbf{L})$ can be calculated as $\prod_{i=1}^N p(y_i | l_i)$ where $p(y_i | l_i)$ is given by a predefined intensity model. The probability of a segmentation $p(\mathbf{L})$ is often modelled based on spatial prior knowledge and smoothness constraints between adjacent labels. Since the target image is fixed, the optimisation problem is independent of $p(\mathbf{I}_{\text{tgt}})$.

Two popular approaches to solve the optimisation problem stated in Equation 2.11 are based on

graph-cuts (GC) (Greig et al., 1989; Boykov et al., 2001; van der Lijn et al., 2008; Wolz et al., 2009) or the expectation-maximisation (EM) algorithm (Van Leemput et al., 1999; Lötjönen et al., 2010; Ledig et al., 2015). A more detailed introduction of the EM approach will be provided in Chapter 3.

The optimisation problem in Equation 2.11 can be rewritten (Greig et al., 1989; van der Lijn et al., 2008) as a minimisation problem by taking the negative logarithm as:

$$\mathbf{L}_{\text{tgt}} = \arg \min_{\mathbf{L}} (-\ln p(\mathbf{I}_{\text{tgt}}|\mathbf{L}) - \ln p(\mathbf{L})) = \arg \min_{\mathbf{L}} (E_{\text{intensity}}(\mathbf{L}) + E_{\text{prior}}(\mathbf{L})) \quad (2.12)$$

This amended formulation allows the definition of a network graph for which a minimum cut can be calculated based on the Ford-Fulkerson max-flow/min-cut algorithm (Greig et al., 1989). In Equation 2.12 the intensity (or data) term, $E_{\text{intensity}}(\mathbf{L}) = -\sum_{i=1}^N \ln p(y_i|l_i)$, quantifies the agreement of the image data $\mathbf{I}_{\text{tgt}}$ with the intensity model. A common choice for the intensity model of $p(y_i|l_i)$ is a Gaussian probability distribution with label-specific parameters $(\mu_{l_i}, \sigma_{l_i})$ (van der Lijn et al., 2008; Wolz et al., 2010b). The second term, $E_{\text{prior}}(\mathbf{L})$, incorporates both spatial prior information provided through the probabilistic segmentation estimates (*e.g.* obtained through label fusion) and smoothness constraints which are often modelled through MRFs (van der Lijn et al., 2008; Wolz et al., 2010b). The cost function that is minimised by GC can then be summarised as:

$$E(\mathbf{L}) = \underbrace{-\sum_{i=1}^N \ln p(y_i|l_i)}_{E_{\text{intensity}}} - \underbrace{\sum_{i=1}^N \ln p(l_i)}_{E_{\text{spatial prior}}} + \underbrace{\sum_{i=1}^N \sum_{j \in \mathcal{N}_i} G(l_i, l_j)}_{E_{\text{smoothness prior}}} \quad (2.13)$$

Here $p(l_i)$ is the prior probability that voxel i has label l_i . It can be calculated using label fusion (*cf.* Section 2.1.3) so that $p(l_i) = \mathbf{p}_i(l_i)$. Furthermore $\mathcal{N}_i$ is the set of voxels neighbouring voxel i and $G(l_i, l_j)$ penalises non-smooth label configurations ($G(l_i, l_j) > 0$ for $l_i \neq l_j$). The individual energy terms can be multiplied with weighting factors to control their individual contribution (Song et al., 2006; van der Lijn et al., 2008; Wolz et al., 2010b).

GC optimisation was first introduced for binary segmentation problems (Greig et al., 1989; Boykov et al., 2001; Boykov and Kolmogorov, 2004) for which a global optimum can be found. However, the optimisation problem can also be formulated as a multiway cut problem and employed for the segmentation of multiple labels (Boykov et al., 2001; Boykov and Kolmogorov, 2004; Song et al., 2006). For multiway cut problems, Boykov et al. (2001) presented an algorithm that efficiently calculates approximate solutions of the global minima with optimality bounds. GC

optimisation typically assigns one strict label to each voxel.

2.1.4.2 Learning-based refinement

Learning-based refinement is based on the assumption that a substantial fraction of falsely labelled image voxels is due to systematic bias of the employed segmentation method (Wang et al., 2011a). Systematic bias often originates in different definitions of manual segmentation protocols or inaccurate translation of the manual protocol to the automatic method (Wang et al., 2011a). The assumption made by Wang et al. (2011a) is similar to the one made by Heckemann et al. (2006): In the work by Heckemann et al. (2006) the authors described segmentation inaccuracies in the context of multi-atlas label propagation as a combination of random variability of propagated labels and systematic errors.

The main contribution of Wang et al. (2011a) is the proposal to explicitly learn systematic segmentation errors with respect to reference segmentations using machine learning. Specifically, features based on image intensities and label context are extracted in a local neighbourhood of each voxel and combined with spatial information. These features are subsequently employed to train error detection and error correction classifiers using AdaBoost (Freund and Schapire, 1995) with respect to reference segmentations (Wang et al., 2011a). These classifiers have been shown to allow a significant reduction of systematic bias in the context of many applications. In the original work, Wang et al. (2011a) presented substantial improvements when the method is employed to correct automatically calculated hippocampus segmentations, brain masks and brain tissue segmentations.

A schematic illustration of the method is provided in Figure 2.6. More details on learning-based refinement can be found in Wang et al. (2011a).

2.1.5 Remarks

Recently several segmentation challenges were held in conjunction with international conferences. The methods with particularly good performance for segmenting brain data were often based on JLF (*cf.* Section 2.1.3.5) and corrected for systematic segmentation bias (*cf.* Section 2.1.4.2) (Landman and Warfield, 2012; Asman et al., 2013a).

The careful design and implementation of each individually outlined building block, *brain extraction, atlas alignment, label fusion, label refinement*, is essential and critical to realise a segmentation

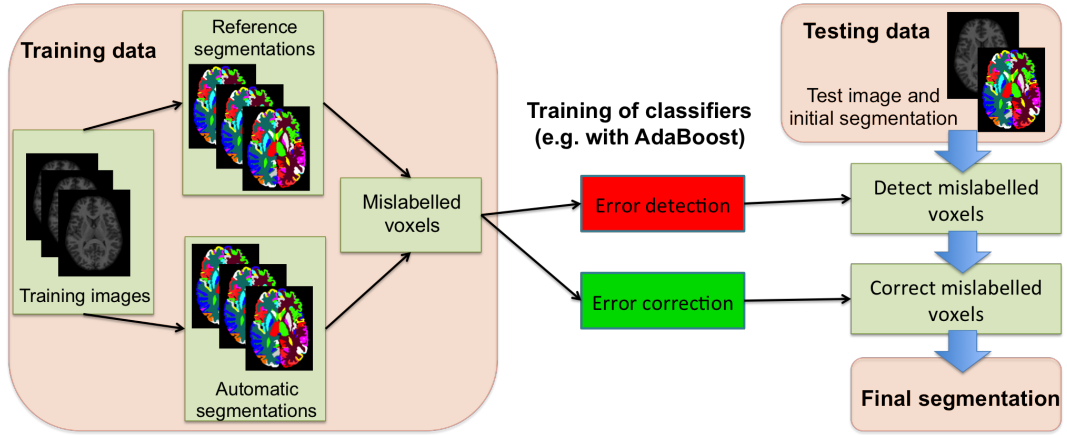


Figure 2.6: Schematic process of training classifiers for the detection and correction of systematic segmentation bias. This illustration corresponds to the variant named *explicit bias correction* in Wang et al. (2011a). Note that for binary segmentation tasks the error detection and error correction classifiers are identical. Illustration adapted and modified from Wang et al. (2011a).

approach that is both accurate and robust. Failure or inaccuracies within a single of these 'modules' is likely to negatively impact the final segmentation result. It is tempting to consider each of these blocks separately to reduce the complexity of the task. This further allows the independent implementation, tuning and validation of each module. It should be mentioned, however, that there is some evidence that more complex approaches that model and solve several building blocks jointly (*e.g.* registration and segmentation) can increase accuracy (Ashburner and Friston, 2005). Nevertheless, in practice the theoretical advantages of these joint models might be outweighed by the above-mentioned merits of a modular approach.

2.2 Brain atlases from structural MRI

Atlases of the human brain can be considered as maps that associate a spatial position within the brain with structural or functional information such as a specific anatomical structure. Early brain atlases were created based on a single subject or a small number of selected subjects (Mazziotta et al., 1995). A widely used example is the Talairach space created from postmortem specimen of a single subject (Talairach and Tournoux, 1988; Mazziotta et al., 1995). While these early atlases provided great insight into brain anatomy they were limited in their ability to represent the large intersubject variability of the human brain.

To address this problem, probabilistic population atlases were created from a population of brain images that were aligned with and segmented in a common reference space, *e.g.* in Shattuck et al. (2008). Popular stereotaxic spaces are given by the MNI-305¹ (Evans et al., 1993) and the MNI-152² (Mazziotta et al., 1995) brain templates, which are available through the Montreal Neurological Institute (<http://www.mni.mcgill.ca/>, last accessed 29 June 2015) (MNI). Both templates were created by averaging the contributing MR intensity images (305, and 152 respectively) in a common space after linear image registration (Evans et al., 1993; Mazziotta et al., 1995). The MNI-152 template was built by the International Consortium for Brain Mapping (ICBM) from images acquired with better contrast and at a higher resolution than the images on which the MNI-305 template is based². The stereotaxic spaces described by these template images can serve as target to align individual atlases of an atlas database (Hammers et al., 2003; Shattuck et al., 2008), such as the ones described in this section. The propagated atlas labels can then be averaged using label fusion techniques (*cf.* Section 2.1.3) to create a probabilistic brain atlas. The characteristics of the resulting probabilistic atlas, such as spatial variability of individual structures, directly depends on the chosen template space (*e.g.* MNI-305, MNI-152) and the registration approach (*e.g.* affine or nonrigid transformation model). As a consequence, a variety of application-specific probabilistic atlases can be created from atlas databases such as the ones described in the following. A broader description and history of atlases of the human brain can be found in Mazziotta et al. (1995); Shattuck et al. (2008).

In this thesis the focus is on non-probabilistic brain atlases that can be aligned with a specific target image. For image segmentation these individual brain atlases need to be aligned with the unsegmented image. This approach is thus computationally more expensive than the alignment with a single (probabilistic) atlas only. However, accuracy of the estimated target segmentation can be substantially improved based on *subject-specific* atlases created using label fusion (*cf.* Section 2.1.3).

In the following, five publicly available non-probabilistic brain atlases based on T1w MR images are introduced. The reviewed brain atlases were selected because they are widely used, publicly available, and contain manually annotated anatomical labels for both cortical and non-cortical structures. An overview over key characteristics of these atlases is given in Table 2.2.

¹<http://www.bic.mni.mcgill.ca/ServicesAtlases/MNI305>, last accessed 29 June 2015.

²<http://www.bic.mni.mcgill.ca/ServicesAtlases/ICBM152Lin>, last accessed 29 June 2015.

However, next to this selection of brain atlases there are alternatives. For example, Klein et al. (2009) employed next to the IBSR18 (*cf.* Section 2.2.2) and LBPA40 (*cf.* Section 2.2.4) atlases, the CUMC12 (Columbia University Medical Center) and MGH10 (Massachusetts General Hospital) atlases to evaluate 14 established registration algorithms. Furthermore, Klein and Tourville (2012) recently created a detailed protocol to label human cortices. Based on this protocol MR images from 101 healthy participants, with origins in nine different publicly available datasets, were segmented. The expert labels for all images³ were obtained by manually editing label maps that were calculated using FreeSurfer (<http://surfer.nmr.mgh.harvard.edu/>, last accessed 29 June 2015, Fischl and Dale (2000); Fischl et al. (2002, 2004)) (FreeSurfer) to correspond with the protocol.

2.2.1 AAL atlas

The *Automated Anatomical Labeling (AAL) brain atlas*⁴ (Tzourio-Mazoyer et al., 2002) distinguishes 116 ROIs and is based on a single MR template image with a resolution of $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$. The template⁵ was obtained by averaging 27 T1w MR brain images of a male individual to increase the signal to noise ratio (SNR) (Holmes et al., 1998; Tzourio-Mazoyer et al., 2002). The MR template was then segmented manually into distinct ROIs based on two-dimensional (2D) axial slices. The ROIs were outlined only on every second axial slice (Tzourio-Mazoyer et al., 2002). The atlas segmentations available for download⁴ are, compared to the MR template, of a lower resolution of $2\text{ mm} \times 2\text{ mm} \times 2\text{ mm}$. The AAL atlas was constructed with a focus on measuring activation patterns within the defined ROIs in functional imaging. To avoid missing some structure-related activity due to a too conservative definition of ROIs and to account for intersubject variability, structures were outlined quite generously, even beyond the GM boundary (Tzourio-Mazoyer et al., 2002).

In total the template is divided into 116 anatomical structures including the cortical gyri but also subcortical structures such as hippocampus, amygdala, caudate nucleus, putamen, pallidum, thalamus and 26 labels subdividing the cerebellum (Tzourio-Mazoyer et al., 2002). Structures in the left and right hemisphere are treated separately. Ventricles were not segmented. The visualisation of the atlas in Figure 2.7 clearly shows the generous definition of ROIs including more than one tissue type.

³available at <http://mindboggle.info/>, last accessed 29 June 2015.

⁴<http://www.gin.cnrs.fr/spip.php?article217>, last accessed 29 June 2015.

⁵<http://www.bic.mni.mcgill.ca/ServicesAtlases/Colin27Highres>, last accessed 29 June 2015.

Table 2.2: Overview of key characteristics of the five selected brain atlases.

Atlas	N	male/ female	age (mean [min; max])	ROIs	C1	C2	Literature	URL
AAL	1	1/0	young adult	116	no	no	Tzourio-Mazoyer et al. (2002)	AAL-1/2
IBSR18	18	4/14	38 [7; 71]	34	yes	yes	-	IBSR-1
Hammers	30	15/15	31* [20; 54]	83	no	no	Hammers et al. (2003, 2007); Gousias et al. (2008)	Hammers-1
LPBA40	40	20/20	29.2 [19; 39]	56	no	no	Shattuck et al. (2008)	LPBA-1
NMM	30 ⁺	10/20	34.3 [18; 90]	134	yes	yes	Marcus et al. (2007); Worth and Tourville (2013)	NMM-1/2/3

*: median; ⁺: atlas consists of 35 images as five subjects were scanned twice

C1: cortical ROIs contain only grey matter tissue.

C2: all brain tissue including ventricles is annotated. *E.g.* AAL, Hammers or LPBA atlases contain not annotated WM.

AAL-1: <http://www.gin.cnrs.fr/spip.php?article217>

AAL-2: <http://www.bic.mni.mcgill.ca/ServicesAtlases/Colin27Highres>

IBSR-1: <http://www.nitrc.org/projects/ibsr>

Hammers-1: <http://wp.doc.ic.ac.uk/brain-development/brain-atlases/>

LPBA-1: http://www.loni.usc.edu/atlas/Atlas_Detail.php?atlas_id=12

NMM-1: <http://Neuromorphometrics.com/>

NMM-2: <http://www.cma.mgh.harvard.edu/manuals/segmentation/>

NMM-3: <http://www.braincolor.org>

(all last accessed 13 November 2015)

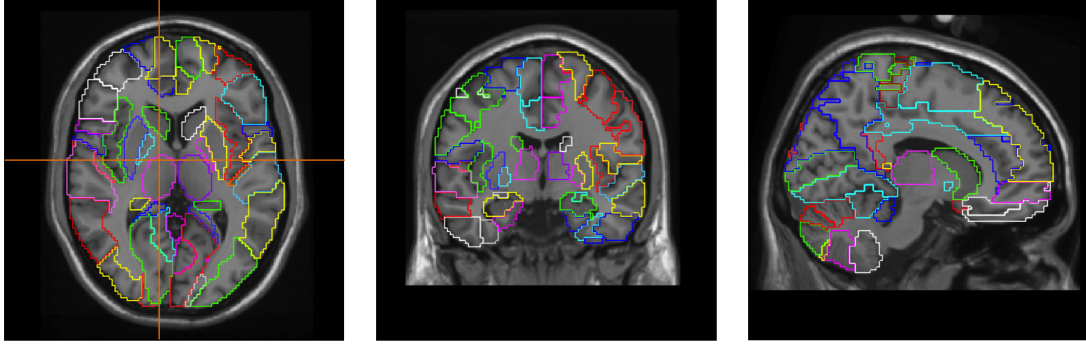


Figure 2.7: The AAL atlas in axial (left), coronal (middle) and sagittal (right) view plane. T1w MR template with overlaid segmentation contours in a colour scheme that provides a good contrast between adjacent labels. The cross-hair in the axial view indicates the position of the illustrated coronal and sagittal slice.

2.2.2 IBSR18 atlas

The *Internet Brain Segmentation Repository (IBSR)* V2.0⁶ atlas consists of T1w MR images acquired from 18 subjects (14 male / 4 female, mean age [min; max] of 14 subjects: 38 years [7; 71]⁷). The slice thickness of all images is 1.5 mm. However, the in-plane resolution varies from $0.8371 \text{ mm} \times 0.8371 \text{ mm}$ to $1 \text{ mm} \times 1 \text{ mm}$. All images were positionally normalised by rotating the images into the Talairach orientation. Expert segmentations of 34 anatomical structures⁸ were created manually.

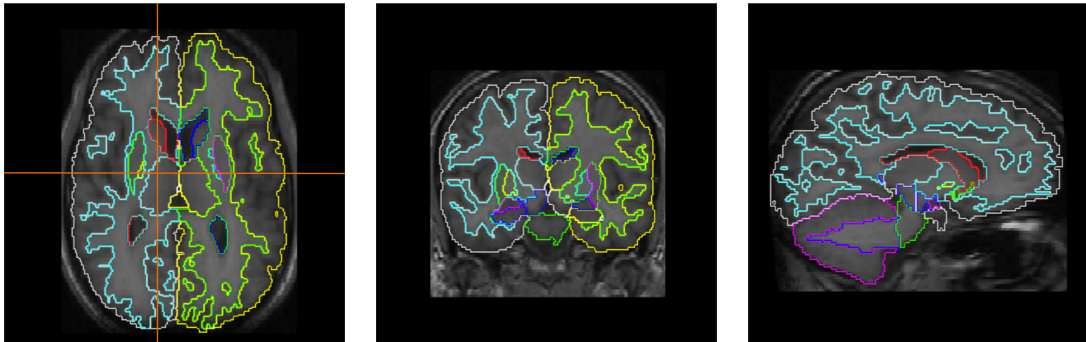


Figure 2.8: An example of a subject (IBSR01, male, 37 years) of the IBSR atlas in axial (left), coronal (middle) and sagittal (right) view plane. T1w MR template with overlaid segmentation contours.

⁶<http://www.nitrc.org/projects/ibsr>, last accessed 29 June 2015.

⁷No exact age provided for four subjects.

⁸The readme file describes 39 anatomical structures. In contrast to this description labels for cerebral exterior (left/right), cerebellum exterior (left/right) and amygdala anterior (left/right) are missing while CSF is an additional label not mentioned in the description ($39 - 6 + 1 = 34$).

The annotated structures include for example hippocampus, amygdala, caudate, pallidum, putamen, thalamus, inferior lateral ventricle, lateral ventricle. The cerebellum is split into cerebellum cortex and cerebellum WM. Structures in the left and right hemisphere are discriminated. Cortical annotations include cortical GM only. An example of a subject of the IBSR atlas is illustrated in Figure 2.8.

2.2.3 Hammers atlas

The *Hammers brain atlas*⁹ (Hammers et al., 2003, 2007; Gousias et al., 2008) consists of 30 T1w MR images with corresponding segmentations distinguishing 83 ROIs. The images were acquired from young healthy adults (15 male/female, median age [min; max] of all subjects: 31 years [20; 54]) and resliced to an isotropic resolution of 0.9375 mm $\times$ 0.9375 mm $\times$ 0.9375 mm (Hammers et al., 2003; Gousias et al., 2008). All images were acquired at 1.5T with the same scanner (Gousias et al., 2008). Originally, Hammers et al. (2003) annotated 49 distinct anatomical regions of 20 subjects. Delineations were done in native space. This work was subsequently extended to 30 subjects and by further subdividing the 49 regions into a total of 83 ROIs (Hammers et al., 2007; Gousias et al., 2008). Following a well defined protocol (Hammers et al., 2003), the MRIs were manually annotated on 2D slices in structure-specific orientations. Subsequent control of difficult cases was conducted by a trained specialist (Hammers et al., 2003). An intrarater reliability study indicates good reliability (Hammers et al., 2007). Hammers et al. (2003) further created a probabilistic brain atlas by aligning the individual segmentations to the MNI-152 template space using Statistical Parametric Mapping (<http://www.fil.ion.ucl.ac.uk/spm/>, last accessed 29 June 2015, Ashburner and Friston (1997, 1999, 2005)) (SPM).

Each of the 30 MRIs is divided into 83 ROIs subdividing the cortex, ventricles and non-cortical structures such as the hippocampus, amygdala, caudate nucleus, pallidum, putamen, thalamus. There is no finer separation within the cerebellum. Structures in the left and right hemisphere form separate ROIs. GM and WM are combined in one common label for most cortical ROIs (Hammers et al., 2003). An example of a subject of the Hammers atlas is illustrated in Figure 2.9.

⁹<http://wp.doc.ic.ac.uk/brain-development/brain-atlases/>, last accessed 13 November 2015.

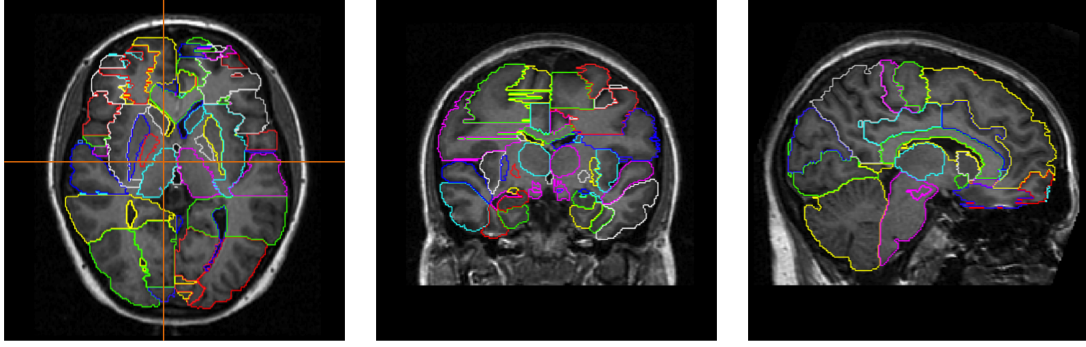


Figure 2.9: An example of a subject (a01) of the Hammers atlas in axial (left), coronal (middle) and sagittal (right) view plane. T1w MR template with overlaid segmentation contours.

2.2.4 LPBA40 atlas

The *LONI Probabilistic Brain Atlas (LPBA)*¹⁰ consists of T1w MR images acquired from 40 healthy volunteers (20 male/female, mean age [min; max] of all subjects: 29.2 years [19; 39]) (Shattuck et al., 2008). All images were acquired with the same scanner at 1.5T and subsequently *rigidly* aligned to the MNI-305¹ (page 77) (Evans et al., 1993) template space. Images were further resampled to an isotropic resolution of $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$ (Shattuck et al., 2008). Manual delineation of the images into 56 distinct ROIs was then performed in template space following a detailed protocol (Shattuck et al., 2008). Rater-reliability was assessed. Several versions of a probabilistic atlas were produced by employing different techniques (*e.g.* SPM) to *nonrigidly* align the individual segmentations to the MNI-305 template space (Shattuck et al., 2008). However, the individual 40 annotated non-probabilistic atlases are available.

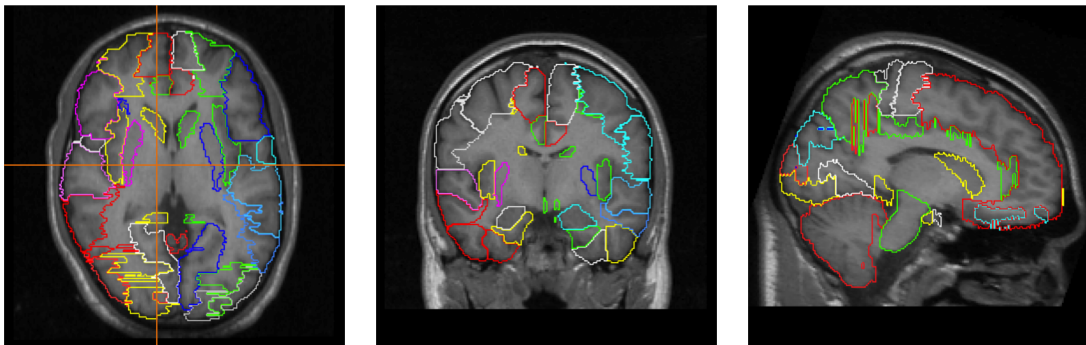


Figure 2.10: An example of a subject (S01) of the LPBA atlas in axial (left), coronal (middle) and sagittal (right) view plane. T1w MR template with overlaid segmentation contours.

¹⁰http://www.loni.usc.edu/atlas/Atlas_Detail.php?atlas_id=12, last accessed 29 June 2015.

The atlas distinguishes hippocampus, caudate, putamen, cerebellum and brainstem but mainly parcellates cortical ROIs (Shattuck et al., 2008). Amygdala, pallidum and thalamus are absent. There is no finer separation within the cerebellum. Structures in the left and right hemisphere form individual ROIs. Cortical structures potentially contain WM that lies between sulci or in the vicinity of cortical GM (Shattuck et al., 2008). An example of a subject of the LPBA atlas is illustrated in Figure 2.10.

2.2.5 NMM atlas

The *Neuromorphometrics (NMM) brain atlas*¹¹ consists of a continuously growing number of manually annotated T1w MR brain images (Worth and Tourville, 2013). The currently available annotated images can be obtained with an academic license for an annual fee. Recently, a subset of 35 images was used as 'gold standard' and made freely available in the course of the "MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling" (Landman and Warfield, 2012). In the following a description of these 35 images is provided.

The 35 images were taken from the Open Access Series of Imaging Studies (OASIS) database and acquired from 30 individuals (Marcus et al., 2007). Repeat scans were acquired of five of the subjects in a second session within 90 days of the original scan (Marcus et al., 2007). All images were acquired at 1.5T with the same scanner (Marcus et al., 2007). Images from the OASIS database were corrected for bias field inhomogeneities and positionally normalised¹². The MR images with corresponding segmentations are resampled to a resolution of 1 mm × 1 mm × 1 mm. A description of relevant subsets including information on gender and age is provided in Table 2.3.

Table 2.3: Overview of relevant NMM subsets with respective age and gender information.

subset	# of images / subjects	gender (# male / # female)	age (mean [min; max])
COMPLETE	35 / 30	10 / 20	34.3 [18; 90]
ATLAS	30 / 30	10 / 20	34.3 [18; 90]
REPEAT	10 / 5	2 / 3	24.6 [20; 29]
TRAINING	15 / 15	5 / 10	23 [19; 34]
TEST	20 / 15	5 / 10	45.7 [18; 90]

The manual segmentation into 138 anatomical structures has been carried out by experts¹³ ac-

¹¹<http://www.neuromorphometrics.com/>, last accessed 29 June 2015.

¹²<http://www.neuromorphometrics.com/wp-content/uploads/2013/07/DescriptionofLabeledScans.pdf>, last accessed 29 June 2015.

¹³provided by Neuromorphometrics, Inc. (<http://Neuromorphometrics.com/>, last accessed 29 June 2015) under academic subscription.

cording to publicly available protocols¹⁴. All manual segmentations were quality controlled by another expert¹² (page 83). As suggested by Landman and Warfield (2012), the small regions *vessel*

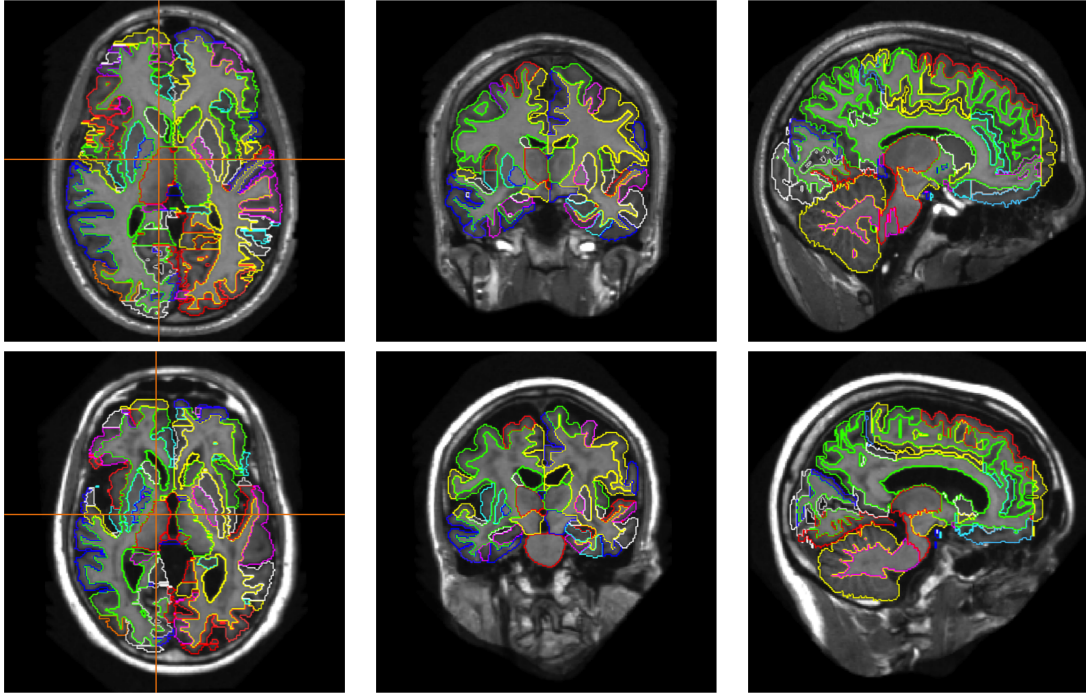


Figure 2.11: Two examples of subjects of the NMM atlas in axial (left), coronal (middle) and sagittal (right) view plane. T1w MR images of a male subject aged 20 years (OAS1_0285, top) and a female subject aged 90 years (OAS1_0083, bottom) with overlaid segmentation contours.

and *cerebral exterior* were excluded in both the left and the right hemisphere. Thus the atlas comprises effectively 134 ROIs of which 36 are non-cortical and 98 cortical. The non-cortical structures comprise several ROIs such as amygdala, caudate nucleus, hippocampus, pallidum, putamen, thalamus. The cerebellum is subdivided into WM, cerebellum exterior and the cerebellar vermal lobules. Cortical ROIs contain a single tissue type (cortical GM) only. Without any further subdivision, cerebral WM is pooled together in one ROI for the left and right brain hemisphere respectively. The 134 ROIs contain 63 distinct anatomical structures which have symmetric counterparts in their opposite hemisphere, in total 126 paired ROIs. The remaining eight unpaired structures are: 3rd ventricle, 4th ventricle, brain stem, CSF, optic chiasm, cerebellar vermal lobules I-V, cerebellar vermal lobules VI-VII, cerebellar vermal lobules VIII-X. A complete list of all ROIs is provided in Appendix B. Examples of brain MR images of a young male subject (20 years) and a rather old female subject (90 years) is illustrated with overlaid manual segmentation outlines in Figure 2.11.

¹⁴<http://neuromorphometrics.org:8080/Seg/> and <http://www.braincolor.org>, last accessed 29 June 2015.

2.3 Biomarkers from structural MRI in AD

2.3.1 Motivation and potential

Non-invasive brain MRI has great potential to provide informative biomarkers for characterising the early onset of AD even before clinical symptoms are apparent or irreversible neuronal damage has occurred (Scheltens et al., 2002; Mueller et al., 2005; Fennema-Notestine et al., 2009). Assuming disease modifying treatment would be available, it is crucial to identify patients well before a clinical diagnosis can be made.

Furthermore, clinical trials can meanwhile greatly benefit from MRI biomarkers as they allow the definition of more rigorous inclusion criteria leading to more homogeneous study groups (Scheltens et al., 2002; Fennema-Notestine et al., 2009; Frisoni et al., 2010; Klöppel et al., 2012). It is assumed that failure of recent trials can be in part attributed to the inhomogeneity of study groups, including patients that for example suffer from a dementia different to AD (Frisoni et al., 2010; Falahati et al., 2014). MRI biomarkers could support the differential diagnosis of various types of early dementia. With imaging information available, MCI subjects that are less likely to convert to AD or have frontotemporal dementia (FTD), vascular dementia (VD) or Lewy body dementia (LBD) could be excluded from a trial. This leads to a study cohort with an increased number of patients that actually convert to AD, which in turn can increase the statistical power of a trial.

There is a vast number of studies that have shown the correlation of biomarkers calculated from brain MRI with the state or disease progression of an AD patient. It is well known that MTL structures are among the first regions affected by the disease (Braak and Braak, 1991). Early involvement of the hippocampus, amygdala and entorhinal cortex is consistently shown in the literature (Braak and Braak, 1991; Jack Jr. et al., 1997; Wenk, 2003; Blennow et al., 2006; Dubois et al., 2007). It is assumed that biomarkers derived from these individual structures can be very sensitive to dementia. These biomarkers might however not be sufficiently specific for AD (Frisoni et al., 2010; Cuingnet et al., 2011). A uniform approach, considering pathological changes of multiple structures within the *whole* brain might increase specificity in dementia diagnosis and support a differential assessment of various dementia types (Klöppel et al., 2008; Frisoni et al., 2010).

2.3.2 Literature overview

There is a substantial amount of published articles studying biomarkers derived from structural MRI in the context of AD. A selection of relevant literature is described in the following. A more complete overview can be found in published surveys such as Scheltens et al. (2002), Ramani et al. (2006), Frisoni et al. (2010), Cuingnet et al. (2011), Falahati et al. (2014) or Bron et al. (2015) and references therein.

2.3.2.1 Brain tissue biomarkers

Klöppel et al. (2008) calculated a probabilistic tissue segmentation of subjects with histopathologically confirmed dementia and HC subjects. After aligning the GM maps to a template space, the authors trained a linear support vector machine (SVM) to classify small cohorts with high accuracy ($\geq 90\%$) in AD and HC. The potential of structural imaging to differentiate FTD and AD was also shown ($\approx 89\%$ accuracy). In Fan et al. (2008), Misra et al. (2009) and Davatzikos et al. (2011) the authors followed a similar approach and aligned previously extracted tissue maps to a template space but recorded the required extent of deformation in regional volumetric maps (Goldszal et al., 1998). The findings confirmed involvement of MTL structures when comparing stable and progressive MCI patients (Misra et al., 2009).

2.3.2.2 Morphometry biomarkers

Similar approaches are based on voxel-based morphometry (VBM) (Ashburner and Friston, 2000; Karas et al., 2004; Chételat et al., 2005), DBM (Ashburner and Friston, 1999) or tensor-based morphometry (TBM) (Hua et al., 2008; Koikkalainen et al., 2011) and often used to study group differences. Here, images or derived information such as tissue segmentations are aligned to a reference space in which statistical tests on tissue probability maps (VBM) or the respective transformations (DBM, TBM) are performed. Significant group differences were shown in global GM volume between HCs and patients with AD using VBM (Karas et al., 2004). Local atrophy in the MTL structures was shown for MCI subjects. In Koikkalainen et al. (2011) the authors used multi-template TBM to classify HC vs. AD subjects (86%), and sMCI vs. pMCI subjects (72%). In Hua et al. (2008), TBM was used in a large study to confirm that brain atrophy in selected ROIs correlates with and is predictive of clinical measurements. It was shown that brain atrophy in MCI converters is higher than in those that do not convert. Teipel et al. (2007) analysed the deforma-

tion maps with respect to a template space using principal component analysis (PCA) and found significant regional differences between patients with AD, MCIs and HCs.

2.3.2.3 Hippocampal biomarkers

Due to its early involvement in dementia, the focus of many published methods lies on the segmentation of the hippocampus to quantify its volume or shape (Dickerson et al., 2001; Shen et al., 2002; Barnes et al., 2008; Colliot et al., 2008; Chupin et al., 2009a,b; Lötjönen et al., 2010; Wolz et al., 2010a; Coupé et al., 2011; Coupé et al., 2012b; Tong et al., 2013). Of further interest is the assessment of the temporal change (*e.g.* hippocampal atrophy), which is discussed in Section 2.3.2.7.

In Dickerson et al. (2001) both hippocampus and entorhinal cortex were manually segmented. The authors found significant group differences between hippocampal volumes for AD, MCI and HC subjects. Further findings indicated that volumes of the entorhinal cortex have higher predictive value to predict conversion from MCI to AD than hippocampal volumes. In Fennema-Notestine et al. (2009) the authors analysed the whole brain using FreeSurfer and showed group differences in hippocampal volume and cortical thickness for HCs and patients with AD. Chupin et al. (2009a) employed an automatic method driven by both anatomical and probabilistic priors to segment the hippocampus. Accuracies ranging from 76% to 80% were reported for AD vs. HC classification. In Colliot et al. (2008), the authors classified patients with AD ($N = 25$) and HCs ($N = 25$) with 84% accuracy, and HCs and MCI patients ($N = 24$) with 73% accuracy based on hippocampal volume. Further, significant group differences in hippocampal volumes for AD, MCI and HC groups were confirmed. Gerardin et al. (2009) calculated automatic hippocampus segmentations and derived coefficients of spherical harmonics (SPHARM) as shape features. It was shown that shape features can potentially improve classification results obtained using hippocampal volumetry only. The authors separated AD patients and HCs with 88% accuracy on a small subset of the ADNI cohort and compared their results to Klöppel et al. (2008) (90%).

2.3.2.4 Cortical thickness biomarkers

Cho et al. (2012) transformed the cortical thickness data of a given subject, extracted using FreeSurfer, into its spatial frequency components. On the same dataset as the one used in Cuingnet et al. (2011) high classification accuracy for both AD vs. HC (sensitivity (SENS): 82%, specificity

(SPEC): 93%) and sMCI vs. pMCI (SENS: 76%, SPEC: 63%) were reported. In Eskildsen et al. (2013) thickness features were extracted from cortical ROIs that were identified as being discriminative for the prediction of AD. The study showed that predicting the conversion from MCI to AD becomes more challenging with increasing time to conversion.

2.3.2.5 Comparison and combination of several biomarkers

In Cuingnet et al. (2011) a variety of approaches based on tissue probability maps, cortical thickness or hippocampal volumetry/shape were compared on an identical subset (N = 509) from the ADNI cohort. High classification accuracies for the separation of patients with AD and HCs (SENS: 81%, SPEC: 95%) were shown, with methods exploring the whole-brain outperforming those focused on the hippocampus. Wolz et al. (2011) also investigated the potential of combining different types of biomarkers derived from structural MRI (hippocampal volume, cortical thickness, TBM and manifold-based learning) to increase classification accuracy. On 834 subjects from the ADNI cohort the authors separated patients with probable AD from HCs with SENS: 93% / SPEC: 85% and pMCIs from sMCIs with SENS: 67% / SPEC: 69%. The authors further provided classification results on the same dataset that was used in Cuingnet et al. (2011), which were comparable.

2.3.2.6 Machine learning and patch-based biomarkers

Coupé et al. (2012b) presented an increased classification accuracy on the dataset used in Wolz et al. (2011) by employing a patch-based grading approach (Coupé et al., 2012a) to estimate the AD disease state. Recently Tong et al. (2014) proposed an innovative method based on multiple instance learning (MIL) using extracted intensity patches as features. Very high classification accuracies were presented on the same dataset that was used in Wolz et al. (2011) (AD/HC, SENS: 86% / SPEC: 93% and pMCI/sMCI, SENS: 69% / SPEC: 74%). In Guerrero et al. (2014), the authors performed population modelling and classification using manifold coordinates learned from automatically defined brain regions. On a dataset from ADNI, however, different to the one used in Wolz et al. (2011), good visualisation properties and competitive classification results were presented (AD/HC, SENS: 86% / SPEC: 85% and pMCI/sMCI, SENS: 75% / SPEC: 67%).

2.3.2.7 Longitudinal biomarkers

Many studies have shown that structural biomarkers such as volume or shape correlate with AD disease progression. However, intersubject variability can lead to substantial overlap with the healthy population and thus limit the discriminative power of these features (Barnes et al., 2008). Approaches analysing the *temporal change* of the whole brain or individual structures aim to reduce this variability by calculating subject specific atrophy rates, which are usually given in per cent per year. The measurable GM atrophy is a consequence of a substantially accelerated, regionally selective loss of neurons in patients with AD (Bondareff et al., 1989; West et al., 1994; Bobinski et al., 1996; Wang et al., 2003).

Scahill et al. (2002) used fluid registration (Freeborough and Fox, 1998; Crum et al., 2001) to analyse serial MRIs of patients with AD and HCs. The authors found both correlations of global brain atrophy with disease progression and increased rates of hippocampal atrophy in patients with AD (Scahill et al., 2002). Due to its early involvement in AD many studies focus on longitudinal change of hippocampal volume or shape (Fox et al., 1996; Wang et al., 2003; Barnes et al., 2004, 2007; Wolz et al., 2010b; Leung et al., 2010a). In Wang et al. (2003) the authors used a deformation based approach to confirm significant hippocampal shape and volume differences between clinical groups of varying severity of dementia as measured by the Clinical Dementia Rating (CDR). Wolz et al. (2010b) employed 4D segmentation based on GC to show the potential of hippocampal atrophy rates calculated over 12 months or 24 months respectively to predict conversion from MCI to AD. In Barnes et al. (2007, 2008); Leung et al. (2010a) the authors employed the boundary shift integral (Freeborough and Fox, 1997) to measure hippocampal atrophy rates and showed good agreement with manually obtained atrophy rates.

2.3.2.8 Classification of biomarkers

When performing classification using a number of features there are different established classifiers for vector-valued features available. The most popular classifiers are SVMs (Cortes and Vapnik, 1995; Vemuri et al., 2008; Magnin et al., 2009; Falahati et al., 2014). Alternatives are random forests (RFs) (Breiman, 2001; Gray et al., 2013) or linear discriminant analysis (LDA) (Wolz et al., 2011). Wolz et al. (2011) observed higher classification accuracies using LDA instead of SVM. However, Aguilar et al. (2013) and Falahati et al. (2014) conclude that most classifiers perform similar given identical feature sets for classification.

2.3.3 Limitations and challenges

Most of the published studies can only be compared qualitatively. In many cases different features, different methodology to extract the features and different classification techniques are used (Falahati et al., 2014). In addition, methods are applied to different cohorts or different subsets thereof. A further limitation of most studies is the lack of histopathologically confirmed ground truth diagnoses, which can only be obtained with brain autopsy. This is why most studies fall back on clinically obtained diagnoses that are potentially incorrect (Bron et al., 2015).

Recently, Bron et al. (2015) organised the Computer-Aided Diagnosis of Dementia (<http://caddementia.grand-challenge.org>, last accessed 29 June 2015) (CADDementia) challenge and provided an identical reference cohort on which classification approaches could be objectively evaluated. Based on T1w MR scans, authors can submit classification results (HC, MCI, AD) while being blinded to the clinically obtained ground truth. Of particular importance for routine use in clinical practice is the interpretability of biomarkers (Falahati et al., 2014). Many recently developed methods heavily rely on machine learning techniques, *e.g.* learned manifolds (Wolz et al., 2011; Guerrero et al., 2014) or MIL (Tong et al., 2014). Even though these methods are highly accurate, their interpretation might be difficult and complicate their acceptance for clinical application. Thus, it seems desirable to calculate features that are on the one hand easy to interpret but on the other hand as informative as features obtained through sophisticated machine learning techniques.

While many studies focus on the separation of patients with AD, pMCI, sMCI and HCs there is a high need for an improved differential diagnosis of the various dementia types (Klöppel et al., 2012). This could support both patient specific treatment decisions but also the recruitment process of clinical trials. It is known that a differential diagnosis of FTD and AD is possible based on biomarkers derived from structural MRI (Klöppel et al., 2008; Frisoni et al., 2010). In addition, atrophy in MTL regions could be discriminative to separate LBD from other dementia types (Barber et al., 1999). Furthermore, a better characterisation of the MCI stage seems desirable. There is strong evidence that different anatomical brain structures are affected at different stages of the disease (Dickerson et al., 2001).

However, there is consensus that methods or biomarkers based on individual MTL structures are most likely not sufficiently informative to address the above-mentioned challenges (Frisoni et al., 2010). One interesting avenue is to employ recently published, highly detailed and accurate brain

atlases to form robust whole-brain segmentation methods. This would allow for the quantification of easily interpretable biomarkers such as volumes, shapes, cortical thickness and atrophy rates. Discriminative power could then be achieved by combining information from numerous structures over the whole brain within a single classifier.

2.4 Biomarkers from structural MRI in TBI

2.4.1 Motivation and potential

Accurate quantitative assessment of the neuroanatomic changes occurring during and after TBI is crucial to assist the understanding of TBI disease progression. There are several studies investigating microstructural changes based on DWI because of its particular sensitivity to disease-related alterations in the WM tracts (Benson et al., 2007; Bendlin et al., 2008; Kinnunen et al., 2011). ROI analyses in conjunction with DWI (Bendlin et al., 2008; Warner et al., 2010a) or PET imaging (Ramlackhansingh et al., 2011) could enable a better understanding of the secondary processes that cause ongoing brain atrophy in the chronic phase of TBI. It is becoming clear that the key to a better understanding of TBI disease progression is the combination of imaging information obtained from multiple modalities (Sharp and Ham, 2011; Irimia et al., 2012; Shenton et al., 2012). However, as the focus of this thesis is the extraction of anatomical ROIs the following review will be restricted to studies based on structural MR imaging.

It is well known that a reduction of total brain volume and cerebral atrophy are common sequelae of TBI (Blatter et al., 1997; Bigler, 2001; Gale et al., 2005; Bendlin et al., 2008; Ding et al., 2008). There is also increasing consensus that the volume of structural ROIs has the potential to support the outcome prediction of TBI (Bigler, 2001; Warner et al., 2010b,a; Irimia et al., 2012). The early identification of affected brain regions that are likely to degenerate due to the primary injury could thus be key to an effective disease treatment (Irimia et al., 2012). Immediate and targeted treatment, enabled and supported by neuroimaging, could improve the final outcome of the disease but also reduce financial costs through shorter hospital stays (Irimia et al., 2011). In Bigler (2001) the authors further pointed out that the volumetric measurement of subcortical structures might reveal irregularities that would be difficult to catch by visual inspection only.

TBI is a very heterogeneous disease as it substantially depends on the type of injury (*e.g.* vehicle accident, fall, assault), severity of injury (*e.g.* vehicle speed, fall height, assault weapon) and loca-

tion of the impact. It is due to this heterogeneity that comparably large sample sizes are required to show significant differences in research studies or treatment effects in clinical trials (Irimia et al., 2012). Performing a quantitative analysis on potentially hundreds of patients based on manual expert measurements is, however, barely feasible due to both time and financial constraints. Methods to automatically extract biomarkers from MR brain images are thus a critical building block to enable large-scale TBI studies (Irimia et al., 2012). Nevertheless, current research on robust methods to automatically process MR images of injured brains is very limited (Irimia et al., 2011, 2012).

2.4.2 Literature overview

In the following an overview is given over studies that have investigated the potential of neuroimaging in the context of TBI. Particular focus is put on measures derived from structural MR brain images. A more detailed summary can be found in overview articles such as Bigler (2001), Irimia et al. (2012) or Shenton et al. (2012).

2.4.2.1 Morphometry biomarkers

Substantial group differences in GM density between healthy control subjects and patients with TBI were confirmed using VBM (Salmond et al., 2004; Gale et al., 2005). Salmond et al. (2004) performed VBM to compare MR images from 22 patients acquired at least six months post injury to a matched set of control subjects. In this study reduced GM density in thalamus, basal forebrain, hippocampal formation and regions of the neocortex were identified (Salmond et al., 2004). The authors explicitly mentioned the difficulty of reliably aligning the MR images to the template space given potential pathology in the images. To increase robustness, only linear alignment was performed (Salmond et al., 2004). In Gale et al. (2005), the authors employed VBM to compare the GM density of nine patients with TBI (mild to severe injury) to age and gender matched controls. Based on follow-up MR images acquired around one year after the injury, the authors found a significant decrease in GM concentration in a multitude of brain regions. These regions included the cerebellum, frontal and temporal cortices, but also subcortical structures (Gale et al., 2005). In Kim et al. (2008) it is stated that the above-mentioned VBM studies might have limited sensitivity. As limitations, inaccuracies when aligning the subjects to the reference template and performing the tissue segmentation on brain images with substantial atrophy are mentioned. The authors suggested to perform TBM instead and to employ an advanced registration technique (SyN, (Avants et al.,

2008)) that allows the recovery of large deformations and the estimation of a population specific template (Kim et al., 2008). Based on a study cohort of 29 patients with at least moderate TBI and 20 HCs, local reductions in WM and subcortical regions such as thalamus, corpus callosum and caudate were shown in the TBI group.

2.4.2.2 Structural biomarkers

Most of the few existing studies that analysed structural morphometric measures (Strangman et al., 2010; Warner et al., 2010a,b) were based on the segmentation techniques available in FreeSurfer and have investigated small patient cohorts (Warner et al., 2010a,b). Warner et al. (2010a) analysed the relation of axonal injury quantified from DTI with structural volumes in the chronic phase (8 months after injury) of a TBI. Structural volumes of hippocampus, amygdala and thalamus but also of cortical ROIs were stronger correlated with white matter integrity at the chronic than at the acute time point (Warner et al., 2010a). This suggests that white matter integrity can change due to secondary processes far beyond the acute phase (Warner et al., 2010a). In Strangman et al. (2010), 50 patients that sustained TBI were enrolled in a memory rehabilitation program and their individual progress recorded. The study investigated the predictive value of structural brain volumes with respect to the outcome of the rehabilitation (Strangman et al., 2010). Ramlackhansingh et al. (2011) used ROIs segmented from structural MRI and PET to demonstrate that inflammatory processes remain active for months or years following a brain trauma.

Several studies (Strangman et al., 2010; Warner et al., 2010a,b; Ramlackhansingh et al., 2011) have identified structures, including thalamus and hippocampus that are affected by TBI and are of significant value when predicting clinical outcome.

2.4.2.3 Longitudinal biomarkers

Bendlin et al. (2008) performed a longitudinal analysis of 46 patients with TBI with respect to 36 matched HCs using VBM. DTI and T1w MRIs were acquired around two months post injury and after 12.7 months respectively. The authors found that the vast majority of fibre bundles were affected in TBI (Bendlin et al., 2008). Further, both structural integrity quantified from DTI and WM/GM density calculated from T1w images declined while scores related to neuropsychological function improved (Bendlin et al., 2008). In Sidaros et al. (2009) longitudinal changes in the months following a severe TBI were investigated. In this study, 24 patients were compared to 14

HCs. Specifically, Sidaros et al. (2009) performed TBM on the deformations required to align the corresponding MR images taken at around eight weeks and 12 months post injury. The authors found both an increased reduction in brain volume when compared to the HC group and regional involvement of brainstem, thalamus, corpus callosum, putamen and cerebellum (Sidaros et al., 2009). Regions that are known to be susceptible to DAI were particularly affected by late atrophy (Sidaros et al., 2009). In Warner et al. (2010b) the authors analysed the correlation of structural brain atrophy of 25 patients who had sustained a DAI with functional outcome. Several brain structures showed significantly increased structural atrophy when compared to a control group of 22 age and gender-matched controls eight months post injury (Warner et al., 2010b). Irimia et al. (2011) compared TBI related changes assessed from images acquired with multiple MR sequences in three representative patients. The authors used semi-automatic methods for brain tissue and lesion classification and created 3D models of TBI pathology.

2.4.3 Limitations and challenges

Many of the conducted studies investigated group differences between patients with TBI and healthy subjects (Salmond et al., 2004; Kim et al., 2008; Bendlin et al., 2008; Sidaros et al., 2009; Warner et al., 2010b). However, given that TBI is a very heterogeneous disease, a more personalised analysis is desirable (Shenton et al., 2012). Comparing heterogeneous groups of patients with TBI to HCs might be the reason that large sample sizes are necessary to find significant group differences (Irimia et al., 2012). As discussed in Shenton et al. (2012), biomarkers quantified from anatomical ROIs could be compared to reference values from HCs in statistical tests to find brain regions affected by the injury. Here, the segmentation of these ROIs is crucial as it allows the calculation of structural volumes or atrophy but also derived measures from, for example, DWI. Using DWI, microstructural changes can be measured based on ROI-specific statistics of FA or MD. Structural ROIs could further enable a connectivity analysis based on fibre tracts calculated using tractography. Based on these measurements, ideally extracted from multiple modalities, a patient-specific disease profile could be created and a derived surrogate measure employed for group analyses (Shenton et al., 2012). Measuring longitudinal change of the biomarkers is also hoped to enable a better understanding of the disease progression. For example, the causal relation of secondary degenerative processes to the primary injury is of high interest (Irimia et al., 2012). Further, disease models could be created that support the assessment of the disease progression, outcome prediction and especially

quantification of treatment effects. It is hoped that disease dynamics can be better characterised to improve the understanding of relevant criteria for favourable patient recovery (Shenton et al., 2012).

The automatic structural segmentation of MR brain images of patients with TBI remains, however, a difficult endeavour as most existing methods lack robustness towards TBI-related changes in anatomy (Irimia et al., 2011, 2012). In the presence of gross pathologies such as hemorrhagic lesions/oedema (in the acute phase) or substantial atrophy (in the chronic phase) most of the established segmentation techniques yield unsatisfying results. While in AD brain changes are consistent with disease progression, MR brain images of patients with TBI can show inconsistent and gross pathological change. It is this high variability and extent of brain change following a moderate or severe TBI that makes the segmentation task so demanding.

A more comprehensive discussion of current and potential future research directions is provided in Irimia et al. (2012) or Shenton et al. (2012).

3 Robust whole-brain segmentation: Multi-Atlas Label Propagation using Expectation-Maximisation (MALP-EM)

This chapter is based on:

C. Ledig, R. A. Heckemann, A. Hammers, J. C. Lopez, V. F. J. Newcombe, A. Makropoulos, J. Lötjönen, D. Menon and D. Rueckert, “Robust whole-brain segmentation: application to traumatic brain injury”, *Medical Image Analysis*, 21(1), pp. 40-58, 2015.

C. Ledig, R. Wolz, P. Aljabar, J. Lötjönen, R. A. Heckemann, A. Hammers, and D. Rueckert, “Multi-class brain segmentation using atlas propagation and EM-based refinement”, *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, pp. 896-899, 2012.

Abstract: *Existing segmentation methods are of limited robustness when applied to MR brain images with substantial pathological changes (cf. Section 2.4.3). In the following, this challenge is addressed by proposing a framework for the robust and fully-automatic segmentation of MR brain images called Multi-Atlas Label Propagation with Expectation-Maximisation based refinement (MALP-EM). The presented approach is based on a robust registration approach (MAPER), highly performant label fusion (JLF) and intensity-based segmentation refinement using EM. The framework is devised to be applicable for the segmentation of brain images with gross changes in anatomy. It is proposed to account for consistent registration errors by relaxing anatomical priors obtained by multi-atlas label fusion and to employ a weighting scheme to locally combine anatomical atlas priors and intensity-refined posterior probabilities. The method is evaluated on a benchmark dataset used in a recent MICCAI segmentation challenge. In this context it is shown that MALP-EM is competitive for the segmentation of MR brain images of healthy adults when*

compared to state-of-the-art automatic labelling techniques. To demonstrate the versatility of the proposed approach, MALP-EM is employed to segment 125 MR brain images into 134 regions from subjects who had sustained TBI. A specific protocol is employed to assess segmentation quality if no manual reference labels are available. Based on this protocol, three independent, blinded raters confirmed on 13 MR brain scans with pathology that MALP-EM is superior to established label fusion techniques. The robustness of the proposed segmentation approach is visually confirmed on the full cohort. It is investigated whether derived symmetry-based imaging biomarkers potentially correlate with and predict clinically relevant variables in TBI such as the MCS or GOS. Specifically, it is shown that TBI patients with favourable outcomes can be stratified from non-favourable outcomes with 64.7% accuracy using acute-phase MR images and 66.8% accuracy using follow-up MR images. Furthermore, the calculated biomarkers allowed to differentiate subjects with the presence of a mass lesion or midline shift from those with diffuse brain injury with 76.0% accuracy. The thalamus, putamen, pallidum and hippocampus are particularly affected. Their involvement predicts TBI disease progression.

The automatic segmentation of the brain into multiple distinct ROIs has great potential as it allows the extraction of volumetric or atrophy biomarkers. Furthermore, anatomical ROIs are essential for numerous analyses, such as the investigation of brain connectivity or the calculation of statistics from DWI images.

In this chapter, a retrospective cohort of TBI patients is employed both to validate the proposed framework and to present a proof of concept that disease relevant biomarkers can be automatically extracted. The described method will then be used to analyse large AD and TBI cohorts in Chapter 5 and Chapter 6.

3.1 Introduction

The main motivation for the work described in this chapter was to devise a fully automatic and robust segmentation method that allows accurate measurement of various brain structures in MR brain images in the presence of severe pathologies. A subject with highly altered brain configuration is shown with overlaid automatic segmentations in Figure 3.1 to illustrate the difficulty of the segmentation task. Figure 3.1 shows that established registration techniques such as MAPER

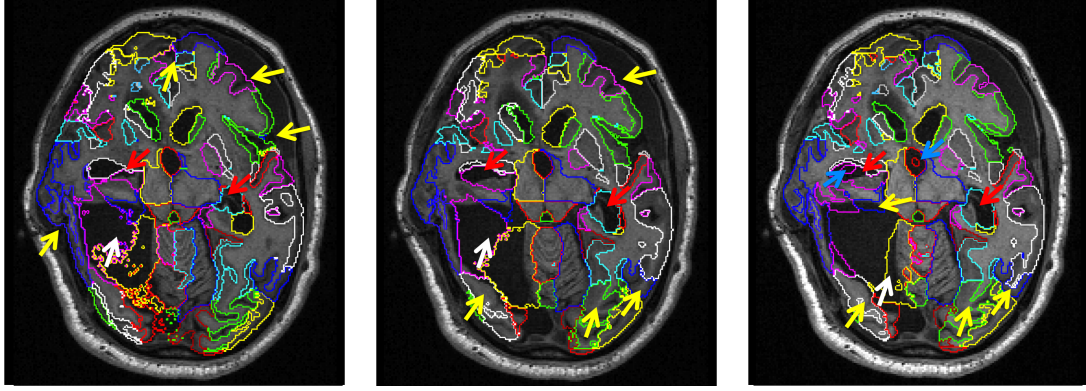


Figure 3.1: Example of segmentation results obtained on a subject with highly altered brain configuration. Segmentations calculated with MAPER using majority voting (left, Heckemann et al. (2010)) and SyN (Avants et al., 2008) from the Advanced Normalization Tools (ANTs) package using either majority voting (middle) or the joint label fusion (right, Wang et al. (2013b)). Red arrows: substantial oversegmentation of the hippocampus; yellow arrows: inaccurate cortex segmentation due to gross brain deformation; blue arrows: ventricles incorrectly labelled as background; white arrows: region of missing tissue prohibits reasonable one-to-one mapping of the atlases.

(Heckemann et al., 2010) or SyN (Avants et al., 2008) struggle to establish a plausible mapping between the available atlas images and an image of an injured brain. Both the presence of gross deformation and the potential absence of brain tissue prevent an accurate anatomical correspondence estimation. Even the application of a leading label fusion technique (Wang et al., 2013b) is not able to correct the substantial and consistent errors of alignment.

Thus there is a high need to address the key need in the management of TBI, that was identified by Irimia et al. (2012): “[...] the key methodological hurdle that must be overcome in order to make structural neuroimaging a powerful tool for predicting TBI outcome is the current paucity of automated image processing methods that can allow researchers to analyse large numbers of TBI CT/MRI volumes without the need for excessive user input or intervention.”

To enable the analysis of large TBI cohorts it is critical to improve on the limitations of existing automatic segmentation methods. A detailed overview over the relevant building blocks for atlas-based brain segmentation was provided in Section 2.1. However, in the following a brief introduction to the literature related to the work described in this chapter is provided.

Atlas propagation techniques rely on the accurate registration of the atlas and unsegmented MR image to determine the spatial transformation of the atlas labels into the target space. This can

be difficult if the target image differs from the available atlases due to the presence of pathology. As described in more detail in Section 2.1.2 several methods were proposed to tackle this problem (Angelini et al., 2007; Bauer et al., 2013). Most of these approaches rely either on a mask excluding the pathological regions (Brett et al., 2001; Stefanescu et al., 2004; Andersen et al., 2010) or tumour growth models (Bach Cuadra et al. (2004); Zacharaki et al. (2009); Niethammer et al. (2011)). Due to the substantial heterogeneity of TBI these approaches are of limited applicability. The manual annotation of masks is prohibitive when large cohorts are to be analysed and tumour growth models do not adequately describe the variety of TBI pathology such as contusions, haemorrhage or haematomas. In this chapter, the registration method MAPER was chosen that has been shown to be highly performant when employed for segmentation of healthy subjects (Landman and Warfield, 2012). While MAPER is robust in the context of AD related atrophy (Heckemann et al., 2010) it does not explicitly account for pathology typical in TBI. Misregistration and corresponding labelling errors are thus expected. In the presented approach these problems are addressed subsequently, after the registration, by using label fusion techniques and exploiting image intensity information.

A particularly successful strategy to fuse the propagated atlas labels into a consensus segmentation estimate is JLF, which was recently proposed by Wang et al. (2013b). In this state-of-the-art approach, as evaluated in Landman and Warfield (2012), segmentation bias is reduced by estimating joint segmentation errors of different atlas pairs (Wang et al., 2013b). Recently, methods based on nonlocal patch-based label fusion have been suggested (Coupé et al., 2011; Rousseau et al., 2011) that rely on affine alignment with a template library and thus relax the requirement for accurate nonrigid registrations. Patch-based methods have been developed further, often with focus on a particular application. For example Tong et al. (2013) used dictionary learning and sparse coding for hippocampal segmentation in patients with AD. A detailed overview of atlas-based label fusion methods is provided in Section 2.1.3 or also by Cabezas et al. (2011).

Atlas-based segmentation can be further improved by incorporating intensity information from the unseen image through a Gaussian mixture model (GMM) (Van Leemput et al., 1999; Fischl et al., 2002) (*cf.* Section 2.1.4). The resulting optimisation problem is often solved using EM optimisation (Van Leemput et al., 1999; Lötjönen et al., 2010; Cardoso et al., 2011; Ledig et al., 2012b; Cardoso et al., 2013b) or GC (van der Lijn et al., 2008; Wolz et al., 2010b). While the EM approach enables simultaneous probabilistic segmentation of multiple brain structures, many graph-

cut based methods yield binary labels for individual structures. Also approaches that propagate atlas labels over a graph (Wolz et al., 2010a; Cardoso et al., 2012) or within clusters (Ribbens et al., 2014) enjoy increasing attention in the community. In the context of brain tissue segmentation it has been further shown that the relaxation of anatomical atlas priors can improve segmentation quality (Cardoso et al., 2011, 2013b). Chupin et al. (2009b) automatically correct atlas registration mismatches of the hippocampus. As an alternative or complement to either approach, Wang et al. (2011a) proposed to use machine learning techniques to learn systematic segmentation errors that are then corrected in a post-processing step.

Atlas-based approaches require a number of brain atlases that, in the ideal case, have been generated by expert manual delineation. In this work an atlas set is used that was the basis of a recent whole-brain segmentation challenge (MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling, Landman and Warfield (2012)). Thus a direct comparison to the methods that were evaluated in this competition can be provided. In general, other whole-brain atlases (*cf.* Section 2.2) are suitable for the proposed approach.

It is illustrated in Figure 3.1 that even state-of-the-art label fusion methods based on established registration techniques can yield unsatisfying segmentation results when applied to MR brain images that show a substantial structural change. The objective of this work to improve on these limitations is pursued on three levels. First, the best features of state-of-the-art atlas-based segmentation tools are combined into a new framework, MALP-EM, by building on MAPER and adding the benefits of JLF and an intensity-based refinement using EM. Second, this method is adapted for the challenges posed by highly altered brain configurations. To achieve this, a prior relaxation scheme is used in regions where accurate alignment of the images is impossible due to missing brain tissue or severe deformation. The proposed approach does not influence or improve the performed atlas to image registration, which can consistently fail in the presence of pathologies. Instead, it is proposed to correct based on image intensities the anatomical atlas priors obtained by label fusion. Further a data-driven and locally adaptive weighting scheme is employed to combine anatomical atlas prior probabilities and intensity-refined posterior probabilities for maximum benefit. Third, the modified MALP-EM algorithm is used to segment 125 MR brain scans of a heterogeneous population of 101 subjects who had sustained TBI. Segmentation accuracy on this TBI cohort is assessed based on a specific protocol that was independently followed by three blinded raters. The

protocol enables an expert to rate hippocampus, thalamus, putamen and occipital pole segmentation quality in the absence of manual reference segmentations. Eventually, biomarkers are derived that are based on an index that quantifies asymmetry between structures appearing both in the left and right brain hemisphere (absolute asymmetry index (Galaburda et al., 1987; Bonilha et al., 2013) (AAI)). It is shown that single time point MR imaging based variables potentially correlate with and predict outcome-relevant clinical variables.

3.2 Materials and methods

3.2.1 Materials

3.2.1.1 Traumatic brain injury cohort

T_1 -weighted MR brain images were obtained of 101 TBI patients and provided through the University Division of Anaesthesia, Cambridge University, UK. The images were acquired using an MPRAGE sequence on a Siemens MAGNETOM TrioTim Syngo with parameters: TR 2300 ms, TE 2.98 ms, TI 900 ms, flip angle 9° , matrix size $256 \times 240 \times 176$ and an isotropic voxel size of $1.0 \text{ mm} \times 1.0 \text{ mm} \times 1.0 \text{ mm}$. Patients underwent MRI in the acute-phase, as part of the follow-up, or both. As only 24 patients had MR scans at both time points, the focus of this chapter is on a cross-sectional analysis. In total 125 datasets were available, including 61 acute and 64 follow-up MR brain scans. Information about the patients' gender and age distributions, the elapsed time between scanning date and injury and the GCS is summarised in Table 3.1. The GCS is a clinical score that quantifies a patient's level of consciousness in the acute stage of the injury (Teasdale and Jennett, 1974). The datasets were further grouped by clinical scores using MCS and the GOS, as shown in Table 3.2. MCS is a score based on the worst acute CT image within 24 hours of injury. MCS takes into account brain pathology such as lesion load, the presence of oedema and midline shift caused by the injury. In contrast, the GOS is a clinical measure categorising the outcome of TBI and is assessed 6 months after injury or once the TBI outcome is considered stable. In Section 3.3.2.3 correlations of these clinical scores with asymmetry biomarkers derived from brain MR scans are described. Details of the definition of MCS and GOS are provided in Appendix C and Appendix D.

Table 3.1: Overview of the available MR images with patient gender, patient age, scan time relative to injury, and GCS.

	acute-phase MR image	follow-up MR image
# of subjects	61	64
gender (# male / # female)	48/13	40/24
age (mean $\pm$ standard deviation)	36.6 $\pm$ 14.9 years	36.1 $\pm$ 14.9 years
time since injury (mean $\pm$ standard deviation)	3.7 $\pm$ 4.2 days	10.0 $\pm$ 7.2 months
GCS (median [min; max])	5 [3;15]	7 [3;15]

Table 3.2: Clinical variables of the 61 acute-phase TBI images (MCS, GOS) and the 64 follow-up MRIs (GOS). n/a: not available.

MCS	n/a	DI I (1)	DI II (2)	DI III (3)	DI IV (4)	EML (5)	NEML (6)
# of subjects per group (acute-phase)	3	4	29	2	0	16	7

GOS	n/a	D (1)	VS (2)	SD (3)	MD (4)	GR (5)
# of subjects per group (acute-phase)	6	8	1	18	17	11
# of subjects per group (follow-up)	0	0	0	21	28	15

3.2.1.2 Atlases

The atlas cohort used in this study consisted of 35 manually annotated MR brain images of 30 subjects of the OASIS database (Marcus et al., 2007). The manual segmentation into 138 anatomical structures has been carried out by experts according to publicly available protocols¹ and were provided by Neuromorphometrics, Inc. (<http://Neuromorphometrics.com/>, last accessed 29 June 2015) under academic subscription. The same atlas cohort was used in the recent “MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling” (Landman and Warfield, 2012). As suggested by Landman and Warfield (2012), the small regions (order of 100 voxels each) “vessel” and “cerebral exterior” were excluded in the experiments in both the left and the right hemisphere, so that effectively 134 structures are investigated. In five of the subjects, repeat scans were acquired in a second session within 90 days of the original scan (Marcus et al., 2007).

The 134 atlas labels comprise 63 anatomical structures which have symmetric counterparts in their opposite hemisphere, in total 126 labels (*cf.* Appendix B). The remaining eight unpaired structures are: 3rd ventricle, 4th ventricle, brain stem, CSF, optic chiasm, cerebellar vermal lobules I-V, cerebellar vermal lobules VI-VII, cerebellar vermal lobules VIII-X.

The NMM atlas is described in more detail in Section 2.2.5.

¹<http://neuromorphometrics.org:8080/Seg/> and <http://www.braincolor.org>, last accessed: 8 December 2014.

3.2.2 Multi-Atlas Label Propagation with Expectation-Maximisation based refinement (MALP-EM)

3.2.2.1 Notation

To present the framework called Multi-Atlas Label Propagation with Expectation-Maximisation based refinement (MALP-EM), the notation introduced in Section 2.1 is employed:

An unsegmented T_1 -weighted MR target image $\mathbf{I}_{\text{tgt}} \in \Omega$ is segmented into $K = 134$ distinct structural regions. The image is indexed as $\mathbf{I}_{\text{tgt}} = \{y_1, y_2, \dots, y_n\}$ where $y_i \in \mathbb{R}^+$, with $i = 1, \dots, n$, denotes the intensity value of the i -th voxel. To incorporate expert knowledge into the segmentation process, M manually annotated brain atlases are employed which are denoted by $\mathbf{A}_m$ with $m = 1, \dots, M$. ϕ_m denotes the calculated transformation from the atlas space of $\mathbf{A}_m$ in the coordinate system of $\mathbf{I}_{\text{tgt}}$ and $\mathbf{A}_m^\phi$ denotes the propagated atlas. A subject specific probabilistic segmentation $\Pi = \{\pi_1, \pi_2, \dots, \pi_n\}$ is created employing multi-atlas label fusion. Π is then relaxed based on image intensities to Π^R and used as spatial prior in the EM framework. Using an EM approach, an intensity-refined probabilistic segmentation of $\mathbf{I}_{\text{tgt}}$ is estimated that is denoted by $Z = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$. Here π_i , π_i^R and $\mathbf{z}_i$ are vectors of size K and the k^{th} component represents the probability that a voxel i belongs to a region k . Thus Π denotes the subject specific probabilistic segmentation before intensity-based refinement and Z after intensity-based refinement respectively. Furthermore, a normal distribution $\mathcal{N}(\mu_k, \sigma_k)$ is abbreviated with $\mathcal{N}_k$ where μ_k is the mean and σ_k the standard deviation of the intensity distribution within label k .

3.2.2.2 Registration and joint label fusion

For each unsegmented image $\mathbf{I}_{\text{tgt}}$, M transformations ϕ_m are obtained by registering M manually generated atlases to the coordinate space of $\mathbf{I}_{\text{tgt}}$. In this study an enhanced registration approach is employed that has been developed as part of MAPER. MAPER incorporates tissue probability maps into a nonrigid registration scheme based on free form deformations (FFDs) (Rueckert et al., 1999; Modat et al., 2010).

A probabilistic map π_k of each anatomical structure k is then formed from the M transformed atlases $\mathbf{A}_m^\phi$ using the JLF strategy presented by Wang et al. (2013b). The implementation of JLF is employed that is publicly available at https://www.nitrc.org/projects/picssl_malf/ (Version 1.2, last accessed: 8 December 2014) using standard parameters. JLF has been shown to be

a leading label fusion technique (Landman and Warfield, 2012). This procedure is illustrated in Figure 3.2.

Joint label fusion combined with the non-local patch-based concept ranks among the most accurate approaches for label fusion (*cf.* Section 2.1.3.5). However, based on the parameter choice (*e.g.* patch size, search-window size) and number of fused atlases (M) this procedure is computationally expensive and requires several hours on a single core for a single subject. With the aim to make the presented framework more applicable in practice, an alternative, computationally more efficient fusion strategy is proposed in the following.

3.2.2.3 Locally Gaussian weighted label fusion

As alternative to JLF a LWF approach is presented. In contrast to Artachevarria et al. (2009) where the weighting of individual patches is determined using SSD it is proposed to employ a Gaussian weighted sum of squared differences (GSSD) instead. In this approach the spatial distance of voxels from the voxel of interest is taken into account when comparing neighbourhoods. Further, GSSD can be calculated very efficiently with Gaussian convolution. For K structures and M propagated atlases the probabilistic atlas at voxel i for label k is calculated as:

$$\pi_{ik} = \frac{\sum_{m=1}^M w_m^1(i) \delta(I_m^A(i), k)}{\sum_{k'=1}^K \sum_{m=1}^M w_m^1(i) \delta(I_m^A(i), k')} \quad (3.1)$$

With the difference image $\mathbf{D}_m$ between the image $\mathbf{I}_{\text{tgt}}$ and atlas image $\mathbf{I}_m^A$ defined as $\mathbf{D}_m = (\mathbf{I}_{\text{tgt}} - \mathbf{I}_m^A)^2$, $w_m^1(i)$ and $\delta(\cdot, \cdot)$ is given as:

$$w_m^1(i) = [G_\sigma \star \mathbf{D}_m]_i^{-1}, \quad \delta(v, w) = \begin{cases} 1 & \text{if } v = w \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

Here G_σ is a zero-centred Gaussian with parameter σ and $\star$ denotes the convolution.

However, the accuracy of locally weighted fusion strategies heavily relies on the assumption that patches of the propagated atlas MR images $\mathbf{I}_m^A$ and unseen image $\mathbf{I}_{\text{tgt}}$ are comparable using a certain measure. The GSSD is obtained on the intensity normalised MR images. In this work the required intensity normalisation is obtained by linearly rescaling the image intensities, similar to Nyul and Udupa (1999). Specifically, the image intensities are rescaled to intensity values in the range of 0 to 255 where the rescaling factor is determined based on the robust minimum and maximum estimates

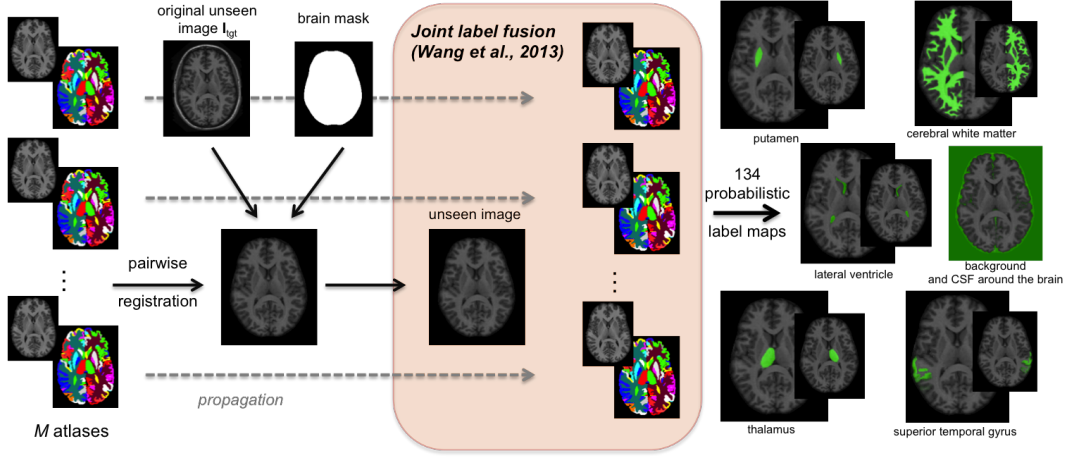


Figure 3.2: Schematic process of the calculation of the subject specific spatial priors Π for an unsegmented target image I_{tgt} . After brain extraction and bias correction, the available M atlases are registered to the space of I_{tgt} . Using these transformations, label maps and corresponding T_1 -weighted MR images are mapped to the space of I_{tgt} . The label maps are then averaged into probabilistic priors for the individual structures using the joint label fusion (Wang et al., 2013b). A subset of the 134 probabilistic labels is shown in green.

of the image intensities (2% and 98% percentiles) (Smith, 2002). To further improve robustness and avoid interference from high-intensity non brain voxels that might be present due to errors in the brain extraction procedure, the robust estimates are calculated using an eroded version (5 iterations of erosion by one voxel) of the brain extracted image.

3.2.2.4 Relaxation of probabilistic priors

Segmentation refinement based on image intensities relies heavily on the probabilistic priors Π . As a consequence, the segmentation at voxel i cannot be refined if all atlases agreed on a certain label k (π_{ik} close to or equal to 1). In general, this is a sensible constraint assuming that at least a subset of the propagated atlases votes for the correct label. When segmenting MR images showing significant pathologies, however, there is evidence that this assumption is no longer justified. Especially in regions that undergo large deformations, for example the inferior lateral ventricles when the target region is enlarged due to injury, swelling or atrophy, unanimous bias in all individual segmentations can be observed. Label fusion approaches can thus return substantial mislabelling of subcortical grey matter structures such as the hippocampus, as well as of cortical regions. This problem is illustrated in the top left image of Figure 3.3. The intensity-based EM-refinement is restricted by these high prior label probabilities Π and thus unable to entirely correct this systematic error.

This problem can be tackled by calculating relaxed priors Π^R from the label probabilities Π . Specifically the probabilistic priors are relaxed based on the probabilistic label fusion estimates and the actual image intensities. Assuming a Gaussian distribution, a common parameter set $(\mu_{\text{CSF-like}}, \sigma_{\text{CSF-like}})$ of eight “CSF-like” structures can be estimated. The set of structures denoted as “CSF-like” is defined as {background (essentially external CSF), 3rd ventricle, 4th ventricle, CSF, right/left inferior lateral ventricle, right/left lateral ventricle}. Furthermore, an individual parameter set (μ_k, σ_k) is estimated for each structure k based on the probabilistic prior segmentation Π :

$$\mu_k = \frac{\sum_i \pi_{ik} y_i}{\sum_i \pi_{ik}}, \quad \sigma_k = \sqrt{\frac{\sum_i \pi_{ik} (y_i - \mu_k)^2}{\sum_i \pi_{ik}}} \quad (3.3)$$

In the actual relaxation step a fraction α_{ik} of the prior probability π_{ik} is redistributed from a structure k to one of the eight CSF-like structures k_{CSF} . At an image voxel i , k_{CSF} is determined as the CSF-like structure with the highest prior probability or set to background otherwise:

$$k_{\text{CSF}} = \begin{cases} \arg \max_{k \text{ is CSF-like}} \pi_{ik} & \text{if } \pi_{ik} \neq 0 \text{ for at least one } k \in \text{CSF-like} \\ \text{background} & \text{else} \end{cases} \quad (3.4)$$

Then α_{ik} is calculated based on the probability that the voxel with intensity y_i comes either from the intensity distribution $\mathcal{N}_k^\Pi$ estimated in label k or $\mathcal{N}_{\text{CSF-like}}^\Pi$ accordingly. Specifically, α is set to:

$$\alpha_{ik} = \begin{cases} 0 & \text{if } \mathcal{N}_k^\Pi(y_i) \geq \mathcal{N}_{\text{CSF-like}}^\Pi(y_i) \\ \max(0, \min(0.5 - \pi_{ik_{\text{CSF}}}, \pi_{ik})) & \text{else} \end{cases} \quad (3.5)$$

This means that it is not allowed for a CSF-like label to exceed 50% probability and only voxels are corrected that have a higher probability of belonging to the CSF-like label k_{CSF} , as defined in Equation 3.4, than to k . Finally the relaxed prior probability Π^R is calculated as:

$$\pi_{ik}^R = \begin{cases} \pi_{ik} + \sum_{l \neq k_{\text{CSF}}} \alpha_{il} & \text{if } k = k_{\text{CSF}} \\ \pi_{ik} - \alpha_{ik} & \text{else} \end{cases} \quad (3.6)$$

The two images on the left in Figure 3.4 show where the voxel priors are relaxed by the proposed method, both for a distinctly deformed brain and for a brain with a normal configuration.

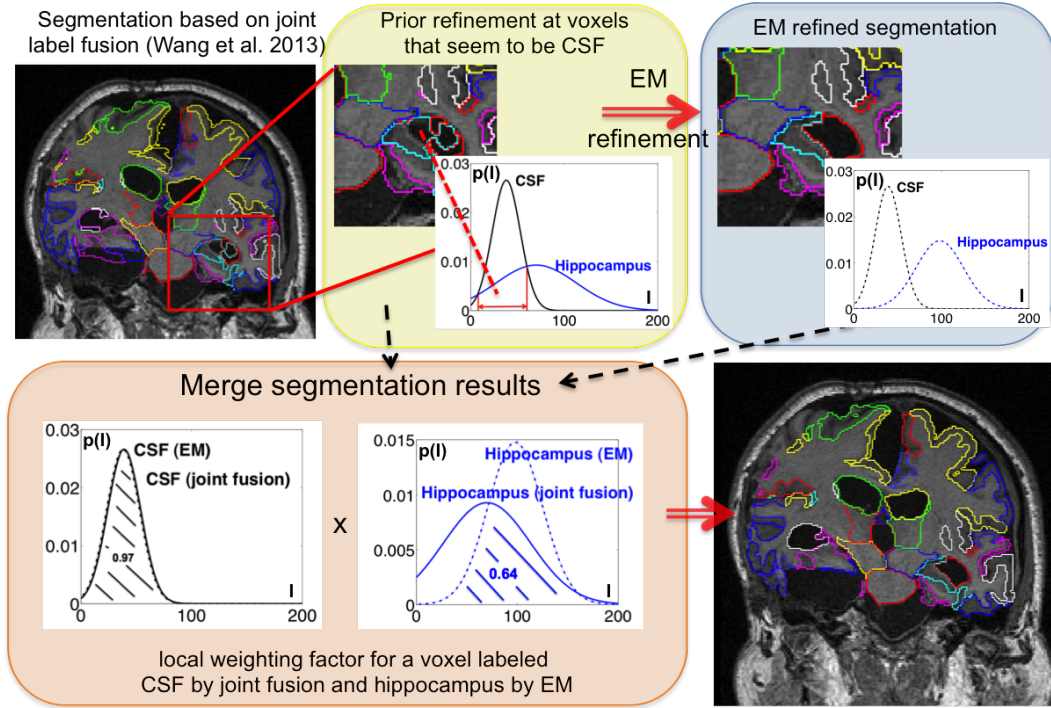


Figure 3.3: Schematic process of the segmentation refinement using prior relaxation, EM-optimisation and spatially weighted combination of probabilistic label maps on the example of the hippocampal region. If registration consistently fails joint label fusion tends to label a significant number of voxels belonging to the inferior lateral ventricle as hippocampus. These wrongly labelled low-intensity voxels lead to a high variance of the estimated intensity distribution within the hippocampus label (top left). The red interval (top left intensity distribution) indicates for which voxels prior relaxation will be carried out. EM-refinement then allows correction of the mislabelled CSF voxels leading to a sharper intensity distribution within the hippocampus (top right). The segmentations obtained using label fusion and EM-optimisation are finally merged into a consensus segmentation (bottom right). This combination is based on spatially varying weights that are calculated based on the overlap of intra-label intensity distributions (bottom left).

The whole pipeline including registration, label fusion and prior relaxation is illustrated in Figure 3.2 and Figure 3.3.

3.2.2.5 Intensity-based label refinement through expectation-maximisation

The relaxed probabilistic priors Π^R are then refined based on the observed intensities of $\mathbf{I}_{\text{tgt}}$ by employing the widely used EM-optimisation presented by Van Leemput et al. (1999). For the sake of readability and consistency with existing literature the notation is followed which was also used in Van Leemput et al. (1999); Cardoso et al. (2011); Ledig et al. (2012b).

Given the parameters $\Phi = \{(\mu_1, \sigma_1), (\mu_2, \sigma_2), \dots, (\mu_K, \sigma_K)\}$ of K structural classes the likelihood of observing, the log-transformed, intensity y_i at voxel i is given as:

$$f(y_i|\Phi) = \sum_k f(y_i|\mathbf{z}_i = \mathbf{e}_k, \Phi) f(\mathbf{z}_i = \mathbf{e}_k) \quad (3.7)$$

It is commonly assumed that the probability of a voxel i to have intensity y_i , $f(y_i|\mathbf{z}_i = \mathbf{e}_k, \Phi)$, given that it belongs to class k , $(\mathbf{z}_i = \mathbf{e}_k)$, is described by a normal distribution (Wells III et al., 1996; Van Leemput et al., 1999; Zhang et al., 2001; Cardoso et al., 2011). Thus $f(y_i|\mathbf{z}_i)$ is modeled as $f(y_i|\mathbf{z}_i = \mathbf{e}_k, \Phi) = \mathcal{N}_k(y_i)$ where $\mathcal{N}_k$ denotes the Gaussian distribution with corresponding parameters (μ_k, σ_k) . The prior probability $f(\mathbf{z}_i = \mathbf{e}_k)$ that a voxel i belongs to structure k is given by the relaxed version of the probabilistic label estimates after multi-atlas label propagation, Π^R (cf. Section 3.2.2.4).

Smoothness of the final segmentation is enforced with a global and stationary MRF, which is integrated using the mean field approximation (Zhang, 1992), following the example of Van Leemput et al. (1999) and Cardoso et al. (2011). This also allows the incorporation of topological knowledge next to the spatial information provided by the prior estimates Π^R . The MRF is implemented by expanding $f(\mathbf{z}_i = \mathbf{e}_k) = \pi_{ik}^R$ to

$$f(\mathbf{z}_i = \mathbf{e}_k | p_{S_i}^{(q)}, G) = \frac{\pi_{ik}^R e^{-U_{MRF}(\mathbf{e}_k | p_{S_i}^{(q)}, G)}}{\sum_{j=1}^K \pi_{ij}^R e^{-U_{MRF}(\mathbf{e}_j | p_{S_i}^{(q)}, G)}} \quad (3.8)$$

The MRF energy function U_{MRF} is calculated based on the probabilistic label estimates in iteration q , $p_{ik}^{(q)}$, in the first-order neighbourhood of voxel i , S_i , as:

$$U_{MRF}(\mathbf{e}_k | p_{S_i}^{(q)}, G) = \quad (3.9)$$

$$\sum_{j=1}^K G_{kj} \left(\sum_{l \in S_i^x} s_x p_{lj}^{(q)} + \sum_{l \in S_i^y} s_y p_{lj}^{(q)} + \sum_{l \in S_i^z} s_z p_{lj}^{(q)} \right) \quad (3.10)$$

Here, G denotes a $K \times K$ matrix defining the connectivity between class k and j and $s = \{\frac{1}{d_x}, \frac{1}{d_y}, \frac{1}{d_z}\}$ accounts for anisotropic voxel spacing in world coordinates. The connectivity matrix G is defined

as:

$$G(k, j) = \begin{cases} 0, & \text{if } k = j \\ \beta, & \text{if structures } k \text{ and } j \text{ share a boundary} \\ \gamma, & \text{if structures } k \text{ and } j \text{ are distant} \end{cases} \quad (3.11)$$

Here β and γ , with $0 \leq \beta \leq \gamma$, are parameters describing the penalty for certain neighbourhood configurations. By assuming that voxels are statistically independent, the probability of observing an image $\mathbf{I}_{\text{tgt}}$, given that the parameters Φ are known, is given by $f(\mathbf{I}_{\text{tgt}}|\Phi) = \prod_i f(y_i|\Phi)$. This model can be solved by interleaving the expectation of the class probabilities $p_{ik}^{(q)}$ and the maximisation of the model by updating the model parameters $\Phi^{(q)}$. It is assumed that the label probabilities $p_{ik}^{(q+1)}$ are known in iteration $(q+1)$ so that the model parameters can be updated as:

$$\mu_k^{(q+1)} = \frac{\sum_i p_{ik}^{(q+1)} y_i}{\sum_i p_{ik}^{(q+1)}} \quad , \quad \sigma_k^{(q+1)} = \sqrt{\frac{\sum_i p_{ik}^{(q+1)} (y_i - \mu_k^{(q+1)})^2}{\sum_i p_{ik}^{(q+1)}}} \quad (3.12)$$

Given the updated model parameters Φ the estimate of the class probabilities in the next iteration is given as:

$$p_{ik}^{(q+1)} = \frac{f(y_i|\mathbf{z}_i = \mathbf{e}_k, \Phi^{(q)})f(\mathbf{z}_i = \mathbf{e}_k|p_{S_i}^{(q)}, G)}{\sum_{j=1}^K f(y_i|\mathbf{z}_i = \mathbf{e}_j, \Phi^{(q)})f(\mathbf{z}_i = \mathbf{e}_j|p_{S_i}^{(q)}, G)} \quad (3.13)$$

Usually the model converges after a few iterations. In this chapter ten iterations were performed to better control the runtime of the algorithm. Furthermore, the parameters for describing the MRF were set to $\beta = 1.0$ and $\gamma = 1.5$.

In order to increase samples for small non-cortical brain structures and thus increase the robustness of the parameter estimate, symmetric brain structures (e.g. hippocampus left/right) are modelled with a single Gaussian distribution. This also allows a more consistent segmentation and thus subsequent volume comparison or symmetry inference. Furthermore, all voxels in cortical brain structures are modelled with a single Gaussian distribution.

3.2.2.6 Weighting scheme for merging fusion-based and EM-based segmentations

Adjacent brain structures often have very similar intensity distributions, which makes an intensity-based refinement of their shared boundary difficult. Examples are adjacent cortical brain regions or the boundary between the hippocampus and the amygdala.

Furthermore, the presented intensity-based optimisation approach tends to calculate well-separated

intensity distributions by reducing the intraclass variances (σ_k^2) within labels. However, this often degrades segmentation results as it does not necessarily reflect manual labelling protocols. For example, the boundary between thalamus and adjacent white matter is determined by geometric characteristics, rather than by intensity characteristics (Hammers et al., 2003). As the boundaries of subcortical structures are not only defined by intensities, their intensity profile tends to have a wider spread (larger σ_k). This has been observed for subcortical structures such as the thalamus, caudate or putamen. For instance, the standard EM-optimisation is likely to relabel high intensity, 'thalamus voxels' in the vicinity of the thalamus/white matter boundary as 'white matter'. This results in a reduced intraclass variance for the thalamus class and higher model likelihood. The segmentation accuracy is reduced, however, as the estimated boundary does not represent the manual protocol of the expert rater.

While intensity-based EM-refinement often provides little or no value for subcortical regions of healthy brains, it is a powerful technique to correct consistent registration failures, *e.g.* in the hippocampal region or in cortical regions. Here, the label estimate of a certain class k often contains several types of brain tissue resulting in large intraclass variance. These intensity distributions can effectively be optimised using the intensity-based refinement as described in Section 3.2.2.5.

In this work it is proposed to not rely exclusively on either label fusion or EM-refined fusion, but to combine their probabilistic estimates into a common segmentation. Here, a simple global formulation of the proposed model is described before introducing the spatially variant extension in Section 3.2.2.7.

Depending on a global weighting factor $\Gamma \in [0; 1]$ the spatial priors obtained by label fusion Π and the posteriors calculated based on intensity-based EM-optimisation Z are combined.

Specifically the final posterior probability z_{ik} is combined with the spatial prior π_{ik} to calculate a new probabilistic estimate z_{ik}^{merged} as:

$$z_{ik}^{merged} = (1.0 - \Gamma)z_{ik} + \Gamma\pi_{ik} \quad (3.14)$$

3.2.2.7 Locally varying weighting parameter Γ

A straightforward extension of this formulation is to model the weighting parameter Γ dependent on the spatial position i . This allows a variable weight for the contribution of either registration-driven

(multi-atlas label propagation) or intensity-driven (EM-refinement) label estimates in the final segmentation. With the known characteristics of EM-refined results (*cf.* Section 3.2.2.6) in mind, the aim is to formulate a model, which favours the geometry-driven and registration-based priors over the intensity-based refinement if there are no indications of substantial registration failures. On the other hand, the model must be flexible and consider the intensity-refined posterior probabilities if it is assumed that the multi-atlas propagation failed. This is often observed if the subject of interest shows severe structural changes of the brain due to disease related atrophy, traumatic deformation, or surgical resection of brain tissue (*cf.* Figure 3.1).

The basic assumption for an automatic choice of Γ is that the EM-refined segmentation Z should get a higher weight with increasing deviation from the segmentation obtained through label fusion Π .

Here, it is assumed that if a label has a similar intensity distribution before and after the intensity-based refinement, the result obtained through the label fusion Π is reliable for this label, and thus Γ_i should be close to 1. In contrast, if for example the hippocampal label in Π erroneously contains ventricular CSF, the intensity distribution has a rather large standard deviation, because the label contains two tissue types. However, after intensity-based refinement the intensity distribution of the label in Z is rather sharp, because the mislabelling of CSF is corrected due to the intensity-based refinement. In this case – two or more intensity distributions within a label based on Π and Z – the aim is to set $\Gamma_i \ll 1$. This means that the more the EM-refined segmentation deviates from the prior the more it contributes to the final segmentation estimate.

To model this behaviour, Γ_i is chosen dependent on the most likely labels assigned to a certain voxel by the label fusion, $k_{\max,i} = \arg \max_k \pi_{ik}$, and the EM-refinement, $z_{\max,i} = \arg \max_k z_{ik}$. Specifically, the overlap of the normal distributions is used which are estimated on label $z_{\max,i}$ in both Z and Π , and for $k_{\max,i}$ accordingly. Specifically, Γ_i is calculated as:

$$\Gamma_i = \int_{-\infty}^{+\infty} \min(\mathcal{N}_{k_{\max,i}}^{\Pi}(y), \mathcal{N}_{k_{\max,i}}^Z(y)) dy \int_{-\infty}^{+\infty} \min(\mathcal{N}_{z_{\max,i}}^{\Pi}(y), \mathcal{N}_{z_{\max,i}}^Z(y)) dy \quad (3.15)$$

This weighting approach is exemplified in Figure 3.3. In this example, the common scenario is shown, in which joint label fusion labels a voxel as hippocampus, and the intensity-based refinement approach labels the same voxel as CSF. The two images on the right in Figure 3.4 illustrate exemplary weights $(1 - \Gamma_i)$ for MR images of brains showing little and substantial structural changes.

3.3 Experiments and results

The goal of this work was to devise a robust segmentation framework that can be employed to segment brain MR images with potentially changed brain structure. Specifically the aim was to segment a cohort of TBI patients and to extract biomarkers that can be correlated with clinical variables.

However, before applying the proposed methodology to clinical data in Section 3.3.2 quantitative experiments were conducted that investigate the method’s performance and characteristics on a well-studied benchmark dataset. For this dataset reference labels, which were manually annotated by experts, are available. This allows the calculation of label overlaps, the conduction of a test-retest analysis and the comparison of the method to other state-of-the-art approaches in Section 3.3.1.

3.3.1 Quantitative evaluation on a benchmark dataset using manual labels

For evaluating MALP-EM, the dataset provided in the course of the “MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling” (Landman and Warfield, 2012) was used (*cf.* Section 2.2.5 and Section 3.2.1.2). The dataset consists of 35 T_1 -weighted MR images with corresponding labels created manually by experts. As in the Grand Challenge, the cohort was divided into a training set of 15 subjects (10 female, 5 male, age: 23 ± 4.3 (mean $\pm$ SD) years, minimum age 19, maximum age 34) and a test set of 15 subjects (10 female, 5 male, age 45.7 ± 24.4 , minimum age 18, maximum age 90). Including the 5 repeat scans, the test set consists of 20 images. These subsets are also summarised in Table 2.3 in Section 2.2.5.

3.3.1.1 Label overlaps

In total seven different approaches were compared:

- MAPER: Standard MAPER with MVF as proposed by Heckemann et al. (2010);
- MALP-LWF (Multi-Atlas Label Propagation with LWF): Segmentations obtained through locally weighted label fusion using GSSD as similarity measure (*cf.* Section 3.2.2.3) (Artaechevarria et al., 2009; Ledig et al., 2012b).

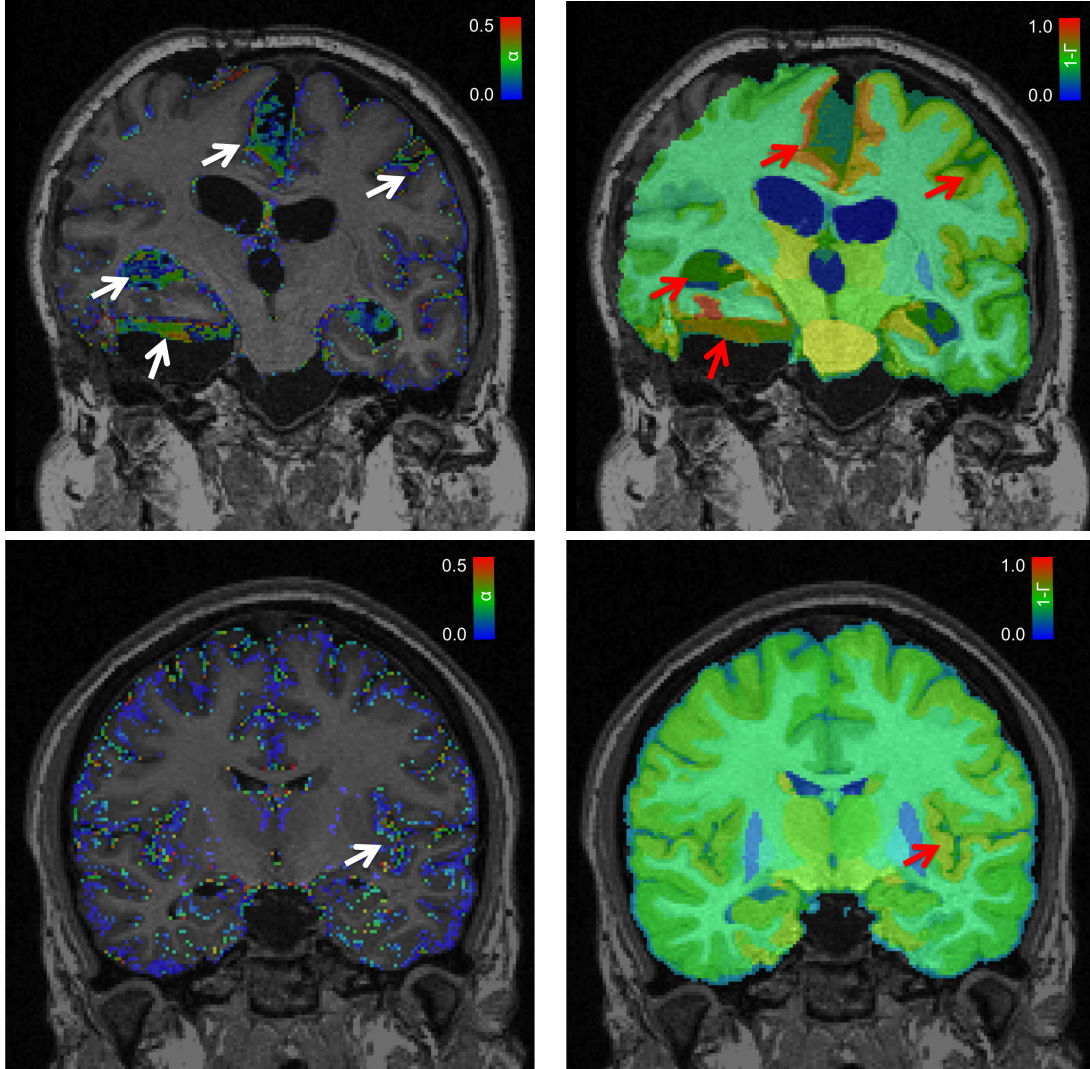


Figure 3.4: Illustration of the relaxation weights (left column) and the spatially varying combination weights as $1 - \Gamma_i$ (right column) for the two example subjects in Figure 3.8 (top row, subject with pathological changes) and Figure 3.10 (bottom row, subject with a close-to-normal brain configuration). It is noted that in cortical regions and regions where the label fusion fails, both increased prior relaxation (white arrows) and combination weights (red arrows) favouring EM-based segmentations are apparent.

- MALP-JLF (Multi-Atlas Label Propagation with JLF): Segmentations obtained through joint label fusion using the implementation of Wang et al. (2013b) with standard parameters².
- MALP-EM $_{\Gamma_i}^{LWF}$ (Multi-Atlas Label Propagation with Expectation-Maximisation based refinement): MALP-LWF followed by the proposed prior relaxation (Section 3.2.2.4), the EM-refinement (Section 3.2.2.5) and the spatially varying merging strategy (Section 3.2.2.7);
- MALP-EM $_{\Gamma_i}$: MALP-JLF followed by the proposed prior relaxation (Section 3.2.2.4), the EM-refinement (Section 3.2.2.5) and the spatially varying merging strategy (Section 3.2.2.7);
- MALP-EM $_{\Gamma_i}$ -BC: MALP-EM $_{\Gamma_i}$ with additional learning-based segmentation bias correction² as proposed by Wang et al. (2011a). This is considered as the setup yielding the highest accuracy on the benchmark dataset.
- PICSL-BC (PICSL research group - Bias Correction): best performing method in the Grand Challenge (Landman and Warfield, 2012) using SyN registration from the ANTs software with JLF (Wang et al., 2013b) and bias correction (Wang et al., 2011a).

As discussed in Section 2.1.3 and Section 3.2.2.2 the two configurations based on locally weighted fusion, MALP-LWF and MALP-EM $_{\Gamma_i}^{LWF}$, are expected to be outperformed by the approaches based on JLF. However, as they are substantially more computationally efficient the aim is to investigate whether they could be a viable alternative in practice. An overview over the compared methods is provided in Table 3.3.

Table 3.3: Overview over all compared methods and their respective building blocks.

Method	Registration	Label Fusion	Additional Processing
MAPER	MAPER (<i>cf.</i> Sec. 3.2.2.2) (Heckemann et al., 2010)	Majority Voting	none
MALP-LWF	MAPER	LWF (<i>cf.</i> Sec. 3.2.2.3)	none
MALP-JLF	MAPER	JLF (Wang et al., 2013b)	none
MALP-EM $_{\Gamma_i}^{LWF}$	MAPER	LWF	proposed (<i>cf.</i> Sec. 3.2.2.4, 3.2.2.5, 3.2.2.6, 3.2.2.7)
MALP-EM $_{\Gamma_i}$	MAPER	JLF	proposed (<i>cf.</i> Sec. 3.2.2.4, 3.2.2.5, 3.2.2.6, 3.2.2.7)
MALP-EM $_{\Gamma_i}$ -BC	MAPER	JLF	proposed + bias correction (Wang et al., 2011a)
PICSL-BC	ANTs (SyN) (Avants et al., 2008)	JLF	bias correction

Each of the 20 test images was segmented into 134 regions. To evaluate the quality of the respective segmentations, Dice overlaps (SIs, Dice (1945)) were calculated with the available manual segmentations. SI values for the different segmentation methods are shown in Table 3.4. Individual SI values of non-cortical structures are shown in Table 3.5.

²Implementation publicly available at https://www.nitrc.org/projects/picsl_malf/ (Version 1.2, last accessed: 8 December 2014).

The results were also compared to the best performing method in the Grand Challenge (Landman and Warfield, 2012) called PICS�-BC. PICS�-BC employs the joint label fusion presented in Wang et al. (2013b) and the learning-based wrapper method presented in Wang et al. (2011a) where segmentation bias with respect to the gold-standard segmentations is learned. Moreover, PICS�-BC has been evaluated on the same images (including the same split into training and test images) using the same 134 regions. No significant differences (Student’s two-sided paired t-test, $p > 0.01$) were found between the proposed flexible MALP-EM $_{\Gamma_i}$ and PICS�-BC for the averaged similarity indices over all classes. Applying additional segmentation bias correction (Wang et al., 2011a) to MALP-EM $_{\Gamma_i}$ significantly ($p < 10^{-4}$) improved segmentation results. Since normal distribution cannot be assumed for similarity indices, hypothesis testing was repeated using the non-parametric Wilcoxon signed-rank test. All differences shown in Table 3.4 remained significant at least at $p < 10^{-2}$.

Furthermore, it was found that MALP-JLF performs significantly better than MALP-LWF. However, employing the proposed prior relaxation and intensity refinement the difference in accuracy is less pronounced. No significant differences could be found for subcortical structures between MALP-EM $_{\Gamma_i}^{\text{LWF}}$ and MALP-EM $_{\Gamma_i}$. Given the substantially reduced computation time (locally weighted fusion (GSSD): ≈ 15 minutes; joint label fusion: ≈ 7 hours), LWF is a potential alternative in practice. In the following of this chapter the more accurate setup, MALP-EM $_{\Gamma_i}$ based on joint label fusion, will be investigated.

Table 3.4: Similarity indices (SI) in % averaged (unweighted) over 20 subjects for all 36 non-cortical classes, all 98 cortical classes and all 134 classes. The methods that were compared are MAPER using majority vote, MALP-LWF using locally weighted fusion, MALP-JLF using joint label fusion, MALP-EM $_{\Gamma_i}^{\text{LWF}}$ /MALP-EM $_{\Gamma_i}$ with a local merging strategy and optional segmentation bias correction (BC), and PICS�-BC. *, ** = significantly (*: $p < 10^{-2}$; **: $p < 10^{-4}$) different to the method in the column to the left. bold = significantly best results.

	MAPER	MALP-LWF	MALP-JLF	MALP-EM $_{\Gamma_i}^{\text{LWF}}$	MALP-EM $_{\Gamma_i}$	MALP-EM $_{\Gamma_i}$ -BC	PICS�-BC
SI 36 subcort.	82.0	82.5*	82.7**	82.9	82.9	83.4*	83.8
SI 98 cort.	72.4	72.9**	73.2**	73.7*	73.8*	74.9**	73.9*
SI 134 classes	74.9	75.5**	75.8**	76.1*	76.3*	77.2**	76.5*

3.3.1.2 Evaluation of the influence of the weighting factor Γ

As illustrated in Figure 3.5, a weighting factor of $\Gamma = 0.8$ yields the best segmentation results on the training dataset. This result shows that joint label fusion yields accurate labels for the majority

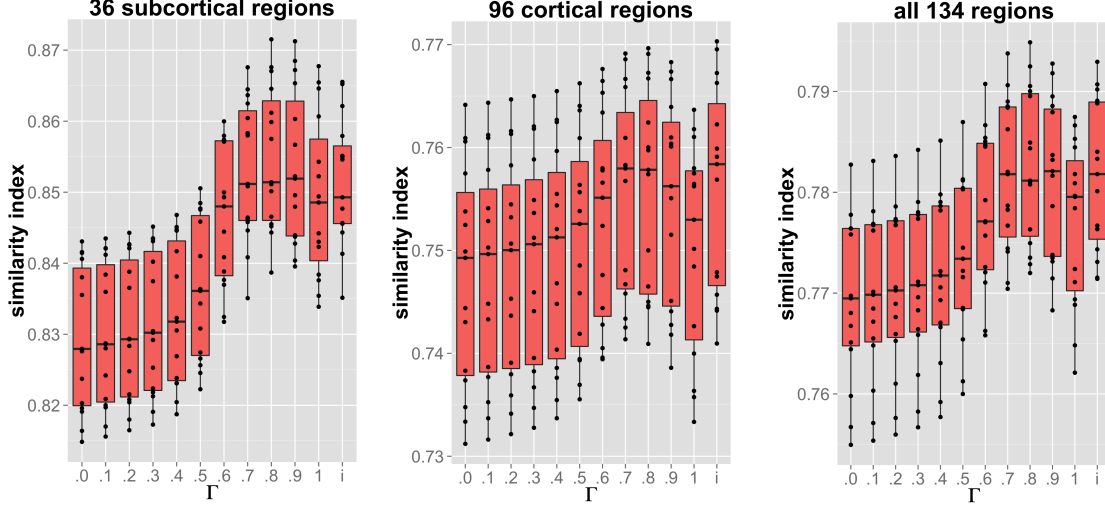


Figure 3.5: MALP-EM $_{\Gamma}$ applied to the 15 training images with varying Γ . Here $\Gamma = 1.0$ is equivalent to exclusive joint label fusion and i denotes the spatially varying choice of Γ_i . Median similarity index with first and third quantiles of 36 non-cortical structures (left), 98 cortical structures (middle) and all 134 regions (right).

of voxels. For voxels with a high uncertainty (more than one label has a high non-zero probability) the EM-refined result should be considered as additional weighting. Using the more flexible model with a spatially varying Γ_i , comparable overlaps to $\Gamma = 0.8$ are observed. This is encouraging, because a global and fixed Γ leads to a stricter model that is assumed to perform well on healthy, normal data while only a data-driven choice of Γ_i allows the flexibility to cope with MR images of potentially injured brains of TBI subjects.

3.3.1.3 Test-retest reliability

To investigate the consistency of segmentations calculated with the proposed method, its test-retest reliability was assessed. This characteristic was quantified using the 5 subjects in the set for whom repeat images are available. The intraclass correlation coefficient (ICC) (two-way random single measures, absolute agreement) following Shrout and Fleiss (1979) was used. Specifically, the reproducibility was assessed of label volumes calculated on images of the same subject at different time points (scan interval less than 90 days). The assumption is that brains of healthy subjects do not change substantially within short periods. ICC is widely used to quantify test-retest reliability (Kempton et al., 2011; Nugent et al., 2013).

On the manual segmentations an average ICC of 0.80 ± 0.28 for non-cortical and 0.78 ± 0.25 for cortical regions was calculated. Using the proposed method MALP-EM $_{\Gamma}$, an average ICC of

0.97 ± 0.04 for non-cortical and 0.94 ± 0.08 for cortical regions was obtained. These results are slightly better than the average ICC obtained using joint label fusion only (non-cortical: 0.96 ± 0.06 , cortical: 0.94 ± 0.09).

Furthermore, the relative volume difference (Δ_{vol}) between a structure's volume at two time points, V_{t_1} and V_{t_2} , was assessed. In this work the relative volume difference is defined as:

$$\Delta_{vol} = 100\% \frac{|V_{t_1} - V_{t_2}|}{0.5(V_{t_1} + V_{t_2})} \quad (3.16)$$

On the manual segmentations an average Δ_{vol} of $8.3 \pm 7.5\%$ for non-cortical and $12.3 \pm 8.4\%$ for cortical regions was calculated. Using the proposed method MALP-EM $_{\Gamma_i}$ an average Δ_{vol} of $2.4 \pm 1.4\%$ for non-cortical and $4.1 \pm 2.3\%$ for cortical regions was obtained. These results are similar to the average Δ_{vol} obtained using joint label fusion only (non-cortical: $2.7 \pm 2.5\%$, cortical: $3.8 \pm 2.2\%$). More extensive quantitative results can be found in Table 3.5.

Table 3.5: SIs in % averaged over 20 test images. ICC (two-way random single measures, absolute agreement, Shrout and Fleiss (1979)) and relative volume differences (Δ_{vol}) based on 5 subjects with available repeat scans for all 36 considered non-cortical structures. Significantly different SI compared to column to the left indicated at $p < 10^{-2}$ (*) and $p < 10^{-4}$ (**).

	similarity index (SI)				intraclass correlation (ICC) / relative volume difference (Δ_{vol})		
	MAPER	MALP-JLF	MALP-EM $_{\Gamma_i}$	MALP-EM $_{\Gamma_i} - BC$	Manual	MALP-JLF	MALP-EM $_{\Gamma_i}$
3rd ventricle	85.2	85.6	79.8**	85.8**	0.855 / 17.8±11.4%	0.994 / 3.6±2.4%	0.978 / 5.9±3.3%
4th ventricle	86.7	87.3*	87.1	87.1	0.991 / 3.0±2.3%	0.999 / 1.5±0.7%	0.998 / 1.6±1.6%
accumbens area R	77.9	78.4	76.7	78.0	0.673 / 13.8±10.5%	0.961 / 3.6±2.6%	0.984 / 2.0±2.6%
accumbens area L	77.3	77.3	77.6	78.7	0.755 / 11.2±10.3%	0.728 / 7.1±6.9%	0.867 / 5.9±2.4%
amygdala R	80.2	80.6	79.4	81.4*	0.605 / 10.9±14.0%	0.893 / 3.6±2.6%	0.929 / 3.7±1.7%
amygdala L	81.3	81.9	81.7	83.1*	0.897 / 6.0±4.1%	0.925 / 4.4±3.2%	0.982 / 2.5±2.1%
brain stem	93.7	93.8*	93.7	94.0**	0.957 / 3.5±1.8%	0.993 / 1.1±0.7%	0.993 / 1.2±0.7%
caudate R	87.0	87.5	86.8	87.9*	0.955 / 3.1±1.8%	0.992 / 1.2±1.0%	0.991 / 1.4±0.9%
caudate L	87.0	87.8*	87.4	88.0	0.956 / 3.4±2.4%	0.998 / 0.6±0.3%	0.996 / 0.8±0.5%
cerebellum exterior R	92.5	92.9**	93.4**	93.5	0.985 / 2.1±1.4%	0.993 / 1.2±1.7%	0.995 / 1.2±1.4%
cerebellum exterior L	92.1	92.6**	93.1*	93.2	0.978 / 2.2±2.2%	0.989 / 1.6±1.9%	0.988 / 2.0±1.5%
cerebellum white matter R	88.9	89.1	90.5**	89.9	0.978 / 3.8±2.6%	0.992 / 1.4±1.7%	0.995 / 1.3±1.0%
cerebellum white matter L	89.0	89.2*	90.6**	90.2	0.884 / 5.3±5.3%	0.972 / 2.1±2.1%	0.975 / 2.2±1.7%
cerebral white matter R	93.3	93.3	93.7*	94.0	0.952 / 3.1±1.8%	0.992 / 1.1±1.2%	0.993 / 1.0±1.0%
cerebral white matter L	93.2	93.2	93.6*	93.9	0.969 / 2.2±1.9%	0.995 / 0.9±0.9%	0.996 / 0.8±0.8%
cerebrospinal fluid	77.7	79.8	77.3*	81.0**	0.794 / 11.6±9.9%	0.859 / 4.7±3.9%	0.824 / 4.0±6.1%
hippocampus R	85.1	86.3*	86.4	86.8	0.815 / 8.4±7.3%	0.997 / 0.9±0.5%	0.991 / 1.5±1.0%
hippocampus L	85.2	86.5**	86.3	86.9*	0.870 / 7.6±7.6%	0.996 / 1.0±0.5%	0.996 / 0.9±1.1%
inf lat ventricle R	55.6	63.2**	68.8**	70.9	0.794 / 20.4±8.6%	0.993 / 4.4±3.3%	0.992 / 3.8±2.8%
inf lat ventricle L	55.5	62.2**	67.1**	67.2	0.983 / 12.9±7.6%	0.991 / 5.9±4.7%	0.996 / 4.1±3.8%
lateral ventricle R	91.9	92.5*	93.0	93.0	0.999 / 9.4±8.3%	1.000 / 2.6±1.6%	1.000 / 3.1±2.1%
lateral ventricle L	92.3	93.0*	93.5	93.3	0.999 / 9.1±7.0%	1.000 / 2.8±1.6%	1.000 / 2.4±1.7%
pallidum R	86.6	87.5**	87.6	87.6	0.468 / 8.8±1.4%	0.948 / 2.4±1.4%	0.937 / 2.8±1.7%
pallidum L	85.1	86.7**	86.6	86.7	0.640 / 4.7±3.3%	0.917 / 2.2±1.7%	0.914 / 2.1±2.4%
putamen R	91.0	91.3	91.1	90.8	0.961 / 2.9±1.5%	0.995 / 1.0±0.6%	0.990 / 1.4±1.0%
putamen L	90.8	91.1	91.1	91.0	0.978 / 2.0±1.8%	0.980 / 1.4±1.8%	0.972 / 1.6±2.2%
thalamus proper R	91.7	92.1*	91.4*	92.0**	0.950 / 2.4±1.6%	0.990 / 1.0±0.9%	0.985 / 1.4±0.9%
thalamus proper L	91.9	92.1	91.5*	91.9*	0.876 / 3.5±3.0%	0.988 / 0.9±0.9%	0.978 / 1.4±1.0%
ventral DC R	88.6	88.8*	88.1**	88.9**	0.835 / 4.6±2.9%	0.996 / 1.0±0.3%	0.988 / 1.8±0.8%
ventral DC L	88.7	88.7	88.1*	88.9*	0.862 / 4.8±2.8%	0.994 / 1.0±0.7%	0.985 / 1.6±1.4%
optic chiasm	52.0	49.1	53.9	43.5	-0.355 / 40.5±38.4%	0.801 / 14.4±10.3%	0.947 / 5.9±3.5%
cerebellar vermal lobules I-V	81.9	82.3	82.7	83.2	0.470 / 11.5±11.1%	0.993 / 2.3±1.3%	0.990 / 2.6±1.8%
cerebellar vermal lobules VI-VII	77.0	77.8	78.1	79.7**	0.867 / 4.9±2.2%	0.902 / 3.1±1.9%	0.859 / 3.5±2.8%
cerebellar vermal lobules VIII-X	87.0	87.5	87.6	87.7	0.962 / 5.0±3.8%	0.995 / 1.6±1.1%	0.998 / 1.2±0.8%
basal forebrain R	43.9	44.5	44.6	47.2	0.722 / 12.2±6.9%	0.880 / 3.7±2.8%	0.945 / 2.1±2.0%
basal forebrain L	45.4	45.0	45.0	46.8	0.040 / 18.8±7.5%	0.926 / 3.8±3.4%	0.887 / 3.5±1.8%

3.3.2 Segmentation and analysis of a TBI cohort

MALP-EM_T (in the following referred to as MALP-EM) was used to automatically segment 125 MR brain scans of TBI subjects with potential pathology. This cohort is described in Section 3.2.1.1. For the segmentation all available atlas datasets were used, except the 5 repeat images, *i.e.*, a total of 30 (*cf.* dataset ATLAS in Table 2.3 in Chapter 2).

All MR images were corrected for intensity inhomogeneities using the N4 algorithm (Tustison et al., 2010). The images were further brain extracted with a fully-automatic in-house brain extraction method called Extended Tissue Classification. In this method, first, an EM classifier based on Van Leemput et al. (1999) was applied for producing a coarse tissue segmentation, which is used as an initial brain mask. Thereafter, a deformable model-based approach combined with morphological operations was applied to tune the mask.

3.3.2.1 Quantitative evaluation on TBI datasets using expert validation scores

A specific scoring protocol was devised by J. C. Lopez (JCL), R. A. Heckemann (RAH) and A. Hammers (AH) to semi-quantitatively assess the quality of automatically generated segmentations of TBI images. The protocol was developed for four paired regions that frequently show morphological change in patients with TBI (hippocampus, thalamus, putamen, and occipital pole). Thalamus, putamen and occipital cortices were selected because these structures are frequently implicated in TBI and its sequelae (Warner et al., 2010a; Strangman et al., 2010; Ramlackhansingh et al., 2011). The hippocampus was added because it is a challenging structure to segment (*cf.* Figure 3.1). Consideration of the hippocampus is biologically justified, as it is typically involved in dementia, which in turn is a frequent long-term consequence of severe TBI. The protocol calls for the raters to assign a score on a six point scale (0, worst to 5, best). Three experienced raters (JCL, 1 year of clinical service; RAH, 12 years of clinical service; AH, 16 years of clinical service) developed the protocol by consensus, using 9 images with corresponding segmentations calculated with both MALP-EM and joint label fusion (MALP-JLF). All raters had basic (JCL) or advanced (RAH, AH) training in neuroanatomy, neuropathology, radiology, and neuroimaging. The 9 images had been selected from the TBI cohort using an ad-hoc approach that ensured that the sample was broadly representative (MCS 2-6; 4 baseline and 5 follow-up scans). The detailed protocol is published as supplementary material to Ledig et al. (2015).

Table 3.6: Mean (standard deviation (SD)) of the expert scores for the assessed segmentation quality of hippocampus, thalamus, putamen and occipital pole. Significant improvement is indicated (*: $p < 0.05$; **: $p < 10^{-4}$). ICC (two-way random single measures, absolute agreement) between all available raters.

	hippocampus		thalamus		putamen		occipital pole	
	MALP-JLF	MALP-EM _{T_i}	MALP-JLF	MALP-EM _{T_i}	MALP-JLF	MALP-EM _{T_i}	MALP-JLF	MALP-EM _{T_i}
rater A	2.81(1.30)	3.23(1.48)	2.42(1.27)	3.12(1.63)*	2.81(2.25)	3.04(2.20)	3.42(1.39)	3.81(1.58)*
rater B	3.00(1.13)	3.23(1.03)	2.88(1.07)	3.62(1.20)**	3.81(1.30)	3.96(1.37)	2.58(0.90)	3.08(0.93)*
rater C	2.04(0.96)	2.65(0.89)*	2.00(0.69)	3.08(1.02)**	2.85(1.19)	3.19(1.27)*	-	-
all raters	2.62(1.20)	3.04(1.18)**	2.44(1.09)	3.27(1.32)**	3.15(1.69)	3.40(1.69)*	3.00(1.24)	3.44(1.33)*
ICC	0.64		0.53		0.67		0.23	

Based on this protocol the three independent raters assessed 13 images using the tool `rview` from the Image Registration Toolkit (<https://github.com/BioMedIA/IRTK>, last accessed: 8 December 2014) (IRTK). All raters were blind to the method (MALP-JLF or MALP-EM). Results of both methods were presented in a balanced, randomised fashion. The raters viewed both methods' results of each subject back to back in order to break the tie if the two scores were equal. None of the raters was directly involved in the development of MALP-EM. The set of test scans did not intersect with the set of scans used for protocol development. Ten scans were randomly chosen with the constraint that five scans be of subjects with $MCS < 4$ and five with $MCS \geq 4$. In addition, a non-rater chose three further subjects to ensure scans with severe pathology were represented in the evaluation set. Structures in the left and right brain hemisphere were rated separately.

The findings obtained through this expert validation confirm that MALP-EM is superior to JLF in the segmentation of brain images of TBI patients with severe pathology. The average expert scores are shown in Table 3.6. The distribution of the scores for the individual structures is shown in Figure 3.6. Further, Figure 3.7 shows the fraction of test images on which a method performs better than the other.

To further assess interrater variance the ICC (two-way random single measures, absolute agreement) between the available raters was calculated. The calculated ICCs are summarised in Table 3.6. Moderate interrater agreement was observed between all three raters for hippocampus, thalamus and putamen. The interrater agreement for the occipital pole, which was rated by two raters of rather different experience, is at a lower level. However, both raters agreed that MALP-EM yields significantly improved results on this structure.

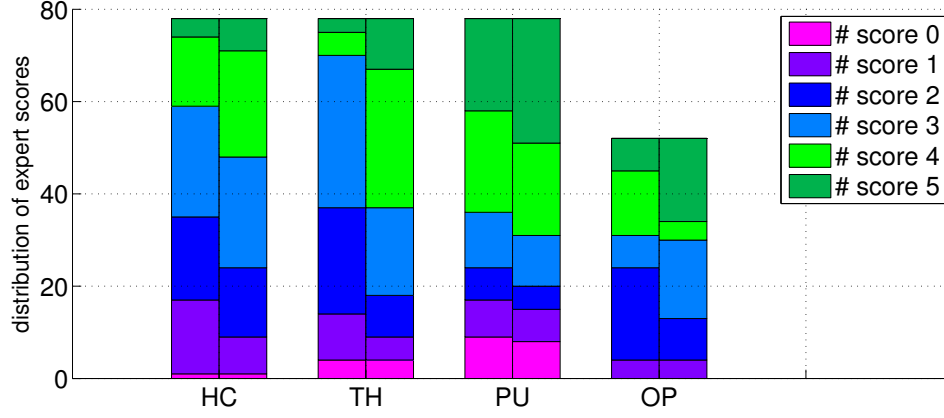


Figure 3.6: Distribution of expert scores for investigated structures. Note that structures in the left/right brain hemisphere were rated separately. This results in 78 ratings per structure ($3 \text{ raters} \times 13 \text{ subjects} \times 2 \text{ hemispheres}$). Occipital pole was not rated by rater C. Comparison of MALP-JLF (left bars) and MALP-EM (right bars).

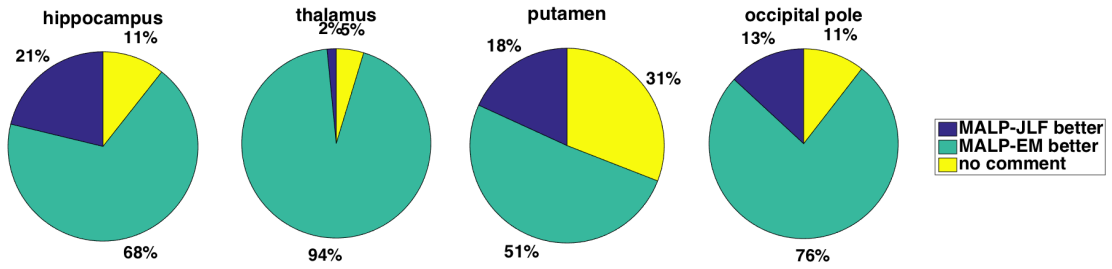


Figure 3.7: Fraction of test images on which a method performs better than the other³. Raters B and C were asked to rate which segmentation was better, even if scores were tied.

³Fractions might not exactly add up to 100% due to rounding effects.

3.3.2.2 Qualitative confirmation of the robustness using MALP-EM on images with pathology

Segmentation quality was considered sufficient if no major parts of the brain were missing due to a overly restrictive brain extraction, and the cortical grey matter/white matter and visually dominant non-cortical boundaries (e.g. ventricles/grey matter) were matched by label boundaries. Inclusion or exclusion of structures that are not present in the atlases (e.g. lesions or contusions) was not regarded as failure. In a small subset ($\approx 5 - 10\%$) of the processed images, local inaccuracies in the shape of brain extractions were accepted in the cortical region, e.g. Figure 3.8, if most cortical and especially subcortical structures were segmented reasonable, considering the pathology.

Five segmentations of insufficient quality were identified after visual inspection. One failure originated in misregistration due to significant intensity inhomogeneities that remained after the

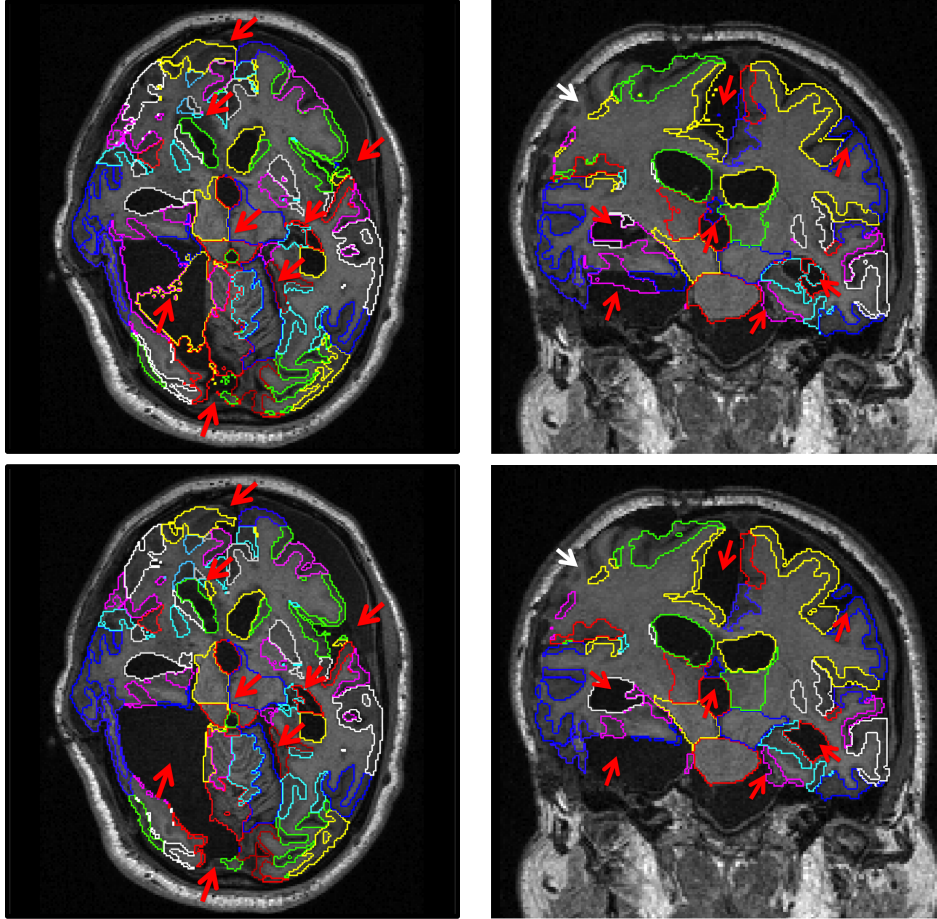


Figure 3.8: Segmentations obtained with MALP-JLF (top) and MALP-EM (bottom) of a subject with a markedly changed brain configuration on MRI (male, age: 45 years, GOS = 3, follow-up MRI, taken one month after injury, axial(left)/coronal(right) view). While the label fusion clearly fails to match various intensity boundaries, for example in the hippocampal region, MALP-EM is able to correct this problem to a large extent due to the strong intensity contrast of CSF and grey matter structures. Red arrows highlight improvements obtained using MALP-EM over pure joint label fusion. White arrows highlight errors due to inconsistent brain extraction.

N4 bias correction. On this single subject the bias correction was reapplied using the generated brain mask to further reduce inhomogeneities. Registration and segmentation were subsequently successful; therefore the image was retained. For another three subjects the generated brain mask was of insufficient quality. One of the corresponding scans was acquired from a subject with a follow-up image for which the brain extraction was fine. Thus the brain mask of the follow-up time point was used to extract the brain at the acute stage. The subsequent segmentation result was satisfactory. The remaining two subjects (*cf.* Figure 3.9), which showed a substantially changed brain structure and the MR images had very poor quality, could not be processed. A single image

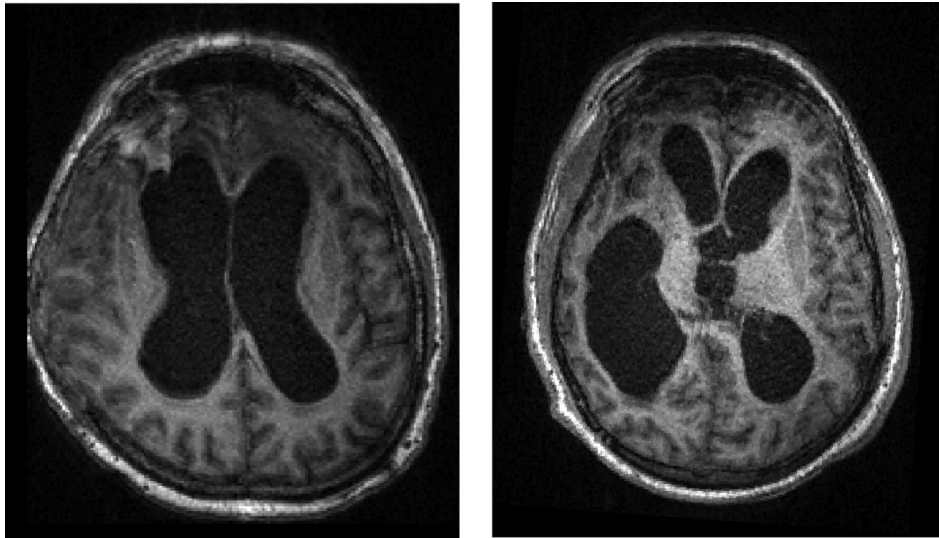


Figure 3.9: Axial slices of the two subjects for which the brain extraction was not successful. left: male, age: 56, GOS = 3, MCS = 5, follow-up MR image taken two months after injury, right: male, age: 16, GOS = 3, MCS = 5, follow-up MR image taken seven months after injury.

was excluded due to consistency problems in the NIfTI format file after image conversion. In conclusion, none of these failures were directly related to MALP-EM. Overall 122 out of 125 available scans could be processed successfully.

On visual inspection, all segmentations of these 122 images were considered reasonable, allowing for pathology. Visual examples of calculated segmentation results are shown in Figure 3.8. The image pair illustrates the advantages of MALP-EM over sole label fusion in images with substantial pathology. Figure 3.8 also clearly reveals improved segmentation results obtained with MALP-EM at boundaries of anatomical regions with large intensity contrast. Improvements in both the hippocampal region and at the cortical grey matter/cerebrospinal fluid boundary are particularly striking.

Labels obtained with MALP-EM on images with little pathology (*cf.* Figure 3.10) are visually convincing for both non-cortical and cortical structures in most instances. When substantial pathology is present (Figure 3.8), some inaccuracies were observed. A frequent problem is unlabelled cortical grey matter due to imperfections of the brain mask. Voxels excluded during brain extraction are not reconsidered during the segmentation process. A subject for which this problem is most striking is illustrated in Figure 3.8. However, this subject was kept for the analysis since even in subjects with significant pathology only a few cortical and no subcortical structures are affected

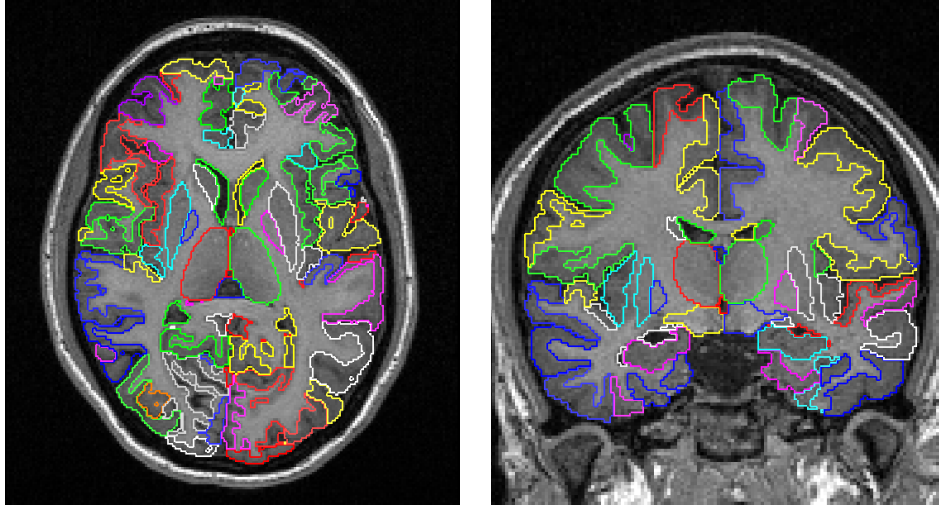


Figure 3.10: Typical segmentation result obtained using MALP-EM $_{\Gamma_i}$ of a subject with a close-to-normal brain configuration in axial(left)/coronal(right) view (female, age: 17 years, GOS = 5, MCS = 2, follow-up MRI, taken nine months after injury).

by this problem. In regions showing severe deformations or atrophy, such as the hippocampal region in the subject shown in Figure 3.8, the nonrigid atlas alignment may consistently fail. Due to the proposed prior relaxation step and spatially varying label combination it was possible to relax this problem and improve upon non-intensity based approaches. It can be concluded from visual comparison that MALP-EM is superior to standard label fusion.

3.3.2.3 Separation of GOS and MCS groups using absolute asymmetry indices

To assess the clinical usefulness of the presented method, the available TBI subjects were clinically classified based on morphometric results. The classification of the brain images performed by an experienced clinician was employed as a gold-standard reference. Correlations with two clinical variables were assessed. First, correlations with a widely used clinical scheme primarily devised for categorising admission X-ray CT images, applied to acute-phase MR images (MCS) and second, one of the most common clinical outcome scores used in TBI (GOS). The focus was on the particularly relevant differentiation between patients who would not be able to live an independent life (GOS < 4) and those with a more favourable outcome. The comparison with the GOS provided a means of estimating the prognostic value of automatically segmented acute-phase MR images. For the MCS, images were additionally dichotomised into those without (MCS < 4 $\hat{=}$ DI I, DI II and DI III) or with (MCS $\geq$ 4 $\hat{=}$ DI IV, EML, NEML) significant mass effect and midline shift.

Table 3.7: Classification results in % obtained separating $MCS < 4$ vs. $MCS \geq 4$ and $GOS < 4$ vs. $GOS \geq 4$ based on absolute asymmetry indices of either acute-phase (MCS and GOS) or follow-up (GOS) MR images only. Results shown are averaged over 1000 cross validation runs. bold = best.

	based on acute-phase MR images						based on follow-up MR images		
	MCS < 4 (Negatives) vs. MCS ≥ 4 (Positives)			GOS < 4 (P) vs. GOS ≥ 4 (N)			GOS < 4 (P) vs. GOS ≥ 4 (N)		
	subcort.	cortical	all	subcort.	cortical	all	subcort.	cortical	all
balanced accuracy	60.8	76.0	72.5	64.7	61.5	61.8	66.8	59.3	62.6
sensitivity	39.6	60.9	56.7	45.4	44.4	44.4	47.4	42.5	42.1
specificity	82.0	91.2	88.2	84.0	81.8	78.5	86.3	76.0	83.1
subjects per group	34 vs. 23			27 vs. 28			19 vs. 43		

MALP-EM was employed for individual and independent cross-sectional experiments at the acute stage (60 subjects) and at the follow-up stage (62 subjects).

As classifier a LDA implemented through the MATLAB function `classify` was used. For validation, 1000 repetitions of a 10-fold cross validation were performed. As feature structural asymmetry was quantified of the paired 63 structures (*cf.* Section 3.2.1.2 and Appendix B). The AAI was employed based on a structure’s volume (V) in the left and right hemisphere, which is defined as:

$$AAI = 100\% \frac{|V_{\text{left}} - V_{\text{right}}|}{0.5(V_{\text{left}} + V_{\text{right}})} \quad (3.17)$$

Specifically, the sum of the absolute asymmetry indices of either all 14 non-cortical structures, all 49 cortical structures or all 63 structures was used.

The results for distinguishing two MCS ($MCS < 4$, $MCS \geq 4$) and GOS ($GOS < 4$, $GOS \geq 4$) groups respectively, using either acute-phase or follow-up images, are summarised in Table 3.7. Since the clinical variables MCS and GOS were missing for 3 (MCS), respectively 5 (GOS) subjects, the number of subjects was reduced in each classification experiment accordingly. As groups were unbalanced the balanced accuracy (Brodersen et al., 2010) (bACC) measure, the average of sensitivity and specificity, was employed to report classification accuracy. Table 3.7 shows that the presented method yields 76% accuracy in distinguishing groups of MCS categories based on acute-phase images. The Marshall classification system is not a linear scale as it takes both midline shift and the size of lesions into account (*cf.* Appendix C). However, in the classification experiment it was possible to discriminate between classes without ($MCS < 4$) and with ($MCS \geq 4$) significant midline shift or mass effect.

Furthermore, it could be estimated from a single acute-phase MRI whether a TBI patient will be able to live an independent life ($GOS \geq 4$) or not with 64.7% accuracy. In comparison, when pre-

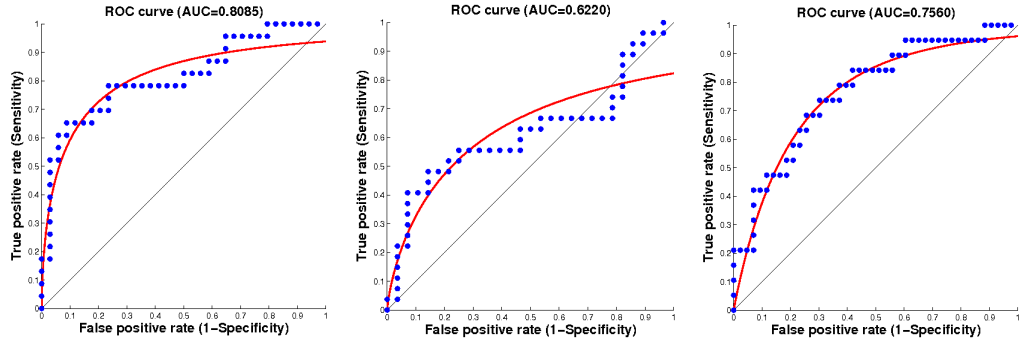


Figure 3.11: Receiver operating characteristic (ROC) curves for classifying subjects according to MCS using the sum of cortical AAI (left), and according to GOS at baseline (middle) and follow-up time point (right) using the accumulated non-cortical AAI.

dicting outcome based on the MCS score at baseline 59.3% accuracy was calculated. Based on the segmentations of non-cortical structures in the follow-up images, 66.8% accuracy was achieved in GOS classification. The classification results are summarised in Table 3.7. The high specificity for MCS classification shows that the presented method does very well in detecting normal appearing brains at the acute stage. The high specificity for GOS classification confirms that the presented approach is able to predict a favourable outcome of a TBI. These findings suggest that structural brain asymmetry could be a sufficient criterion to indicate an unfavourable disease outcome. On the other hand, symmetry seems to be a necessary criterion for favourable disease outcome. It is not, however, a sufficient criterion to rule out an unfavourable outcome. Receiver operating characteristic (ROC) curves for these classification experiments are shown in Figure 3.11.

A detailed summary of results for individual non-cortical structures for both MCS and GOS classification is provided in Table 3.8, Table 3.9 and Table 3.10, including p-values for group separation. These results suggest structural asymmetry of non-cortical brain structures does not correlate well with MCS.

P-values were calculated for group separation using MALP-JLF without the proposed processing. Unlike MALP-EM, this setup did not reveal any significant symmetry differences between GOS groups for the thalamus (at the acute time point) or for the caudate, hippocampus and inferior lateral ventricle (at the follow-up time point). All the structures that show significant symmetry differences between groups of clinical variables in the MALP-JLF setup were also found in the MALP-EM setup.

Table 3.8: Classification results in % (10-fold cross-validation, 1000 runs) obtained separating MCS < 4 vs. MCS ≥ 4 based on AAI of acute-phase MR images. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.01$).

Classification of MCS < 4 (Negatives) vs. MCS ≥ 4 (Positives) (acute scans)							
structure	bACC	SENS	SPEC	mean AAI (SD) for MCS < 4 and MCS ≥ 4		p-value	significance
all non-cortical	60.8	39.6	82.0	175.9 (88.4)	308.9 (231.1)	3.50e-03	++
all cortical	76.0	60.9	91.2	759.3 (151.6)	1072.2 (354.8)	2.74e-05	++
all	72.5	56.7	88.2	935.3 (200.4)	1381.1 (556.7)	7.25e-05	++
accumbens area	62.8	45.8	79.9	16.4 (13.6)	32.7 (35.8)	1.90e-02	+
amygdala	50.6	23.7	77.5	15.2 (16.0)	24.7 (44.2)	2.55e-01	o
caudate	50.6	31.2	69.9	6.9 (7.0)	12.2 (17.8)	1.18e-01	o
cerebellum exterior	66.0	52.6	79.3	4.2 (4.2)	8.7 (7.1)	3.67e-03	++
cerebellum white matter	49.9	39.3	60.4	10.9 (8.4)	15.3 (13.5)	1.37e-01	o
cerebral white matter	48.4	35.1	61.8	2.7 (1.7)	4.6 (7.0)	1.22e-01	o
hippocampus	37.0	8.5	65.6	8.5 (7.0)	12.3 (26.5)	4.21e-01	o
inf lat ventricle	68.5	55.5	81.4	24.0 (19.1)	51.9 (43.7)	1.74e-03	++
lateral ventricle	55.5	36.4	74.6	24.9 (18.7)	47.4 (48.7)	1.78e-02	+
pallidum	59.6	28.1	91.1	9.5 (15.3)	22.0 (31.7)	5.26e-02	o
putamen	57.2	36.2	78.2	11.9 (22.0)	16.3 (23.1)	4.63e-01	o
thalamus	55.0	30.7	79.4	5.6 (8.5)	10.3 (14.5)	1.22e-01	o
ventral DC	38.6	38.0	39.1	8.3 (7.8)	8.2 (5.5)	9.88e-01	o
forebrain	54.9	34.8	75.1	27.1 (24.6)	42.2 (50.3)	1.37e-01	o

Table 3.9: Classification results in % (10-fold cross-validation, 1000 runs) obtained separating GOS < 4 vs. GOS ≥ 4 based on AAI of acute-phase MR images. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.01$).

Classification of GOS < 4 (Positives) vs. GOS ≥ 4 (Negatives) (acute scans)							
structure	bACC	SENS	SPEC	mean AAI (SD) for GOS < 4 and GOS ≥ 4		p-value	significance
all non-cortical	64.7	45.4	84.0	272.8 (213.8)	183.2 (121.4)	6.02e-02	o
all cortical	63.1	44.4	81.8	954.2 (372.6)	821.7 (198.6)	1.04e-01	o
all	61.5	44.4	78.5	1227.0 (556.6)	1004.8 (290.4)	6.77e-02	o
accumbens area	59.7	44.0	75.4	29.7 (33.9)	15.0 (11.0)	3.36e-02	+
amygdala	58.9	33.5	84.3	25.4 (41.4)	13.4 (14.8)	1.56e-01	o
caudate	54.7	37.6	71.9	9.9 (13.3)	7.9 (12.4)	5.64e-01	o
cerebellum exterior	60.0	73.4	46.5	5.2 (5.1)	7.7 (6.6)	1.27e-01	o
cerebellum white matter	43.9	36.0	51.9	13.6 (12.6)	12.8 (9.6)	7.78e-01	o
cerebral white matter	47.2	35.1	59.3	3.8 (6.3)	3.1 (2.5)	5.51e-01	o
hippocampus	51.0	26.3	75.7	12.6 (23.6)	7.8 (9.4)	3.20e-01	o
inf lat ventricle	42.6	34.3	50.8	35.0 (34.8)	33.1 (33.9)	8.34e-01	o
lateral ventricle	56.4	39.8	72.9	39.1 (40.6)	24.3 (30.1)	1.28e-01	o
pallidum	57.8	26.1	89.5	21.4 (29.8)	8.7 (15.5)	5.21e-02	o
putamen	60.5	39.1	82.0	18.6 (28.3)	9.8 (14.6)	1.48e-01	o
thalamus	58.2	34.2	82.1	10.8 (15.1)	4.6 (5.5)	4.61e-02	+
ventral DC	37.4	39.5	35.2	8.4 (5.3)	8.4 (8.4)	9.99e-01	o
forebrain	52.1	33.3	70.9	39.0 (48.6)	26.6 (24.5)	2.36e-01	o

In an additional set of 1000 rounds of the 10-fold cross-validation, the p-value for the group separation on the training set was determined in each run using an unpaired two-sided Student's t-test for each of the 63 symmetry features (AAI). A histogram of the 10 most significant structures in each run was calculated. Figure 3.12 shows the histogram for the GOS separation and thus the regions that are particularly correlated with the disease outcome. The plots show the 10 consistently

Table 3.10: Classification results in % (10-fold cross-validation, 1000 runs) obtained separating $GOS < 4$ vs. $GOS \geq 4$ based on AAI of follow-up MR images. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.01$).

Classification of $GOS < 4$ (Positives) vs. $GOS \geq 4$ (Negatives) (follow-up scans)							
structure	bACC	SENS	SPEC	mean AAI (SD) for $GOS < 4$ and $GOS \geq 4$		p-value	significance
all non-cortical	66.8	47.4	86.3	340.0 (219.8)	187.5 (101.1)	3.84e-04	++
all cortical	59.3	42.5	76.0	950.9 (357.7)	792.2 (166.8)	1.97e-02	+
all	62.6	42.1	83.1	1290.9 (522.9)	979.6 (227.2)	1.70e-03	++
accumbens area	64.8	36.9	92.8	54.4 (58.6)	25.2 (40.4)	2.66e-02	+
amygdala	57.2	40.1	74.3	20.5 (19.4)	12.3 (15.0)	7.54e-02	o
caudate	55.7	42.1	69.2	12.9 (13.9)	7.5 (6.9)	4.54e-02	+
cerebellum exterior	57.2	43.8	70.6	6.7 (4.4)	4.5 (4.9)	1.07e-01	o
cerebellum white matter	59.9	54.1	65.7	10.2 (7.4)	6.7 (4.6)	2.62e-02	+
cerebral white matter	51.7	23.8	79.7	7.7 (15.3)	4.9 (9.3)	3.75e-01	o
hippocampus	57.8	38.3	77.4	16.4 (16.1)	8.4 (6.9)	7.84e-03	++
inf lat ventricle	57.4	46.8	67.9	61.7 (48.1)	38.2 (25.1)	1.42e-02	+
lateral ventricle	64.4	47.4	81.4	33.5 (31.1)	14.9 (10.3)	8.10e-04	++
pallidum	57.4	42.1	72.7	22.9 (23.1)	11.6 (11.5)	1.20e-02	+
putamen	59.8	31.9	87.7	32.2 (51.2)	8.0 (10.2)	4.06e-03	++
thalamus	59.0	33.1	85.0	24.2 (24.1)	10.8 (16.4)	1.33e-02	+
ventral DC	36.7	32.5	40.9	8.0 (4.2)	7.9 (5.0)	9.25e-01	o
forebrain	47.8	42.4	53.1	28.7 (24.9)	26.5 (21.3)	7.25e-01	o

most relevant structures for GOS group separation using acute-phase (left) or follow-up (right) MR images. This experiment reveals that asymmetry in subcortical structures is particularly correlated with poor patient recovery. Notably, asymmetry in the thalamus, pallidum, hippocampus, putamen and occipital pole was found to discriminate TBI patients with favourable from non-favourable outcome. Both thalamus and hippocampus are known to be involved in TBI disease progression (Bigler, 2001) and were found to have predictive value in previous studies based on MR imaging (Strangman et al., 2010; Warner et al., 2010b,a; Irimia et al., 2012). The presented results also confirm the findings of Ramlackhansingh et al. (2011), where inflammation markers following a head trauma were significantly raised in the thalamus, putamen and occipital cortices.

3.4 Discussion and future work

In this work a framework called Multi-Atlas Label Propagation with Expectation-Maximisation based refinement (MALP-EM) was introduced for robust segmentation of MR brain images. Building on state-of-the-art registration and label fusion techniques, it was proposed to relax spatial priors obtained through multi-atlas label propagation and to combine segmentation results obtained with registration- and intensity-based approaches to exploit individual benefits. For prior relaxation incorrect label priors were detected at low intensity voxels or cisterns and the corresponding proba-

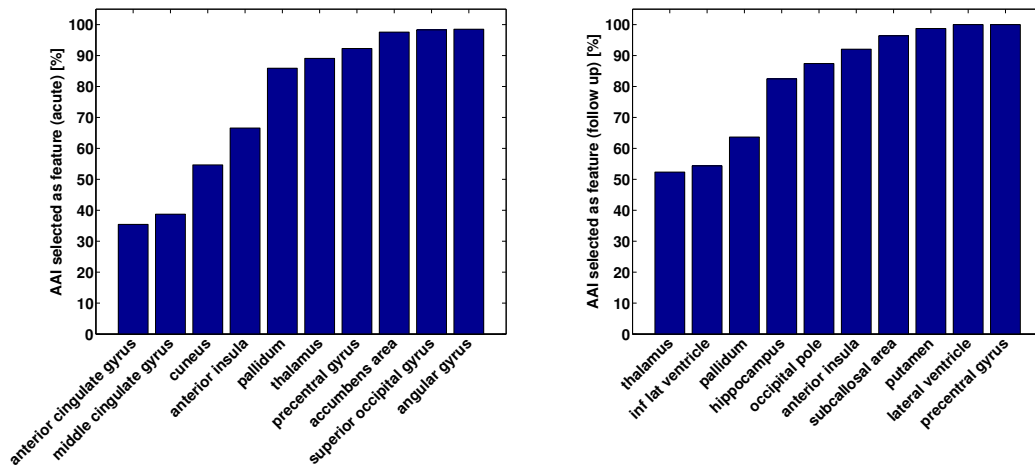


Figure 3.12: Relevance of individual brain structures for GOS group separation. Sorted histogram of how often a structure’s absolute asymmetry index (AAI) was one of the 10 most significant indices in the 10000 (1000 rounds of 10-fold cross-validation) runs. The 10 structures that were picked most often based on acute-phase (left) or follow-up (right) MR images.

bilities redistributed to the most probable CSF-like structure. Here it was assumed that low intensity voxels belong to cisterns that are filled with CSF. Potentially these low intensity voxels could also, especially in TBI patients, result from oedema, haemorrhage, or direct injury. The described relaxation approach is in theory applicable to the redistribution of prior probabilities between other structure pairs. However, here it was only described and implemented for the correction of mislabelled CSF-like structures. The intensity profile of CSF is particularly distinct from other GM or WM regions and thus redistribution less ambiguous.

The employed atlas is built from healthy patients and the described model does not allow for the detection of outliers or the classification of lesions, which is very challenging (Rao et al., 2014). Dependent on the disease, segmentation failures due to pathologies, such as contusions in TBI, could be addressed by an explicit lesion segmentation (Rao et al., 2014) or an outlier detection approach (Asman et al., 2013b). This is, however, left for future work. In general, imaging features derived from automatic segmentations, such as structural volumes, need to be interpreted carefully, when pathologies are present.

It was shown that MALP-EM significantly improves segmentation quality compared to non-intensity refined label fusion. Specifically, it was observed that significant improvements can be

obtained by combining results from joint label fusion and EM-refined fusion using a locally varying weighting factor Γ . This approach is similar to the weighting of different energy terms in the formulation presented by van der Lijn et al. (2008). The formulation described in this chapter allows any combination of segmentation results calculated with independent models or unrelated labelling techniques. In the future it will be interesting to investigate how more sophisticated combination strategies and intensity models can further improve this approach.

Previously, the objective evaluation procedure of the “MICCAI Multi-Atlas-Segmentation Challenge 2012” (Landman and Warfield, 2012) has shown MALP-EM to be among the leading segmentation methods. While the implementation of MALP-EM which was used in the challenge was preliminary and highly tuned, the implementation used for the work in this chapter is more generic and less dependent on parameter settings. Thanks to an improved registration and more sophisticated and general fusion strategy, significantly higher overlaps could be achieved for both MAPER (overall SI: 74.9% vs. 74.1%) and MALP-EM (overall SI: 76.4% vs. 75.8%) than in the Grand Challenge. Additional application of a learning-based segmentation bias correction method (Wang et al., 2011a) further improves the segmentation results (overall SI: 77.2%), yielding small but significant improvements over the best method (PICSL-BC; overall SI: 76.5%) in the Grand Challenge. It must be acknowledged that the development of MALP-EM benefitted from prior experience from participating in the Grand Challenge, where the timeframe for algorithm development and tuning was tight. However, the testing set from the Grand Challenge was not used to tune MALP-EM.

While the segmentation bias correction significantly improves label overlaps on the MICCAI Segmentation Challenge dataset, this post processing technique was not employed to segment the TBI subjects. The reasoning is that it is difficult to justify the application of a correction classifier that was trained exclusively on a homogeneous cohort of healthy subjects to a heterogeneous cohort of brain scans with severe pathology.

To evaluate MALP-EM on a TBI cohort, a specifically developed protocol was employed for the expert assessment of segmentation quality of the hippocampus, thalamus, putamen and occipital pole. The protocol is publicly available as supplementary material to Ledig et al. (2015). Using the protocol-based ratings of three independent experts, it was shown on 13 subjects of the TBI cohort that the proposed modifications based on image intensities improve on pure label fusion.

The proposed method is shown to be robust: 120 out of 125 TBI images were segmented successfully into 134 regions. After manual intervention on the preprocessing step, the successful record

increased to 122/125. This is a solid basis for future research into image-based quantification of MR brain images acquired from patients with head injury. Derived morphometric biomarkers, such as a structural asymmetry index, can serve as features for automatic classifiers, predicting how the image will be rated by an expert (MCS) and prognosticating clinical outcome (GOS). Given that MCS is assessed at the acute stage quantifying brain pathology it seems reasonable that structural asymmetry in acute MRIs correlates well with this score. In contrast to this, GOS is an outcome score assessed several months after the injury. It was expected that MRI features derived from follow-up scans are more consistent with the outcome measure than features available at the acute stage. The segmentation setup for the TBI subjects inherently differs from the intra-atlas experiments (15 training, 20 test images) in that more atlases were used for the segmentation of the TBI subjects, the image source (scanner) differs from the atlas database, and, most importantly, subjects with potentially substantial pathology are to be segmented. Visual inspection confirmed the robustness and advantages of the proposed method under these new challenges (*cf.* Figure 3.8). The generated segmentations of the TBI dataset are a valuable resource for investigating further potential biomarkers for TBI disease progression.

The work presented in this chapter also motivates further research and discussion about how meaningful or generalisable high Dice overlaps on a homogeneous cohort of healthy patients are. More informative similarity measures are desirable (*cf.* Chapter 7). In many studies, *e.g.* Alzheimer’s disease or TBI, the subjects of interest show high variability in both brain appearance and disease burden. More restrictive models might lead to a high labelling accuracy on subjects that are very similar to the atlas cohort. However, they are potentially too rigid to cope with images of subjects with significant pathology. Here more flexible formulations might be desirable, even if they are slightly less performant in intra-atlas cohort validations.

To conclude from the conducted experiments, MALP-EM yields a segmentation accuracy that is on healthy subjects comparable to other state-of-the-art methods while offering sufficient flexibility to cope with gross pathology. Most inaccuracies, as visually confirmed in the segmented TBI datasets, were due to minor problems in the brain extraction. This highlights the necessity of further improvement of fully automatic brain extraction tools, which is a very challenging task for brains in the presence of pathology.

3.5 Conclusion

In this chapter a fully automatic, robust and accurate segmentation framework called MALP-EM was presented. This includes a new paradigm: The spatially weighted combination of probabilistic segmentation results obtained through different techniques into a common segmentation exploiting individual benefits. Extensive quantitative evaluation on a manually annotated atlas cohort of healthy subjects confirmed that MALP-EM significantly improves on existing label fusion techniques. Based on the ratings of three independent experts, MALP-EM is superior to joint label fusion for hippocampus, thalamus, putamen and occipital pole segmentation on TBI brain scans. The benefits regarding robustness through intensity based refinement were demonstrated on 125 MR brain images of TBI subjects. Using MALP-EM 122 out of 125 available TBI brain images were segmented into 134 different anatomical regions. Correlations between asymmetry indices of paired structures and clinical variables were observed, using acute-phase or follow-up MR images. Furthermore, evidence was found that subcortical brain structures such as the thalamus, putamen and hippocampus have strong potential to predict the clinical outcome of individual TBI patients.

4 Longitudinal analysis of MR brain images

In the previous chapter a robust approach called MALP-EM for the segmentation of individual brain MR images was introduced. The aim in this chapter is to extend this approach for the simultaneous segmentation of time series of MR brain images. This is particularly valuable in longitudinal studies where automatically quantified structural atrophy rates are essential to monitor and examine a subject's progression over the course of a disease.

In the context of dementia several methods have been published that place particular focus on the measurement of hippocampal atrophy, a hallmark of AD. With this in mind the performance of two popular approaches, the Boundary Shift Integral (BSI) (Leung et al., 2010b) and four-dimensional (4D) GC (Wolz et al., 2010b), is calculated as reference on a subset of 309 subjects from the ADNI cohort. In the first part of this chapter, Section 4.1, potential extensions to the BSI are presented that exploit the availability of probabilistic hippocampus segmentations. However, this method is, like many other established techniques, limited to measure atrophy within a single structure only. In Section 4.2 this limitation is addressed. Specifically, a 4D extension of the framework introduced in Chapter 3 is proposed that is called Multi-Atlas Label Propagation with Expectation-Maximisation based refinement in 4D (MALP-EM4D). This extension allows the consistent and symmetric estimation of atrophy rates for a large number of structures of the whole brain simultaneously. The method is quantitatively assessed by measuring atrophy of the hippocampus, the ventricles and the whole brain while comparing to state-of-the-art BSI based approaches (Freeborough and Fox, 1997; Leung et al., 2010b). Further, the effects of differential bias correction (DBC) and different intrasubject registration approaches are investigated within the framework.

4.1 PBSI: A probabilistic extension of the Boundary Shift Integral

This section is based on:

C. Ledig, R. Wolz, P. Aljabar, J. Lötjönen and D. Rueckert, “PBSI: A symmetric probabilistic extension of the Boundary Shift Integral”, MICCAI Workshop on Novel Imaging Biomarkers for Alzheimer’s Disease and Related Disorders (NIBAD), pp. 125-132, 2012.

Abstract: *The BSI is a widely used technique to measure brain or hippocampal atrophy in dementia. In this section the BSI method is extended through the use of probabilistic posteriors obtained through a 4D EM segmentation framework. A complete framework is presented that refines spatial probabilistic priors for a baseline and a follow-up MR image simultaneously and incorporates this spatial information in a probabilistic Boundary Shift Integral (PBSI) measure. The method exploits the probabilistic information to calculate an accurate ROI on which the PBSI measure is evaluated. To ensure a consistent estimate, a symmetric strategy is followed by transforming baseline and follow-up images to a common, intermediate coordinate system. Furthermore, the aligned intensity images are corrected for any potential differential bias field. The resulting atrophy measure is used to compute group separation between cognitive normal subjects, patients with MCI and patients with AD. The approach is compared to state-of-the-art methods using 309 subjects from the ADNI cohort. The focus is on a prediction of the conversion from MCI to AD, which is technically challenging but clinically very useful. The presented atrophy rates are consistent with those in the literature and allow good classification accuracy for separating stable and progressive MCI patient groups.*

4.1.1 Introduction

Hippocampal volume loss has been shown to be a good marker for the presence of AD (Jack Jr. et al., 1997; Barnes et al., 2007; Leung et al., 2010a). Decline in hippocampal volume is also predictive at the stage of MCI (Leung et al., 2010a). Accurate atrophy measurements based on MR images could thus assist the diagnosis of AD before a conventional clinical diagnosis is established. Furthermore it could support the assessment of the potential of disease-modifying therapies (Leung et al., 2012). Manual labeling of the hippocampal region is time consuming (Leung et al., 2010a) and leads to intra- and interrater variability. This motivates the need for accurate and au-

tomatic methods. For example, in Chupin et al. (2009a) the scans of single time points were non-rigidly aligned with a probabilistic atlas which was then employed to segment the hippocampus. To quantify hippocampal atrophy automatically in longitudinal studies based on structural MR images several techniques have been proposed: These techniques are either based on the automatic segmentation of the hippocampal area, using for example GC (Wolz et al., 2010b), or on methods based on intensity differences of aligned images such as the BSI (Barnes et al., 2007; Leung et al., 2010a, 2012) (*cf.* Section 2.3.2.7).

In this work a model is presented that exploits advantages of both segmentation- and registration-based approaches, by incorporating subject-specific probabilistic spatial information into the original BSI formulation (Freeborough and Fox, 1997). The method relies on probability maps that are enhanced with a 4D EM algorithm (Van Leemput et al., 1999). This approach refines prior information based on the intensities of the images. Registration bias towards either of the two time points is avoided by following a symmetric registration strategy (Yushkevich et al., 2010; Leung et al., 2012) and the application of differential bias field correction (Lewis and Fox, 2004). The proposed method is evaluated on a subset of the ADNI-1 cohort by comparing classification accuracies, atrophy rates and samples sizes to two established methods based on GC (Wolz et al., 2010b) or the BSI (Leung et al., 2010a).

4.1.2 Method

4.1.2.1 Notation and prerequisites

In the following a pair of T1w MR images is considered, $\mathbf{I}^t$ and $\mathbf{I}^{t+1}$, of the same subject acquired at two different time points, t and $t + 1$. Let the n voxels in an MR image be indexed by $i = 1, \dots, n$. Denoting the intensity at a voxel as $y_i \in \mathbb{R}^+$ an image can be defined as $\mathbf{I}^t = \{y_1, y_2, \dots, y_n\}$. For a given image $\mathbf{I}$ the corresponding probabilistic segmentation is given as $\mathbf{S}_{\text{prob}} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$, where $\mathbf{z}_i$ is a vector of size K and the k^{th} component represents the probability that a voxel belongs to a structural class k . Since hippocampal atrophy is characteristic for AD, the focus in this section is on the segmentation of the hippocampus. The component of $\mathbf{S}_{\text{prob}}$ that is representing the hippocampus label is abbreviated, for a certain time point and thus for an image $\mathbf{I}^t$, as p^t . The rigid transformation from the coordinate system of $\mathbf{I}^t$ to that of $\mathbf{I}^{t+1}$ is denoted as $\mathbf{R}_{\mathbf{I}^t, \mathbf{I}^{t+1}}$. $\mathbf{R}_{\mathbf{I}^{t+1}, \mathbf{I}^t}$ is defined accordingly. Parameters are denoted as lower case Greek letters and sets with Ω .

It is assumed that individual brain masks are available for all baseline images $\mathbf{I}^t$ and that subject-

specific probabilistic priors p_{prior}^t for the hippocampal region in the baseline scan exist. These priors can be calculated with, for example, multi-atlas propagation techniques such as LEAP (Wolz et al., 2010a) or MALP-EM (*cf.* Chapter 3). Subsequent time points $\mathbf{I}^{t+1}$ can be brain extracted using a rigid transformation $\mathbf{R}_{\mathbf{I}, \mathbf{I}^{t+1}}$ of the baseline masks. While this is not symmetric it substantially reduces the computational cost and the need for quality control associated with the calculation of brain masks. Further it ensures that the outline of the brain images from distinct time points is consistent. All brain extracted images are intensity normalised (Nyul and Udupa, 1999; Smith, 2002), since the BSI is defined on intensity differences on normalised images (Freeborough and Fox, 1997).

4.1.2.2 Symmetrising the process

In order to process both time points simultaneously each pair of baseline and follow-up images is rigidly transformed to their common intermediate space. This is important to reduce bias introduced by asymmetric interpolation (Yushkevich et al., 2010) or by randomly defining one of the coordinate systems of $\mathbf{I}^t$ or $\mathbf{I}^{t+1}$ as reference coordinate system (Leung et al., 2012). The process is carried out using the rigid transformation $\mathbf{R}_{\mathbf{I}, \mathbf{I}^{t+1}}^{\text{avg}}$ obtained by averaging the transformations $\mathbf{R}_{\mathbf{I}, \mathbf{I}^{t+1}}$ and $\mathbf{R}_{\mathbf{I}^{t+1}, \mathbf{I}}^{-1}$ as described in Leung et al. (2012). Probabilistic spatial hippocampus priors, which are required for each baseline scan, are then also mapped to this common space. Note that alternatively an affine registration approach could be chosen. The effect of using affine registration instead will not be investigated here, but is further explored in Section 4.2.

4.1.2.3 Posteriors through 4D-EM optimisation

The challenge is to estimate the underlying but unknown probabilistic segmentations $\mathbf{S}_{\text{prob}}^t, \mathbf{S}_{\text{prob}}^{t+1}$ by means of the observed intensities y_i^t, y_i^{t+1} and the available subject-specific probabilistic priors p_{prior}^t and p_{prior}^{t+1} . A common assumption (Van Leemput et al., 1999; Ledig et al., 2015) is that the observed log-transformed intensities of voxels that belong to a certain class k are normally distributed with mean μ_k and standard deviation σ_k . The overall parameters for this intensity model are thus $\Phi = \{(\mu_1, \sigma_1), (\mu_2, \sigma_2), \dots, (\mu_K, \sigma_K)\}$. By assuming the same parameters Φ for both time points, both images are modelled and segmented simultaneously in a 4D EM framework. Smoothness within the segmentations is enforced with a global and stationary MRF (Van Leemput et al., 1999).

The model thus consists of one Gaussian intensity distribution (μ_1, σ_1) , that models the intensity distribution within the hippocampus at both time points, and $K - 1$ Gaussian distributions modelling the intensities in the background. Since different tissue types, including cerebrospinal fluid, white matter and non-hippocampal grey matter, are expected within the neighbourhood of the hippocampus, the parameter defining the number of tissues was set to $K = 4$. In the following, the background is thus modelled with three distinct intensity distributions. Prior spatial information is incorporated via a thresholded version of the available subject-specific probabilistic segmentation of the hippocampus, where the same baseline prior is used for both time points. This yields:

$$p_{\text{prior}}^t(i) = p_{\text{prior}}^{t+1}(i) = \begin{cases} \alpha, & \text{if } p_{\text{prior}}^t(i) > \alpha \\ p_{\text{prior}}^t(i), & \text{else} \end{cases} \quad (4.1)$$

Thresholding the priors at α slightly relaxes the priors, which is important to allow non-hippocampal voxels in the follow-up image to change their label. A mask on which the 4D EM refinement is carried out, can be obtained by dilating a region defined by the spatial priors p_{prior}^t and p_{prior}^{t+1} .

4.1.2.4 Probabilistic Boundary Shift Integral

The volume change Δv between a structure in the follow-up scan $\mathbf{I}^{t+1}$ and the corresponding structure in the baseline scan $\mathbf{I}^t$ is calculated based on a modified version of the BSI, which was proposed by Freeborough and Fox (1997). To incorporate the spatial information provided by the probabilistic segmentations $\mathbf{S}_{\text{prob}}^t, \mathbf{S}_{\text{prob}}^{t+1}$ into the measure, the PBSI is proposed. A novel spatially dependent weighting function $\gamma(i) \equiv \gamma(\mathbf{S}_{\text{prob}}^t, \mathbf{S}_{\text{prob}}^{t+1}, i)$ is introduced that can be defined to adapt the model to the application. The PBSI is calculated on a ROI Ω covering the boundary of the hippocampus with the surrounding tissue and is defined as:

$$\Delta v = \frac{v_{\text{voxel}}}{\xi_{\text{high}} - \xi_{\text{low}}} \sum_{i \in \Omega} \gamma(i) (\text{clip}(\mathbf{I}^{t+1}(i), \xi_{\text{low}}, \xi_{\text{high}}) - \text{clip}(\mathbf{I}^t(i), \xi_{\text{low}}, \xi_{\text{high}})) \quad (4.2)$$

Here the thresholds ξ_{low} and ξ_{high} are to be defined such that ξ_{low} represents intensities outside and ξ_{high} inside the hippocampus respectively. Further, v_{voxel} denotes the volume of a voxel in the

image. The clipping function is defined as:

$$\text{clip}(y_i, \xi_{\text{low}}, \xi_{\text{high}}) = \begin{cases} \xi_{\text{low}}, & \text{if } y_i < \xi_{\text{low}} \\ y_i, & \text{if } \xi_{\text{low}} \leq y_i \leq \xi_{\text{high}} \\ \xi_{\text{high}}, & \text{if } y_i > \xi_{\text{high}} \end{cases} \quad (4.3)$$

Since probabilistic segmentations ($\mathbf{S}_{\text{prob}}^t, \mathbf{S}_{\text{prob}}^{t+1}$) of the hippocampus in the baseline and follow-up scans are available, the ROI Ω can be obtained completely automatically. In order to calculate Ω , the definition given in Freeborough and Fox (1997) is generalised and Ω_{PBSI} is computed based on the probability maps. Here, the fuzzy union and intersection of two probabilistic maps

$$\cup(p_i, p_j) := \max(p_i, p_j) \quad \cap(p_i, p_j) := \min(p_i, p_j) \quad (4.4)$$

and the binarised versions depending on thresholds η and ζ are used:

$$\cap_{\eta}(p_i, p_j) := \begin{cases} 1, & \text{if } \eta \leq \cap(p_i, p_j) \\ 0, & \text{else} \end{cases} \quad \cup_{\zeta}(p_i, p_j) := \begin{cases} 1, & \text{if } \zeta \leq \cup(p_i, p_j) \\ 0, & \text{else} \end{cases} \quad (4.5)$$

The ROI Ω_{PBSI} is then defined as the difference, $\bar{\Omega} \setminus \check{\Omega}$, of a region $\bar{\Omega}$ describing the outer boundary of the ROI and an interior region lying completely within the hippocampus $\check{\Omega}$. The sets $\check{\Omega}$ and $\bar{\Omega}$ are defined as:

$$\check{\Omega} = \varepsilon^{n_e} \left[\cap_{\eta}(\mathbf{S}_{\text{prob}}^t, \mathbf{S}_{\text{prob}}^{t+1}) \right] \quad \bar{\Omega} = \rho^{n_d} \left[\cup_{\zeta}(\mathbf{S}_{\text{prob}}^t, \mathbf{S}_{\text{prob}}^{t+1}) \right] \quad (4.6)$$

The operators ε^{n_e} and ρ^{n_d} denote the n_e - and n_d -times application of the erosion and dilation operators. A common choice is $n_e^{\text{bin}} = n_d^{\text{bin}} = 1$ for both the erosion and dilation operator (Freeborough and Fox, 1997). In the model presented here the erosion can be controlled by varying η for $\check{\Omega}$. The model also allows to control the calculation of $\bar{\Omega}$ via the parameter ζ . The proposed procedure is illustrated in Figure 4.1. Figure 4.2 shows an example of a ROI, computed based on either hard segmentations (Ω_{BSI}) or the proposed method (Ω_{PBSI}).

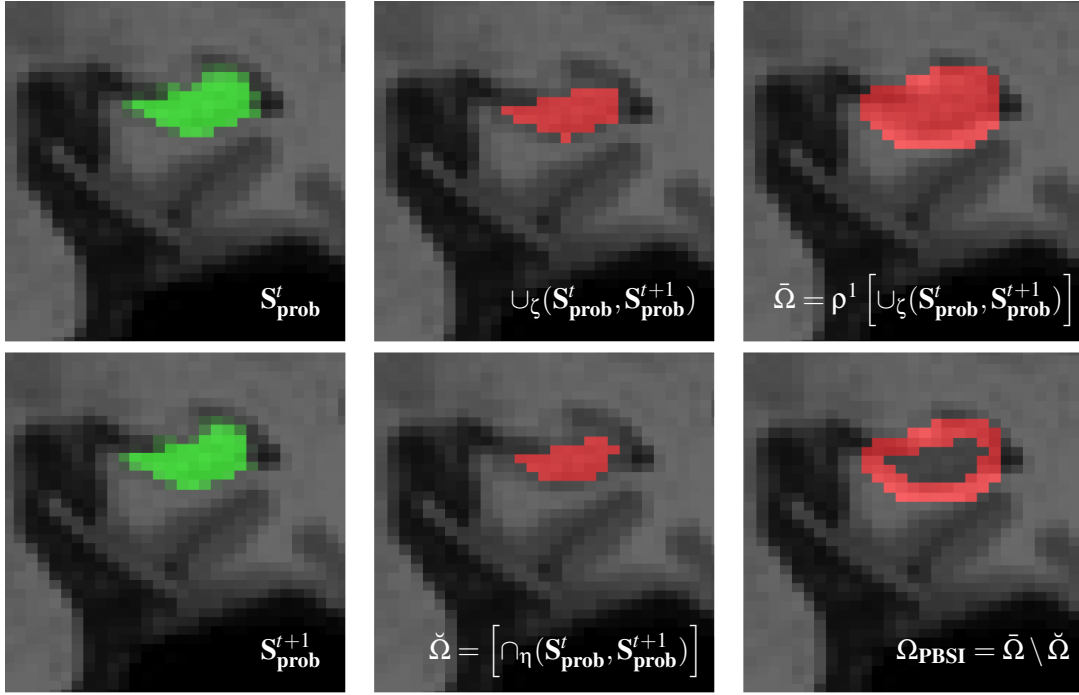


Figure 4.1: Construction of the ROI Ω_{PBSI} which is also shown in Figure 4.2 with parameters $\eta = 0.95$, $n_e = 0$ and $\zeta = 0.90$, $n_d = 1$. Probabilistic hippocampus segmentations (green) and binary masks (red) are overlaid over the respective subject MR image.

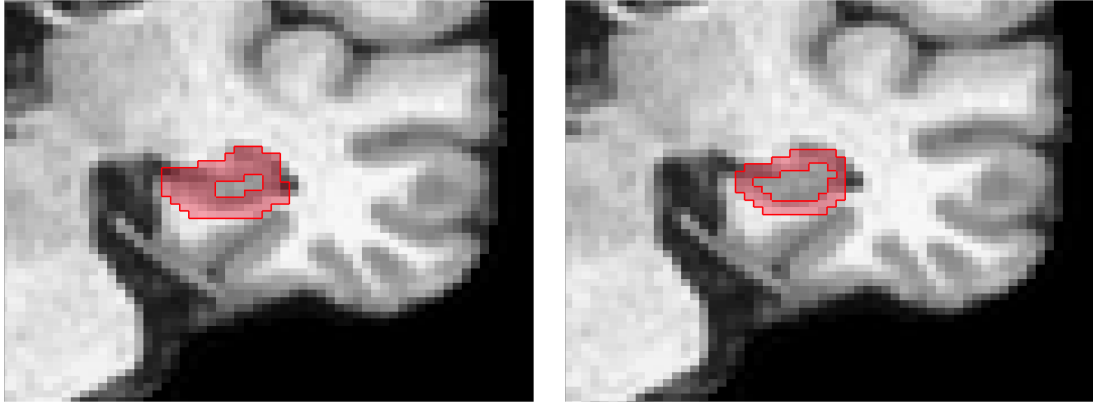


Figure 4.2: Binary ROIs (red) for the calculation of the respective integrals. ROI calculation based on hard labels (Ω_{BSI} , left) [$n_e^{\text{bin}} = 1$ and $n_d^{\text{bin}} = 1$] and based on the proposed method with probabilistic labels (Ω_{PBSI} , right) [$\eta = 0.95$, $n_e = 0$ and $\zeta = 0.90$, $n_d = 1$].

Abbreviation	Description
4D-GC	Results based on 4D graph-cuts from Wolz et al. (2010b)
DRC-HBSI ³	Hippocampus BSI Leung et al. (2010b,a)
4D-EM (no DBC)	volumes of 4D-EM segmentation <i>without</i> DBC
4D-EM	volumes of 4D-EM <i>with</i> DBC (Lewis and Fox, 2004).
$\text{BSI}_1^{\Omega_{\text{BSI}}}$	reimplementation of the BSI, ROI defined based on binary labels
$\text{PBSI}_1^{\Omega_{\text{PBSI}}}$	$\gamma \equiv 1$ and ROI defined based on probabilistic labels
$\text{PBSI}_\gamma^{\Omega_{\text{BSI}}}$	$\gamma = \max(0.5, p'_{\text{prob}}, p_{\text{prob}}^{t+1})$ and ROI defined based on binary labels
$\text{PBSI}_\gamma^{\Omega_{\text{PBSI}}}$	$\gamma = \max(0.5, p'_{\text{prob}}, p_{\text{prob}}^{t+1})$ and ROI defined based on probabilistic labels

Table 4.1: Compared methods.

4.1.3 Experiments and results

The method was evaluated on 309 subjects¹ from the ADNI-1 cohort (Mueller et al., 2005) based on segmentations of the *left* hippocampus. The cohort consisted of the 84 HC, 153 subjects with MCI and 72 patients with AD for which a 12 month follow-up scan and spatial priors calculated with the method described in Wolz et al. (2010a) were available. A further criterion for the definition of this subset was the availability of results from all reference methods summarised in Table 4.1 and those considered in Section 4.2. In addition it was ensured that all subjects were consistently diagnosed throughout the study. This means that subjects were excluded that reverted at some point to a less severe disease state, *e.g.* from MCI to HC or from AD to MCI.

The MCI group was further divided into 68 non-converting, sMCI and 60 converting, pMCI patients respectively. Progressive MCI patients were defined as those subjects converting within two years from diagnosed MCI (at baseline) to diagnosed AD. Stable MCI patients were defined as those being consistently diagnosed as MCI throughout the study, based on the available time points with imaging information. Patients with MCI that converted to AD after two years were thus excluded from the sMCI/pMCI analysis².

The main focus throughout the conducted experiments was on classification accuracy, and the calculation of sample sizes and realistic atrophy rates (*cf.* Section 4.2.3.2 for the definition of *atrophy* that is used in this thesis). In the following, five different approaches are compared against

¹A list of these subjects including diagnosis can be found at: http://www.doc.ic.ac.uk/~cl6311/PhD_thesis/ADNI1_309.csv, last accessed 29 June 2015.

²The definition of sMCI/pMCI is identical to the one which is used in Chapter 5, with one exception: Stable MCI patients who were not diagnosed at least 24 months after baseline were not excluded to avoid a further reduction of study subjects.

³provided by the Dementia Research Centre (DRC) through Dr. Kelvin K. Leung.

volume changes computed on the left hippocampus with the two state-of-the-art methods based on either graph-cuts (4D-GC) (Wolz et al., 2010b) or the hippocampus BSI (DRC-HBSI) (Leung et al., 2010a,b). The atrophy computation based on label volumes obtained by the 4D-EM segmentation is denoted as 4D-EM. This approach is expected to perform similar to 4D-GC which also performs intensity based refinement of spatial priors computed with LEAP (Wolz et al., 2010a). Next to these, atrophy rates are calculated based on a reimplementaion of the standard BSI, with a ROI based on hard labels (Ω_{BSI} , Figure 4.2), and different versions of the PBSI with ROIs being either Ω_{BSI} or Ω_{PBSI} . The weighting function γ was defined as either $\gamma \equiv 1$ ('PBSI₁'), which represents the standard BSI with the ROI determined according to the proposed method, or $\gamma = \max(0.5, p_{\text{prob}}^t, p_{\text{prob}}^{t+1})$ ('PBSI _{γ} '). The latter version decreases the weight of voxels that do not clearly represent hippocampal tissue. The compared methods are summarised in Table 4.1.

4.1.3.1 Preprocessing strategy and parameters

For brain extraction brain masks were employed that were obtained by a fully-automatic procedure called *pincram* as described in Heckemann et al. (2015). Then robustly intensity-scaled versions of the T1w MR images were calculated and the baseline $\mathbf{I}^t$, the 12 month follow-up scans $\mathbf{I}^{t+1}$ and the probabilistic priors obtained through LEAP (Wolz et al., 2010a) mapped to a common space as described in Section 4.1.2.2.

The spatial priors for the left hippocampus were thresholded with parameter $\alpha = 0.8$. The spatial priors for the three different background classes (WM, GM, CSF) were built in an ad-hoc manner based on the probability fraction left undefined by the hippocampal prior ($1 - p_{\text{prior}}(i)$) and the image intensities. As the hippocampus and all background classes are potentially adjacent, the MRF parameters (*cf.* Equation 3.10) were set to 0.15 for all classes. No MRF constraints were imposed *across* both time points. Given that the same spatial hippocampus prior was used for both time points this is necessary to ensure high sensitivity and to avoid penalising the detection of a decrease in label volume over time. For a maximum intensity value of 255, the lower and upper bounds for the clipping function were set to $\xi_{\text{low}} = 75$ and $\xi_{\text{high}} = 125$. As parameters for erosion and dilation $\eta = 0.95$, $n_e = 0$ and $\zeta = 0.90$, $n_d = 1$ were used respectively. Parameters were chosen without explicit training and were kept fixed for all experiments. The chosen parameters are summarised in Table 4.2.

Symbol	Value	Description
α	0.8	upper threshold for spatial hippocampus priors
K	4	hippocampus + three background tissue classes
MRF_{spatial}	0.15	spatial MRF connectivity strength
MRF_{temporal}	0	no temporal regularisation
n_e^{bin}	1	number of erosions of ROI (based on binary labels)
n_d^{bin}	1	number of dilations of ROI
η	0.95	threshold for interior ROI (based on probabilistic labels)
n_e	0	number of erosions of ROI
ζ	0.9	threshold for outer boundary of ROI
n_d	1	number of dilations of ROI
y^{max}	255	maximum image intensity after normalisation
ξ_{high}	125	upper bound of clipping function
ξ_{low}	75	lower bound of clipping function

Table 4.2: Summary of parameter setting.

4.1.3.2 Classification accuracy

For classification a 6-fold cross-validation was performed using LDA. SENS, SPEC and accuracy (ACC) averaged over 100 runs are shown in Table 4.3. On this dataset, the reimplementations of the standard BSI, $BSI^{\Omega_{\text{BSI}}}$, yielded the highest accuracy for all performed classification experiments. Both $BSI^{\Omega_{\text{BSI}}}$ and the PBSI based methods performed slightly better than the reference methods. When directly comparing the proposed BSI-based approaches to DRC-HBSI, it needs to be considered that the preprocessing procedures were different. Differential bias correction was beneficial for the 4D-EM approach and classification results obtained with 4D-EM similar to those calculated with 4D-GC.

4.1.3.3 Atrophy rates

Mean and standard deviation of the volume changes used for classification are shown in Table 4.4. Since volume loss depends on different factors, such as age, it is difficult to accurately predict atrophy rates (Chupin et al., 2009a). However, compared to other methods and rates based on manual segmentations, as for example those in Leung et al. (2010a), $PBSI_1$ and in particular the reimplementations of the BSI seem to consistently overestimate atrophy rates. This may be due to the fact that BSI potentially includes too many voxels at the boundaries. This is especially pronounced in the standard BSI where the ROI is defined using hard labels and no spatial information is consid-

Table 4.3: Classification results for separating AD groups obtained with 4D-GC, DRC-HBSI, 4D-EM, $BSI^{\Omega_{BSI}}$, $PBSI_1^{\Omega_{PBSI}}$, $PBSI_\gamma^{\Omega_{BSI}}$ and $PBSI_\gamma^{\Omega_{PBSI}}$. 4D-EM was also evaluated without DBC. Results shown for the **left hippocampus** in % calculated by averaging 100 runs of a 6-fold cross validation.

Method	Statistics	AD-CN	MCI-CN	AD-MCI	pMCI-sMCI
4D-GC	SENS/SPEC	71/76	59/63	61/64	64/73
	ACC	74	60	63	69
DRC-HBSI	SENS/SPEC	74/77	61/70	66/65	67/68
	ACC	76	64	65	68
4D-EM (no DBC)	SENS/SPEC	70/79	57/64	63/60	62/67
	ACC	75	60	61	65
4D-EM	SENS/SPEC	73/77	58/59	66/65	63/72
	ACC	75	58	66	68
$BSI^{\Omega_{BSI}}$	SENS/SPEC	76/86	62/72	62/69	72/74
	ACC	81	66	67	73
$PBSI_1^{\Omega_{PBSI}}$	SENS/SPEC	72/83	59/72	63/66	71/78
	ACC	78	64	65	75
$PBSI_\gamma^{\Omega_{BSI}}$	SENS/SPEC	69/85	60/74	54/68	63/76
	ACC	77	65	64	70
$PBSI_\gamma^{\Omega_{PBSI}}$	SENS/SPEC	71/86	59/75	58/69	67/79
	ACC	79	65	65	73

ered. Thus a decrease in intensity might contribute to the measured atrophy rate without considering if a voxel belongs to the hippocampus at all. This might also be the reason for the increased classification accuracy as non-hippocampal voxels can provide additional discriminative information. The presented framework relaxes this problem by incorporating spatial information and thus allows the calculation of more realistic rates when compared to the rates found in a meta-analysis of hippocampal atrophy rates in AD (Barnes et al., 2009a).

4.1.3.4 Sample sizes

In a hypothetical two-arm study, the required study population size to measure a specified change in atrophy rate for a given disease group (g) is of particular interest. For a power $(1 - \beta)$ and significance level α the *sample size* can be calculated (Fox et al., 2000; Leung et al., 2010a) as:

$$N = (z_{1-\beta} + z_{1-\alpha/2})^2 \times \frac{(2\sigma_g^2)}{\Delta^2} \quad (4.7)$$

Table 4.4: Mean volume change of the **left hippocampus** in % (top) and sample sizes for atrophy measurement on the MCI and AD group (bottom). Compared methods are 4D-GC, DRC-HBSI, 4D-EM, $\text{BSI}_1^{\Omega_{\text{BSI}}}$, $\text{PBSI}_1^{\Omega_{\text{PBSI}}}$, $\text{PBSI}_\gamma^{\Omega_{\text{BSI}}}$ and $\text{PBSI}_\gamma^{\Omega_{\text{PBSI}}}$. 4D-EM was also evaluated without the DBC. Standard deviation in parentheses. Corrected sample sizes were computed on the excess change over normal aging.

		4D-GC	DRC-HBSI	4D-EM (no DBC)	4D-EM	$\text{BSI}_1^{\Omega_{\text{BSI}}}$	$\text{PBSI}_1^{\Omega_{\text{PBSI}}}$	$\text{PBSI}_\gamma^{\Omega_{\text{BSI}}}$	$\text{PBSI}_\gamma^{\Omega_{\text{PBSI}}}$
Atrophy rates	Normal	-1.1 (1.8)	-1.5 (2.5)	-1.3 (2.0)	-1.5 (2.1)	-2.2 (2.7)	-1.6 (1.9)	-1.2 (1.6)	-1.4 (1.8)
	sMCI	-1.4 (2.3)	-2.7 (3.3)	-1.9 (2.5)	-2.2 (2.4)	-3.6 (3.8)	-2.7 (2.9)	-1.9 (2.1)	-2.2 (2.5)
	pMCI	-3.4 (2.5)	-5.5 (3.5)	-3.7 (2.9)	-4.2 (2.5)	-7.3 (3.7)	-5.7 (3.0)	-4.1 (2.4)	-5.0 (3.0)
	AD	-3.7 (2.4)	-6.0 (3.6)	-4.3 (2.7)	-4.9 (2.8)	-7.7 (4.4)	-6.0 (3.5)	-4.3 (2.9)	-5.3 (3.5)
Sample sizes	sMCI (uncor.)	676	370	439	284	280	297	288	305
	sMCI (CN-cor.)	17313	1829	4081	3141	1838	1988	2320	2229
	pMCI (uncor.)	134	103	154	92	63	69	84	90
	pMCI (CN-cor.)	302	193	367	233	128	136	171	174
	MCI (uncor.)	300	218	270	185	147	155	169	178
	MCI (CN-cor.)	1128	578	968	746	434	443	518	502
	AD (uncor.)	105	94	100	85	80	83	108	107
	AD (CN-cor.)	215	167	202	182	155	157	212	198

Here Δ is the change in atrophy that is to be shown between the treatment groups. In this thesis sample sizes were calculated to detect a 25% change in atrophy rate ($\Delta = 0.25\mu_g$) with 80% power ($z_{0.8} \approx 0.84$) at a 5% significance level ($z_{1-0.05/2} \approx 1.96$). These parameter choices are commonly found in the literature (Leung et al., 2010a; Wolz et al., 2010b; Yushkevich et al., 2010). Following recent communication in the neuroimaging community (Fox et al., 2011) discussing the importance of relating atrophy rates in dementia to normal atrophy during aging, sample sizes were corrected for normal aging. This is important as in the uncorrected case it is assumed that 100% treatment effect would effectively reduce the structural atrophy to zero (Leung et al., 2010a). To explicitly correct for this, Equation 4.7 was evaluated with $\Delta = 0.25(\mu_g - \mu_{\text{healthy}})$ to reduce the maximal treatment effect to the level of normal aging. It must be noted that in Equation 4.7 it is assumed that measurements of healthy atrophy have the same variance as measurements from diseased subjects ($\sigma_g \approx \sigma_{\text{healthy}}$) (Leung et al., 2010a), which, however, usually leads to a more conservative estimate.

Estimated sample sizes for the MCI (pMCI and sMCI) and AD patient groups with respect to normal aging are presented in Table 4.4. Performing DBC reduces sample sizes and yields more realistic atrophy rates for the 4D-EM approach. In terms of sample sizes, the proposed 4D-EM performs competitive with respect to the reference method based on GC, while calculating realistic atrophy rates. All instances of the (P)BSI perform comparable with $\text{PBSI}_\gamma^{\Omega_{\text{PBSI}}}$ producing the most realistic atrophy rates (Barnes et al., 2009a; Leung et al., 2010a).

4.1.4 Conclusion

A symmetric probabilistic extension of the boundary shift integral (PBSI) was presented and employed to analyse a population of 309 subjects with respect to their clinical diagnoses. In the presented experiments accuracies above 70% were obtained for the clinically highly relevant classification of MCI conversion. The calculated sample sizes are comparable to those obtained with state-of-the-art reference methods and relative volume changes are similar to published atrophy rates. The presented model offers a spatially dependent weighting function γ that allows to incorporate prior information into the BSI computation. While the accurate determination of the ROI seems of less importance, the incorporation of spatial information using γ might be the key to robust atrophy measurement. In future it will be interesting to investigate the influence of the parameters and to employ this approach for atrophy measurement on other brain structures.

4.2 Consistent and robust 4D whole-brain segmentation

This section is based on:

C. Ledig, W. Shi, A. Makropoulos, J. Koikkalainen, R. A. Heckemann, A. Hammers, J. Lötjönen and D. Rueckert, “Consistent and robust 4D whole-brain segmentation: application to traumatic brain injury”, IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), pp. 673-676, 2014.

Abstract: *A consistent approach to automatically segmenting longitudinal MR brain images in the presence of pathologies is proposed. Using symmetric intrasubject registration, corresponding scans are aligned. In an EM framework the availability of probabilistic segmentation estimates is exploited to calculate a consistent segmentation for images acquired at multiple time points. A novel technique is introduced to perform symmetric DBC for images in presence of pathologies. To achieve a consistent segmentation across multiple time points, a coupling term is proposed based on local image similarities using a spatially and temporally varying MRF. The consistency of the developed method was demonstrated by segmenting repeat scans into 134 regions. Furthermore, based on the dataset of 309 subjects from the ADNI cohort (cf. Section 4.1) atrophy rates and sample sizes are investigated for the hippocampus, ventricles and the whole brain. Here the proposed method is compared to established reference measurements available through the Ventricle Boundary Shift Integral (VBSI) and Brain Boundary Shift Integral (BBSI). The influence of intrasubject registration (rigid, affine) and DBC is investigated. Using a second clinical dataset, the approach has been further applied to segment baseline and follow-up scans from 56 patients who have sustained TBI. Significant correlations between regional atrophy rates and clinical outcome are reported: Patients with poor outcome showed a much higher thalamic atrophy rate ($-4.8 \pm 4.6\%$) than patients with favourable outcome ($-0.8 \pm 2.2\%$).*

4.2.1 Introduction

As described in Chapter 1 both AD and TBI are serious and common health problems affecting a large number of individuals. In dementia research the accurate measurement of structural atrophy has the potential to predict disease progression but also to monitor drug efficacy in clinical trials. In the context of TBI the quantification of anatomical atrophy is crucial to better understand the

structural changes during the weeks and months following the event of the actual trauma.

Several methods have been published that address the problem of quantifying brain changes over time based on MR images. One popular representative is “CLASSIC” (Xue et al., 2006), a method that uses adaptive clustering for tissue segmentation of MR scans taken at multiple time points. Another longitudinal segmentation method presented by Lorenzo-Valdes et al. (2004) is based on EM optimisation and has been applied to 4D cardiac sequences. This asymmetric approach uses stationary temporal MRFs and affine alignment with a probabilistic spatiotemporal atlas. Several other methods have been proposed that address primarily brain tissue segmentation such as Wang et al. (2011b) which uses level sets. In Bendlin et al. (2008) indications for brain volume loss following a TBI have been identified using tissue segmentation techniques on structural MR scans and DTI. In Wang et al. (2012a,b) the authors longitudinally segment multi-modal image sequences of patients with TBI into distinct tissue and lesion types using personalised atlas construction. The method relies on an appearance model and requires user input as initialisation. This approach was extended in Wang et al. (2013a) to extract biomarkers such as tissue/lesion volumes or cortical thickness measurements and applied to analyse five TBI patients. Wolz et al. (2010b) simultaneously segment the hippocampi of longitudinally acquired images using GC. Here, images of follow-up time points are affinely aligned to the baseline and its subject-specific prior. In Reuter et al. (2012) the authors build on FreeSurfer and propose an unbiased method to analyse longitudinal whole-brain image series. In this approach, an intrasubject template is built from information extracted cross-sectionally from individual time points. Individual time points are then re-processed using both information transferred from the within-subject template and time point specific data. A further discussion of relevant literature can be found in Chapter 2.

Recent research suggests that a fully symmetric process can significantly reduce bias, *e.g.* in atrophy measurement (Leung et al., 2012; Reuter et al., 2012). The advantages of a longitudinal segmentation of several time points over multiple, single time point segmentations are well established (Xue et al., 2006; Wolz et al., 2010b).

In this section a novel method is proposed, an extension of Ledig et al. (2012b, 2015), for consistent segmentation of serial images into many anatomical regions, rather than tissue classes. Image alignment and DBC is performed in a symmetric fashion. To achieve consistency, an approach is presented that can perform DBC in the presence of significant pathology in the images. Further an approach is introduced to determine a spatially and temporally varying, fully data driven temporal

coupling of the longitudinal segmentation based on MRFs. As far as known, the presented method is one the first consistent segmentation approaches that segment a series of intrasubject whole-brain MR images into a large number of anatomical structures while being robust to pathology. The conducted experiments show quantitative evidence for improved segmentation consistency while maintaining high accuracy. The method's performance is confirmed with respect to established atrophy measurements based on the BSI on the subset of 309 subjects from the ADNI cohort that was employed in Section 4.1. Further, it is demonstrated on a cohort of TBI patients that the proposed method is robust and has the potential to quantify imaging biomarkers, specifically atrophy, that correlate well with clinical outcome.

4.2.2 Method

In the following of this section the proposed methods are described for a symmetric, consistent multi time point whole brain segmentation. This 4D extension of the framework presented in Chapter 3 is called MALP-EM4D.

First in Section 4.2.2.1, subject specific probabilistic priors are calculated for scans at individual time points. These priors and the intensity images are transformed to a common intermediate space using a symmetric registration technique (Leung et al., 2012). These priors are then employed to segment longitudinal brain images using a 4D EM optimisation in Section 4.2.2.2. To enable a flexible spatially and temporally varying longitudinal coupling and thus consistent segmentation a temporal coupling weight is calculated based on intensity normalised images in Section 4.2.2.3. An innovative way is described to apply DBC to images that may contain pathologies in Section 4.2.2.4.

4.2.2.1 Spatial priors and symmetric longitudinal image alignment

Assuming preprocessed, brain extracted (pinfram, Heckemann et al. (2015)) and bias corrected (N4, Tustison et al. (2010)) images subject specific probabilistic labels are derived from M available atlases using the multi-atlas label propagation approach described in Chapter 3. For the unsegmented images at T time points, $\mathbf{I}^t$ at $t = 0 \dots T - 1$, individual transformations ϕ_m^t , $m = 1 \dots M$ are calculated by registering M manually labelled atlases to $\mathbf{I}^t$. For the image alignment, MAPER (Heckemann et al., 2010) is employed, an approach that incorporates tissue probability maps into the registration and relies on a non-rigid registration based on FFDs (Rueckert et al., 1999; Modat

et al., 2010). Using the resulting transformations the M label maps are mapped to the space of image $\mathbf{I}^t$. For each time point t , a probabilistic atlas is created using MALP-EM with a locally weighted multi-atlas fusion strategy based on GSSD as described in Chapter 3. In the following, π_{ik}^t denotes the probabilistic prior for label k at voxel i and at time point t .

As shown in Leung et al. (2012); Yushkevich et al. (2010), asymmetric image registration can introduce significant bias into longitudinal image analyses. Thus, a symmetric setup is used based on global rigid or affine (12 DOFs) registration similar to Yushkevich et al. (2010); Leung et al. (2012). All images and priors of all T time points are transformed to a common intermediate space (Leung et al., 2012).

4.2.2.2 4D-EM segmentation

The probabilistic segmentation of the aligned multi time point scans, residing in their common rigid/affine intermediate space, is then estimated simultaneously. Based on the subject and time point specific priors, a 4D extension of the EM algorithm (Ledig et al., 2012b, 2015) is employed.

The same Gaussian distributions, described by Φ , are employed for all time points. Spatial smoothness of the final segmentations is enforced through a global and stationary MRF (Cardoso et al., 2011). To also enforce longitudinal consistency, novel spatially and temporally varying constraints are proposed. The strength of the temporal coupling is formulated as a function of the local intensity similarity of the aligned images. This is crucial both to allow for flexibility in regions with potential pathology and to enforce consistency in regions with no or very little change. A GMM (*cf.* Chapter 3) is assumed to model the probability of observing intensity y_i^t at a voxel i at time point t

$$f(y_i^t | \Phi^{t,(q)}) = \sum_k f(y_i^t | z_i^t = k, \Phi^{(q)}) \frac{\pi_{ik}^t e^{-U_{\text{MRF}}(k | p_{\mathcal{N}_i}^{(q)})}}{\sum_{j=1}^K \pi_{ij}^t e^{-U_{\text{MRF}}(j | p_{\mathcal{N}_i}^{(q)})}} \quad (4.8)$$

where $\mathcal{N}_i$ is the set of first-order neighbours of voxel i , k a label index and z_i^t the labelling of voxel i at time point t .

The MRF energy U_{MRF} that is dependent on the probabilistic label estimates $\mathbf{p}$ at iteration q is represented as:

$$U_{\text{MRF}}(k | p_{\mathcal{N}_i}^{(q)}) = \sum_{j=1}^K G_{kj} v_j, \quad v_k = \sum_{l \in \mathcal{N}_i^{x,y,z}} p_{lk}^{t,(q)} + \sum_{l \in \mathcal{N}_i^t} \gamma_i^{t,l} p_{lk}^{l,(q)} \quad (4.9)$$

Here $\mathbf{G}$ denotes a $K \times K$ matrix defining the connectivity between class k and j . The vector $\mathbf{v}$ of size K denotes the accumulated class probabilities in the 8-neighborhood $\mathcal{N}_i$ of voxel i . The class probabilities of adjacent time points are weighted by temporal weighting factors, $\gamma_i^{t,\pm 1}$. The following sections describe the calculation of $\gamma_i^{t,\pm 1}$, and the DBC. These steps are performed once before the first iteration of the EM algorithm.

4.2.2.3 Voxel-wise coupling weights for temporal consistency

To reduce segmentation bias, it is proposed to couple the segmentation of corresponding images from multiple time points through a spatially and temporally varying MRF. While the spatial MRF weights are stationary, the temporal weights are determined based on the similarity of temporally adjacent intensity normalised images (Nyul and Udupa, 1999; Smith, 2002). At time point t local patch similarities are calculated between time point t and $t + 1$, or t and $t - 1$ respectively. Specifically, the Gaussian weighted sum of absolute differences (GSAD) is calculated where the kernel width determines the “patch size”. From the GSAD value at a given voxel $w_i^{t,t\pm 1}$, the coupling weights $\gamma_i^{t,\pm 1}$ to obtain temporal consistency are derived based on a logistic function of the form:

$$\gamma_i^{t,\pm 1} = c - \frac{c}{1.0 + a \cdot \exp(-b \cdot w_i^{t,t\pm 1})} \quad (4.10)$$

Here c determines the maximal value of γ_i while a and b determine the shape of the S -curve. $\gamma_i^{t,\pm 1}$ is close to c in areas where the data suggests no presence of pathologies or significant structural changes (small w_i) and close to 0 otherwise. Example plots of γ are visualised in Figure 4.3.

4.2.2.4 Differential Bias Correction

Lewis and Fox (2004) proposed a straightforward way to estimate the differential bias fields of aligned longitudinal image pairs by applying a median filter to the difference image, calculated between log-transformed images. This approach is extended by Leung et al. (2012) to multiple time points. However, the basic assumption of this approach, that the differential bias field is given as the low frequency component of the difference image, does not necessarily hold in the presence of pathologies. A violation of this assumption is given by the image pair in Figure 4.5. Here, significantly enlarged ventricles or pathological changes appear as low frequency intensity change and thus falsely influence the bias field estimation with the approach presented in Lewis and Fox

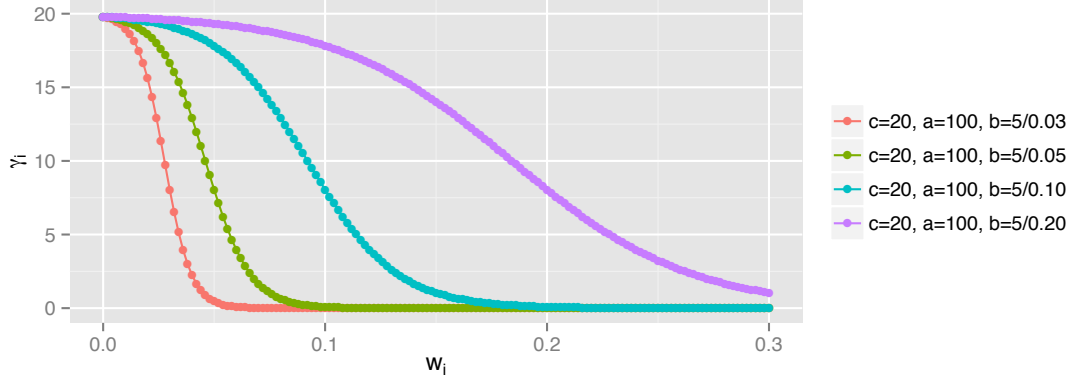


Figure 4.3: Example coupling weights γ_i dependent on local patch similarity w_i for different choices of parameter b . In this work, b is automatically determined based on the image intensities as $b = 5.0/\text{median}[w_i^{t,t+1}]$. The figure shows typical results for intrasubject repeat scans (red), follow-up images of healthy subjects, or of patients with AD or mild TBI (green, blue). For highly pathological cases such as the one illustrated in Figure 4.5 rather high values for b are determined (purple). Note: The low values of w_i are due to the fact that the weights are calculated on the log-transformed image intensities.

(2004); Leung et al. (2012).

An alternative way is proposed to adapt bias field estimation to increase its robustness towards pathologies. The formulation presented in Leung et al. (2012) is extended to estimate pairwise differential bias fields $\mathbf{dbf}_{\text{poly}}^{t,t'}$, for time points t, t' , based on a polynomial model. After estimating a differential bias field $\mathbf{dbf}^{t,t'}$ based on the difference image, a second order polynomial is fit to $\mathbf{dbf}^{t,t'}$ using weighted least squares optimisation to calculate $\mathbf{dbf}_{\text{poly}}^{t,t'}$. As weights the quantities γ are exploited which were previously derived from patch similarities. Employing the temporal coupling weights γ , the influence of pathological areas to the bias field estimation can be reduced. This allows the estimation of a differential bias field based on low frequency intensity changes within brain regions that are not subject to substantial pathological change. A symmetric bias field correction is then performed as described in Leung et al. (2012) using all pairwise estimates $\mathbf{dbf}_{\text{poly}}^{t,t'}$. DBC can be considered as a preprocessing step for the described 4D-EM segmentation (*cf.* Section 4.2.2.2) and is performed once after the voxel-wise coupling weights have been calculated (*cf.* Section 4.2.2.3).

4.2.3 Experiments and results

4.2.3.1 Materials and parameters

As brain atlases the dataset was used which was provided as part of the “MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling” (Landman and Warfield, 2012). It consists of 35 T_1 -weighted MR brain images with isotropic voxel sizes of $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$ from the OASIS database (Marcus et al., 2007). The corresponding 134 anatomical labels have been created manually by experts⁴. The scans have been acquired from 30 subjects (20 female, 10 male, mean $\pm$ SD age: 34.3 ± 20.7 years) where 5 subjects have been scanned twice. The duplicates were acquired within 90 days of the original scan (Marcus et al., 2007). A more detailed description of the atlas dataset can be found in Section 2.2.5.

In addition the proposed method was applied to the same 309 subjects from the ADNI-1 cohort that were used in Section 4.1. Atrophy rates and sample sizes were calculated for the hippocampus, ventricles and the whole brain. Results are compared to DRC-HBSI (*cf.* Section 4.1) measurements, KN-BBSI (Leung et al., 2010b) and VBSI measurements downloaded from ADNI named ‘Fox Lab - BSI Measures, version November 3rd 2014’. These BSI measures were calculated using affine registration and DBC.

From a cohort of patients who had sustained TBI, 56 intrasubject pairs of MR brain images were segmented which were acquired several months apart (mean 6.7 ± 1.2 months). Acquisition parameters were: resolution $240 \times 256 \times 170$; spacing $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$; sagittal slices. More details on this cohort can be found in Chapter 6.

The proposed method was evaluated in three ways: First, consistency and accuracy were evaluated on the 5 pairs of intrasubject repeat scans. Second automatically calculated atrophy rates and sample sizes were compared to those obtained using state-of-the-art BSI based measurements. The effect of DBC and affine vs. rigid intrasubject image alignment was investigated. Finally, the method was applied to the TBI cohort to evaluate the clinical relevance of the calculated atrophy measurements.

To segment the five pairs of repeat scans, which are part of the atlas cohort, the subset of 15 atlases that served as training data in the Grand Challenge was used. These $M = 15$ scans were

⁴provided by Neuromorphometrics, Inc. under academic subscription (<http://Neuromorphometrics.com/>).

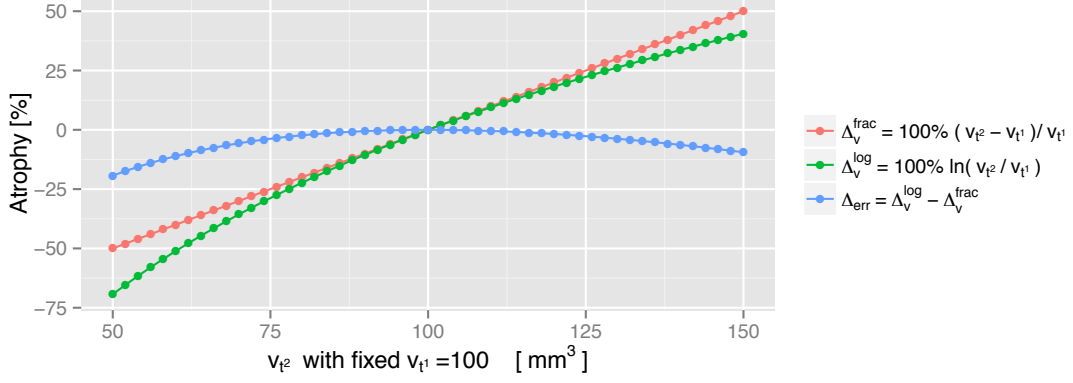


Figure 4.4: Simulated atrophy calculation for fixed baseline volume (v_{t^1}) and varying follow-up volume (v_{t^2}). Calculation based on $\Delta_v^{\text{frac}}(t^1, t^2) = 100\%(v_{t^2} - v_{t^1})/v_{t^1}$ (red) and the logarithmic transform $\Delta_v^{\text{log}}(t^1, t^2) = 100\% \ln(v_{t^2}/v_{t^1})$ (green) with corresponding approximation error (blue).

distinct from the 5×2 repeat datasets. For the segmentation of the TBI data all $M = 30$ atlases were used, excluding the five repeat scans. More details on the atlas subsets is provided in Section 2.2.5.

Parameters: Parameters were set mainly based on results from previous experiments. For the GSSD in the locally weighted atlas fusion σ_{GSSD} was set to 2.5. As spatial MRF weights, entries (k, k) of the connectivity matrix $\mathbf{G}$ were set to 0. Entries (k, j) were set to 1.0 if structures k and j shared a boundary, and to 2.5 otherwise. For the logistic function in Section 4.2.2.3, parameters were chosen heuristically as $c = 20.0$, $a = 100$ and $b = 5.0/\text{median}[w_i^{t,t+1}]$. b was modelled dependent on the median of the occurring GSAD values w_i ($\sigma_{\text{GSAD}} = 1.0$) within the images. This adapts the logistic function to different similarity levels of image pairs in a data-driven fashion. As in Lewis and Fox (2004), a median filter of kernel size 5 was used to estimate the differential bias field.

4.2.3.2 Calculation of atrophy rates

In Section 4.1 atrophy rates were calculated based on the BSI or as $\Delta_v^{\text{frac}}(t^1, t^2) = 100\%(v_{t^2} - v_{t^1})/v_{t^1}$. For the remainder of this thesis, atrophy rates are calculated using the logarithmic transform as $\Delta_v(t^1, t^2) = \Delta_v^{\text{log}}(t^1, t^2) = 100\% \ln(v_{t^2}/v_{t^1})$. This is a very close numerical approximation (*cf.* Figure 4.4, and Table 4.4 and Table 4.7) and has the advantage of being multiplicatively symmetric and transitive in the sense that $\Delta_v(t^1, t^3) = \Delta_v(t^1, t^2) + \Delta_v(t^2, t^3)$. Note that atrophy rate and volume

ICC	manual	MAPER	no coupling	proposed (DBC)	(weighted DBC)
36 subcortical	.80±.28	.95±0.08	.95±.15	.98±.08	.97±.09
98 cortical	.78±.25	.95±0.09	.94±.07	.99±.02	.99±.02
134 structures	.78±.25	.95±0.09	.94±.10	.99±.04	.98±.05

Table 4.5: Intraclass correlation coefficient (ICC) (mean±SD) based on label volumes averaged over 5 image pairs for all 36 subcortical and all 98 cortical regions obtained manually, with MAPER, and the proposed method without and with the temporal constraints as described in Section 4.2.2.3-4.2.2.4. Proposed method evaluated with unweighted, standard DBC and weighted DBC. Bold = significantly better than without coupling (paired, two-sided Student’s t-test, $p < 0.05$).

change is used interchangeable, which means that a *positive* atrophy rate indicates an *increase* in volume.

4.2.3.3 Test-retest reliability

To investigate the consistency of imaging biomarkers derived using the proposed segmentation method it is crucial to evaluate its test-retest reliability. Since five testing subjects in the testing set have duplicate scans this could be quantified. The ICC was employed to quantify test-retest reliability, following Shrout and Fleiss (1979) in the scheme ICC(2,1). ICCs for both manual and automated segmentations are summarised in Table 4.5. Overall, the proposed method yielded a significantly higher ICC than the reference methods.

4.2.3.4 Subject-atlas label overlaps

Enforcing identical segmentations at both time points leads to high ICC values, but high ICCs do not necessarily reflect a high segmentation accuracy which needs to be assessed with measures of overlap. The accuracy of segmenting both time points with and without the temporal constraints was assessed quantitatively. Table 4.6 shows overlap results based on the Dice coefficient (Dice, 1945) (DC) or also called SI. The DC was calculated on the 10 repeat scans by comparing to the gold standard. Enforcing temporal consistency did not deteriorate segmentation accuracy, but rather improved it for some structures. In this work, the DC was calculated in an intermediate coordinate system to which the reference segmentations were transformed using nearest-neighbour interpolation. The presented scores should thus not be directly compared to the Grand Challenge (Landman and Warfield, 2012) or the results presented in Chapter 3.

	36 subcort.	98 cort.	134 structures
proposed (weighted DBC)	82.1	72.6	75.1
proposed (DBC)	82.1	72.6	75.1
no coupling	80.5	72.5	74.7

Table 4.6: Average Dice coefficient (Dice, 1945) (DC) values comparing the proposed approach without or with temporal constraints to the gold-standard segmentations of the 10 images of the 5 pairs of repeat scans. Proposed method evaluated with unweighted, standard DBC and weighted DBC. Bold significantly better than without coupling (paired, two-sided Student’s t-test, $p < 0.001$).

4.2.3.5 Longitudinal analysis of 309 subjects from ADNI

The proposed approach MALP-EM4D was also compared to established measurements calculated with the BSI using the same dataset that was used in Section 4.1 of 309 subjects from the ADNI-1 cohort. Atrophy rates for distinct patient groups and sample sizes were calculated as described in Section 4.1.3.4. Measurements were compared on the left hippocampus (Table 4.7), the ventricles (Table 4.8) and the whole brain (Table 4.9). The proposed method yields realistic atrophy rates for all investigated structures. While sample sizes for the ventricles are slightly lower using MALP-EM4D, the BSI based measurements yielded a better group separation on the left hippocampus and the whole brain. The conducted experiments suggest that an affine registration approach might be preferable to rigid registration to align intrasubject MR images. In Paling et al. (2004) the authors discussed variations in scanner voxel size as possible explanation for reduced measurement bias when using affine registration. However, no significant benefits when using affine registration were found (Paling et al., 2004) so that rigid registration should be considered a reasonable option (Yushkevich et al., 2010). In contrast to the results presented in Section 4.1, where no temporal coupling was used in the EM optimisation, no clear benefit was observed using DBC. The performed experiments indicate that the regularisation imposed by the 4D MRF largely compensates for existing differential bias. Nevertheless, performing DBC slightly decreased measured atrophy rates, suggesting that small changes in intensity signal are removed. Employing a fixed temporal coupling weight ($\gamma = 1$) does not necessarily deteriorate group separation, however it tends to yield unrealistically low atrophy rates due to a loss in sensitivity.

Further MALP-EM4D was employed with affine registration and unweighted differential bias field correction (as in Leung et al. (2012)) to analyse left/right counterparts of the hippocampus and ventricles individually. The results are summarised in Table 4.10. While similar sample sizes

Table 4.7: Mean atrophy rates of the **left hippocampus** in % (top) and sample sizes for atrophy measurement on the MCI and AD group (bottom). Compared methods are the DRC-HBSI (*cf.* Section 4.1), MALP-EM and different instances of the proposed method. Standard deviation in parentheses. Corrected sample sizes were computed on the excess change over normal aging.

Method		DRC-HBSI	MALP-EM	MALP-EM4D fixed affine unweighted	MALP-EM4D varying rigid none	MALP-EM4D varying affine none	MALP-EM4D varying affine unweighted	MALP-EM4D varying affine weighted
Atrophy rates	Normal	-1.5 (2.5)	-1.9 (3.9)	-0.6 (2.1)	-1.2 (1.8)	-1.2 (1.7)	-1.2 (1.6)	-1.2 (1.7)
	sMCI	-2.6 (3.2)	-2.3 (5.7)	-1.4 (2.7)	-2.0 (2.7)	-2.0 (2.7)	-2.0 (2.6)	-2.0 (2.6)
	pMCI	-5.3 (3.3)	-3.1 (4.3)	-2.0 (1.7)	-4.3 (2.8)	-4.3 (2.9)	-4.1 (2.8)	-4.3 (2.9)
	AD	-5.7 (3.4)	-3.2 (4.3)	-2.4 (2.2)	-4.8 (3.8)	-5.0 (3.8)	-4.7 (3.8)	-4.9 (3.8)
Sample sizes	sMCI (uncor.)	364	1496	864	462	437	429	447
	sMCI (CN-cor.)	1816	40622	2568	3399	2695	2850	2743
	pMCI (uncor.)	99	476	167	110	114	117	114
	pMCI (CN-cor.)	187	3004	336	218	219	235	217
	MCI (uncor.)	214	834	453	268	272	269	275
	MCI (CN-cor.)	572	7803	1152	829	793	837	797
	AD (uncor.)	90	447	202	157	148	161	155
	AD (CN-cor.)	162	2561	359	284	258	292	269

Table 4.8: Mean atrophy rates of the **ventricles** in % (top) and sample sizes for atrophy measurement on the MCI and AD group (bottom). Compared methods are the VBSI, MALP-EM and different instances of the proposed method.

Method		VBSI	MALP-EM	MALP-EM4D fixed affine unweighted	MALP-EM4D varying rigid none	MALP-EM4D varying affine none	MALP-EM4D varying affine unweighted	MALP-EM4D varying affine weighted
Atrophy rates	Normal	4.2 (4.1)	3.2 (3.8)	2.2 (2.4)	3.4 (3.5)	3.3 (3.4)	2.9 (3.0)	3.2 (3.3)
	sMCI	5.0 (4.9)	4.1 (4.6)	3.0 (2.8)	4.4 (4.0)	4.3 (3.9)	3.7 (3.4)	4.1 (3.8)
	pMCI	8.8 (4.5)	7.6 (4.3)	5.4 (3.2)	7.7 (4.1)	7.5 (4.0)	6.8 (3.7)	7.3 (4.0)
	AD	9.6 (5.8)	8.3 (5.3)	6.1 (3.9)	8.6 (5.4)	8.3 (5.2)	7.5 (4.6)	8.2 (5.1)
Sample sizes	sMCI (uncor.)	241	308	219	215	206	216	214
	sMCI (CN-cor.)	9550	6472	3595	4262	3988	4208	4140
	pMCI (uncor.)	64	80	85	72	71	77	74
	pMCI (CN-cor.)	236	243	248	228	224	231	232
	MCI (uncor.)	129	155	140	126	124	132	128
	MCI (CN-cor.)	836	751	666	668	652	661	669
	AD (uncor.)	92	101	104	100	98	95	97
	AD (CN-cor.)	297	272	263	272	267	250	260

were observed for corresponding structures in both brain hemispheres, atrophy seems to be more pronounced in the right brain hemisphere on this dataset.

The potential of combining atrophy rates from different structures is summarised in Table 4.11. Here atrophy rates of distinct structures were combined by calculating the arithmetic mean of the individual rates. Atrophy rates for the cortical and deep GM were calculated based on the actual volumes of the tissues. These results are encouraging as they show the potential of combining biomarkers of multiple structures instead of focusing on individual structures only. Sample sizes that were calculated based on the inferior lateral ventricles and all ventricular structures are lower than those obtained with the considered reference methods.

Table 4.9: Mean atrophy rates of the **whole brain** in % (top) and sample sizes for atrophy measurement on the MCI and AD group (bottom). Compared methods are the KN-BBSI (Leung et al., 2010b), MALP-EM and different instances of the proposed method.

Method		KN-BBSI	MALP-EM	MALP-EM4D	MALP-EM4D	MALP-EM4D	MALP-EM4D	MALP-EM4D
Coupling				fixed	varying	varying	varying	varying
Registration				affine	rigid	affine	affine	affine
Diff. bias correction				unweighted	none	none	unweighted	weighted
Atrophy rates	Normal	-0.7 (0.6)	-0.4 (0.7)	-0.5 (0.4)	-0.7 (0.8)	-0.7 (0.8)	-0.6 (0.6)	-0.7 (0.7)
	sMCI	-0.9 (0.7)	-0.6 (0.7)	-0.7 (0.5)	-0.9 (0.8)	-1.0 (0.9)	-0.8 (0.8)	-0.9 (0.8)
	pMCI	-1.6 (0.8)	-0.8 (1.0)	-0.9 (0.6)	-1.2 (0.9)	-1.3 (1.0)	-1.1 (0.9)	-1.2 (0.9)
	AD	-1.7 (1.0)	-0.9 (0.9)	-1.1 (0.6)	-1.5 (1.1)	-1.7 (1.1)	-1.4 (0.9)	-1.6 (1.0)
Sample sizes	sMCI (uncor.)	162	426	149	205	214	220	221
	sMCI (CN-cor.)	3491	3272	3075	3256	3736	3779	3852
	pMCI (uncor.)	63	391	107	154	147	143	144
	pMCI (CN-cor.)	209	1231	598	799	746	668	683
	MCI (uncor.)	112	432	132	203	195	190	194
	MCI (CN-cor.)	634	1985	1271	1772	1668	1502	1550
	AD (uncor.)	82	250	76	128	103	96	100
	AD (CN-cor.)	239	648	278	402	319	291	300

Table 4.10: Mean atrophy rates of **other structures** in % (top) and sample sizes for atrophy measurement on the MCI and AD group (bottom) using MALP-EM4D with affine registration, unweighted DBC and the proposed spatially varying temporal coupling.

Structure		hippocampus R	hippocampus L	inf lat. vent. R	inf. lat. vent. L	lat. vent. R	lat. vent. L
Atrophy rates	Normal	-0.9 (1.9)	-1.2 (1.6)	1.5 (3.3)	2.1 (4.3)	3.1 (3.2)	3.2 (3.3)
	sMCI	-2.0 (3.1)	-2.0 (2.6)	3.0 (4.6)	3.2 (4.2)	4.0 (3.8)	3.9 (3.9)
	pMCI	-3.7 (3.2)	-4.1 (2.8)	6.2 (5.2)	6.4 (5.5)	7.2 (4.3)	7.1 (4.1)
	AD	-5.6 (5.3)	-4.7 (3.8)	8.6 (6.8)	7.5 (5.2)	7.9 (5.2)	7.7 (4.7)
Sample sizes	sMCI (uncor.)	599	429	611	442	230	261
	sMCI (CN-cor.)	1958	2850	2498	3696	4375	8410
	pMCI (uncor.)	186	117	182	182	90	82
	pMCI (CN-cor.)	325	235	319	393	275	266
	MCI (uncor.)	336	269	315	306	144	148
	MCI (CN-cor.)	733	837	717	1018	720	851
	AD (uncor.)	229	161	159	122	107	92
	AD (CN-cor.)	326	292	234	232	291	268

Table 4.11: Mean atrophy rates of **combined structures** in % (top) and sample sizes for atrophy measurement on the MCI and AD group (bottom) using MALP-EM4D with affine registration, unweighted DBC and the proposed spatially varying temporal coupling. Structures were combined by averaging individual atrophy rates (†) or calculating atrophy rates on merged structures (‡). *: includes 3rd and 4th ventricle.

Structure		hippocampus †	inf. lat. vent. †	lat. vent. †	ventricles †	ventricles* ‡	cortical GM ‡	deep GM ‡
Atrophy rates	Normal	-1.1 (1.5)	1.8 (3.5)	3.1 (3.2)	2.5 (3.1)	2.9 (3.0)	-0.9 (1.1)	-0.9 (1.2)
	sMCI	-2.0 (2.4)	3.1 (3.9)	3.9 (3.7)	3.5 (3.4)	3.7 (3.4)	-1.1 (1.3)	-0.9 (1.3)
	pMCI	-3.9 (2.6)	6.3 (4.4)	7.2 (4.0)	6.7 (3.7)	6.8 (3.7)	-1.7 (1.4)	-1.2 (0.9)
	AD	-5.1 (3.8)	8.0 (5.2)	7.8 (4.8)	7.9 (4.5)	7.5 (4.6)	-2.0 (1.6)	-1.4 (1.2)
Sample sizes	sMCI (uncor.)	367	413	226	236	216	342	558
	sMCI (CN-cor.)	1651	2365	5459	2676	4208	11186	1083247
	pMCI (uncor.)	113	123	80	77	77	180	147
	pMCI (CN-cor.)	212	239	251	191	231	877	2316
	MCI (uncor.)	223	233	136	140	132	268	338
	MCI (CN-cor.)	577	636	730	537	661	2521	42739
	AD (uncor.)	139	105	96	81	95	166	189
	AD (CN-cor.)	221	173	269	171	250	581	1659

4.2.3.6 Longitudinal analysis of TBI patients

To evaluate the proposed method in a realistic clinical research setting, the longitudinal size change of individual brain structures was assessed with respect to a standard clinical outcome score, the extended Glasgow Outcome Score (GOSe). All 56×2 TBI brain images were preprocessed using the N4 algorithm to correct for intensity inhomogeneities (Tustison et al., 2010). PinCrAM (Heckemann et al., 2015) was employed for brain extraction. Subsequently MALP-EM4D was applied to the 56 image pairs. Clinical GOSe scores were available (GOSe 3: 4 subjects, GOSe 4: 3, GOSe 5: 4, GOSe 6: 11, GOSe 7: 15, GOSe 8: 19). Visual inspection confirmed the robustness of the method for all processed images. Figure 4.5 shows results calculated on one of the most pathological subjects.

In a preliminary experiment to underline the potential of the proposed method, the TBI subjects were split into two groups⁵. Following Jennett et al. (1981) the subjects were split into those with GOSe 3-5 (severe or lower moderate disability) and GOSe 6-8 (upper moderate disability, good recovery).

Atrophy rates for all 134 structures were calculated and structures identified for which the size change rate was significantly different (two-sided unpaired Student's t-test) between the groups. Paired structures appearing both in the left and right brain hemisphere were combined. The four most significant subcortical structures were cerebral white matter ($p < 10^{-6}$), thalamus ($p \approx 10^{-4}$), lateral ventricles ($p \approx 10^{-4}$) and caudate ($p < 10^{-3}$). This is in line with published literature that suggests atrophy both in the thalamus and caudate (*cf.* Section 2.4). Without temporal coupling group differences were far less marked, *e.g.* for the thalamus and caudate there was no significant difference. An exception was the pallidum which was found significant only without the temporal coupling ($p < 10^{-3}$).

The boxplots shown in Figure 4.5 illustrate the substantial reduction of the size of the thalamus (GOSe 3-5: $-4.8 \pm 4.6\%$ vs. GOSe 6-8: $-0.8 \pm 2.2\%$) and the caudate ($-7.8 \pm 12\%$ vs. $1.1 \pm 4.5\%$). In contrast, ventricles expand faster, due to brain atrophy in more severely injured patients ($15.8 \pm 16.7\%$ vs. $1.6 \pm 8.4\%$). Such a clear group separation is crucial for outcome prediction: 100 iterations of a 6-fold cross validation were run to separate GOSe groups 3-5 and 6-8 based on thalamic atrophy using LDA. While the classification accuracy without temporal coupling was 62%

⁵Note that confounding factors such as age differences between the groups were not considered in this preliminary experiments. This is in contrast to the analysis presented in Chapter 6.

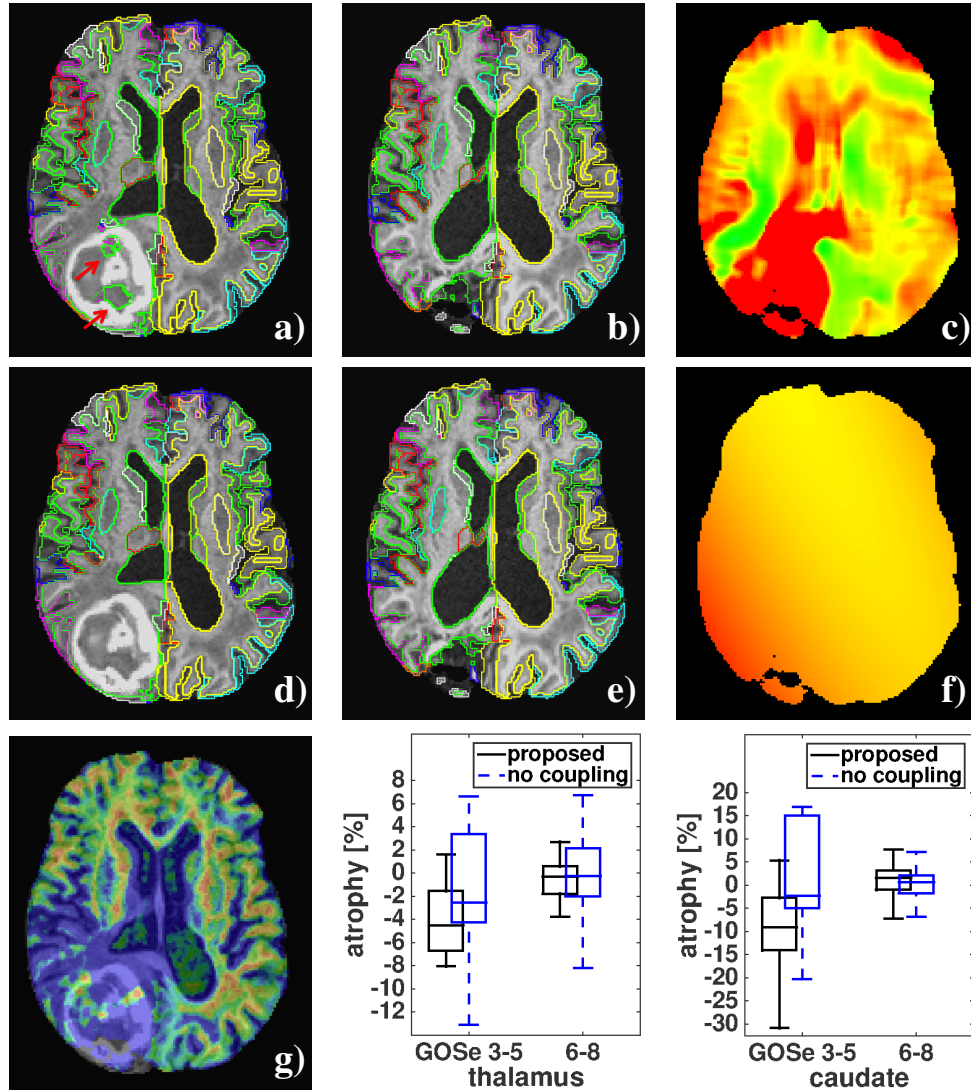


Figure 4.5: Visual results for a TBI subject (TBI066, GOSe = 3) with significant pathologies obtained with the proposed method. (a,d) T_1 -weighted baseline, (b,e) 7.9 month follow-up scan with overlaid segmentations and (c,f) corresponding differential bias fields. Results were calculated with unweighted (a,b) and with weighted (d,e) differential bias field correction. (c) differential bias field estimated as presented in Lewis and Fox (2004), (f) proposed polynomial fitted bias field. (g) corresponding temporal coupling weights according to Equation 4.10 (high weights in red). Boxplots of volume change in the thalamus and caudate calculated with (black) / without (blue) the proposed temporal coupling as described in Section 4.2.2.3-4.2.2.4 for GOSe groups 3-5 vs. 6-8.

(SENS: 62%, SPEC: 62%) a high accuracy of 81% (70%, 83%) was observed using the proposed method with temporal coupling. Similar results were observed for the caudate: without temporal coupling 63% (31%, 71%), with temporal coupling 83% (65%, 87%).

4.2.4 Conclusion

A novel method was presented that enables the symmetric and consistent segmentation of pathological serially acquired MR brain images into a large number of anatomical structures. The segmentation consistency was significantly improved in comparison to manual and state-of-the-art procedures. On the same subset of subjects with potential AD that was used in Section 4.1 it was shown that the proposed method allows the calculation of atrophy rates for multiple structures simultaneously. Computed atrophy rates are realistic and sample sizes comparable to those calculated by established techniques that focus on individual structures only. By application to a novel cohort of patients with TBI, it was confirmed that MALP-EM4D allows the derivation of discriminative prognostic markers in longitudinal imaging. The findings also indicate the method's potential to assist in the challenging task of monitoring the development of pathological conditions. In the future, the potential of the derived measurements as markers of AD (*cf.* Chapter 5) and TBI (*cf.* Chapter 6) disease progression needs to be investigated in greater detail. A further promising avenue would be the integration of a deformation model to estimate the structural disease progression over time.

5 Automated brain morphometry in patients with Alzheimer's disease

Abstract: *In Chapter 3 and Chapter 4 methodology was developed and evaluated to segment T1w MR brain images into numerous distinct anatomical structures based on the manually annotated NMM brain atlas described in Section 2.2.5. In this chapter, these techniques are applied to analyse the ADNI cohort. The aim is the automatic extraction of clinically-relevant features (structural volumes, atrophy rates) from brain images of healthy subjects and patients with probable AD. Specifically, a cross-sectional analysis is performed of images acquired at baseline and a longitudinal analysis using also the MR images acquired during the month-12 (m12) or month-24 (m24) follow-up visits. The presented results demonstrate that the extracted volumetric features are informative and confirm several findings published in the literature. Atrophy rates for numerous distinct brain structures are calculated and their discriminative power investigated to separate both AD patients from healthy controls but also progressive from stable MCI patients. The effect of age, gender and head size on structural volumes is briefly investigated.*

ADNI is a multi-center study (over 50 acquisition sites in North America) with the aim to identify biomarkers that support the early detection of AD and that allow an improved understanding of the progression of the disease. Collected biomarkers include clinical information (*e.g.* demographics, cognitive assessment), imaging biomarkers (MRI, PET, DWI), genetic data (*e.g.* ApoE genotype) and also biospecimen data (*e.g.* from blood or CSF). The first ADNI study, ADNI-1, started in 2004 and was subsequently continued as ADNI-GO and ADNI-2, which is still ongoing. Throughout the three studies more than 1500 participants were enrolled including healthy control subjects and patients diagnosed with different stages of MCI or AD. More information on ADNI can be found at: <http://adni.loni.usc.edu> or <http://www.adni-info.org> (last accessed 29 June 2015).

In the following, an overview over the imaging data is given in Section 5.1. The employed methodology is summarised in Section 5.2. In Section 5.3 the extracted imaging features and the classification setup is described. In Section 5.4 a widely used approach is described to correct for multiple nuisance variables. In Section 5.5.1 the effect of correcting for nuisance factors is investigated and a cross-sectional analysis performed using imaging information acquired at baseline. A longitudinal analysis is conducted in Section 5.5.2. Finally a discussion of the results is provided in Section 5.6 and Section 5.7 concludes this chapter.

5.1 Materials

In this chapter a subset of T1w MR brain images is analysed from all phases (ADNI-1/-GO/-2) of the ADNI study. On 19/02/2015 the image repository and the respective clinical information was updated based on the ADNIMERGE package¹. The repository contained 7128 images of which 1723 were baseline images. As imaging data is still added to the ADNI database the cohort is continuously growing. This is why at the time of this analysis not all of the 7128 images have been processed. Subjects were not explicitly selected and no images were excluded due to reasons related to image processing. Specifically, 5074 images were processed of which 1674 were baseline images. In the following a subset from these 5074 processed images is selected based on clinical information. Here the clinical information from the complete repository was used, *i.e.* also from follow-up images that were not processed.

Certain criteria were defined for the inclusion/exclusion of individual subjects which are summarised below:

- All subjects that reverted at any time point from a more severe to a less severe disease stage, *i.e.* AD $\rightarrow$ MCI or MCI $\rightarrow$ HC, were excluded from both the cross-sectional (N=68) and the longitudinal analysis.
- All subjects with baseline diagnosis early MCI (N=277) or subjective memory complaints (SMC) (N=76) were excluded from both the cross-sectional and the longitudinal analysis.
- The *sMCI* group was defined as those subjects that were diagnosed as MCI (called 'late MCI' in ADNI-GO/-2) at baseline and remained at the MCI stage for at least two years and

¹The ADNIMERGE package is updated daily and summarises commonly used data from all three ADNI studies (ADNI-1/-GO/-2). More information can be found at: <http://adni.bitbucket.org/ADNIMERGE-package.html>, last accessed 15 July 2015.

until the most recent diagnosis which was available. This means that subjects for which this information was not available, *e.g.* because the repository did not contain a corresponding m24 or later visit, were excluded (N=130).

- The *pMCI* group was defined as those subjects that were diagnosed as MCI at baseline and converted within two years follow-up to a probable AD diagnosis.
- Subjects that were diagnosed as MCI at baseline but converted to probable AD more than two years later (N=54) were excluded from the analyses and neither considered as sMCI nor pMCI.
- The m12 image (I373205) of one subject (ADNI_007_S_4568) was reviewed after irregular volume measurements and excluded manually from the analysis due to poor image quality².
- All subjects summarised in Table 5.1 fulfil the above criteria. For the longitudinal analysis, however, all subjects that converted at any time of the study from HC to a symptomatic stage (*e.g.* MCI or even AD) were excluded (N=52).

Table 5.1 gives an overview over the baseline images considered in the conducted analyses. Lists of the processed filenames that also include the unique image identifier and diagnosis are available online³⁴⁵.

Table 5.1: Overview of the processed subjects from the ADNI cohort, including age and clinical information at baseline.

	all	HC/CN	sMCI	pMCI	AD
# of subjects/images at baseline	1069	404	166	177	322
gender (# male / # female)	581 / 488	202 / 202	98 / 68	104 / 73	177 / 145
years of age (median [min; max])	74.6 [48.1; 91.4]	74.2 [59.8; 89.6]	74.4 [55.9; 91.4]	74.3 [48.1; 88.3]	75.8 [55.1; 91.4]
ApoE4 (# 0 / # 1 / # 2) [†]	547 / 407 / 113	293 / 101 / 9	93 / 60 / 13	57 / 91 / 29	104 / 155 / 62
MMSE (median [min; max])	27 [18; 30]	29 [24; 30]	28 [24; 30]	26 [23; 30]	23 [18; 27]
FAQ (median [min; max]) [‡]	1 [0; 30]	0 [0; 6]	1 [0; 21]	5 [0; 21]	13 [0; 30]
CDRSB (median [min; max])	1.5 [0; 10]	0 [0; 1]	1.5 [0.5; 4]	2 [0.5; 5]	4.5 [1; 10]
FieldStrength (1.5T/3T)	653 / 416	223 / 181	112 / 54	129 / 48	189 / 133
# of subjects/images at month 12	802	195	149	168	290
# of subjects/images at month 24	532	168	107	140	117

[†]: not available for 2 subjects, [‡]: not available for 1 subject

ApoE: apolipoprotein E

MMSE: mini-mental state examination

FAQ: functional activities questionnaire

CDRSB: clinical dementia rating scale sum of boxes

²The ADNI cohort provides another (accelerated) m12 image for this subject of good quality. However, this other image was not processed to avoid inconsistencies with the script that chooses an image in case of multiple possibilities.

³http://www.doc.ic.ac.uk/~cl6311/PhD_thesis/ADNI_baseline_1069.csv, last accessed 29 June 2015.

⁴http://www.doc.ic.ac.uk/~cl6311/PhD_thesis/ADNI_m12_802.csv, last accessed 29 June 2015.

⁵http://www.doc.ic.ac.uk/~cl6311/PhD_thesis/ADNI_m24_532.csv, last accessed 29 June 2015.

5.2 Methods

As the preprocessed images were downloaded from ADNI no additional preprocessing was performed, such as smoothing or a correction for intensity inhomogeneities (Tustison et al., 2010). Brain masks were calculated for all available baseline images using pincran (Heckemann et al., 2015). Brain masks were visually reviewed and some were recalculated with an updated pincran atlas database. Follow-up images were brain extracted utilising the corresponding baseline brain masks which were transformed using rigid intrasubject registration. To avoid failures of this rigid registration procedure a threshold was set on the image similarity of the follow-up image and transformed baseline image. Transformations leading to an image similarity below a set threshold were recalculated with updated parameters.

All baseline and follow-up (m12, m24) images were subsequently segmented using MALP-EM as described in Chapter 3. Specifically, the 30 manually annotated brain atlases (*cf.* Section 2.2.5) were propagated based on transformations calculated with the robust registration approach MAPER. For the fusion of the propagated atlases the locally weighted fusion approach was chosen as it is computationally more efficient and of similar accuracy compared to the patch-based joint label fusion (Wang et al., 2013b). The probabilistic segmentation output and the intensity normalised, brain extracted images were then employed to perform the consistent longitudinal segmentation described as MALP-EM4D in Chapter 4. The standard, unweighted differential bias correction approach was applied as described in Lewis and Fox (2004). Implementation and parameters were otherwise identical to the experiments presented in Chapter 4.

5.3 Features and classification

For the *cross-sectional* analysis at baseline structural volumes were extracted for each individual brain label available through the NMM brain atlas (*cf.* Section 2.2.5). Note that the baseline segmentations were calculated with MALP-EM and not MALP-EM4D as the latter exploits information of later scanning time points, which is not available at baseline.

For the *longitudinal* analysis structural volumes were extracted based on the respective MALP-EM4D segmentations and atrophy rates calculated as described in Section 4.2.3.2.

As features all available structural volumes were employed, however, corresponding volumes of the left and right brain hemisphere were merged ($98/2=49$ cortical plus $28/2+7=21$ non-cortical

features). Note that the seven non-cortical structures 3rd ventricles, 4th ventricle, brain stem, CSF, cerebellar vermal lobule I-V, cerebellar vermal lobule VI-VII and cerebellar vermal lobule VIII-X have no symmetric counterpart. An exception are the amygdala, hippocampus, inferior lateral ventricles and the lateral ventricles. As it is expected that these structures are particularly informative their respective left/right volumes were retained as separate features (8 features). This allows to investigate whether there is substantial asymmetric involvement of these structures in the disease progression. Individual structures were further summarised as ventricles, cortical grey matter, deep grey matter, white matter, brain tissue and total brain volume (brain tissue including ventricles/CSF) (6 features). The optic chiasm was excluded from the analysis due to its very small size. In total 86 features were considered, including age and gender.

For classification a 6-fold cross-validation (CV) was chosen using a LDA classifier for individual features. When combining multiple features both SVM and RF classifiers were employed to investigate their respective performance. For individual features (LDA) 100 CV repetitions and for the multi-feature classifiers (SVM, RF) 20 repetitions of the CV were performed. A classification framework was implemented using MATLAB that relies on `classify` (LDA), `TreeBagger` (RF, 100 trees) and `libSVM` (SVM, (Chang and Lin, 2011)). For the SVM classification the features were normalised individually to the range 0 to 1. This avoids that features with larger magnitude have a stronger impact on the classification. The normalisation statistics were calculated based only on the respective training set to avoid exploiting information from the test set and thus the introduction of bias. Both the LDA classifier, which was used for single-feature classification only, and the RF classifier do not require feature normalisation.

Significance levels were quantified as p-values of two-sided, unpaired Student's t-tests. Further, effect sizes were calculated as *Cohen's d*⁶ by dividing the differences of the sample means (absolute value) by their pooled standard deviation (Cohen, 1988; Fennema-Notestine et al., 2009; Wang et al., 2013b). Reporting the effect size in addition to the p-value is important as it quantifies the magnitude of a group difference, while a low p-value only confirms its existence (Sullivan and Feinn, 2012).

⁶According to Cohen (1988) an effect size of $d = 0.2$ can be considered as small, of $d = 0.5$ as medium and of $d = 0.8$ as large.

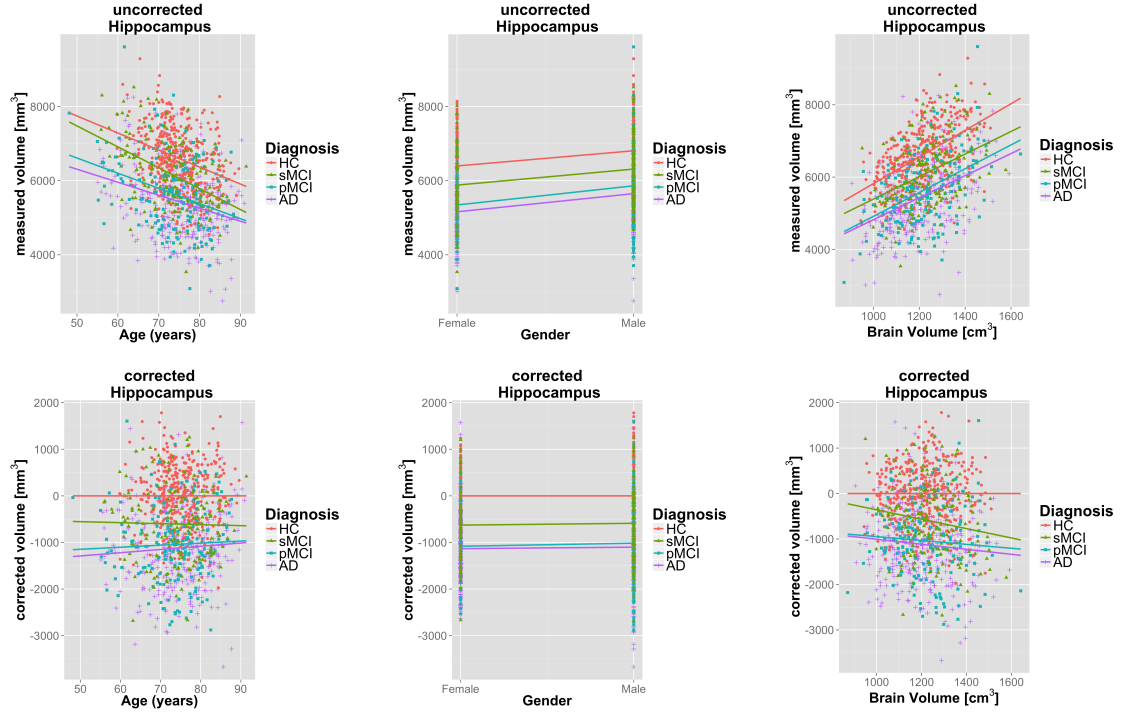


Figure 5.1: Top: Dependence of hippocampal volume on age (left), gender (middle) and brain volume (right). Bottom: Corresponding volumes corrected for nuisance factors age, gender and brain size. Overlaid regression lines for distinct disease groups with corresponding regression lines.

5.4 Correction of nuisance factors

It is known that the volume of many individual brain structures reduces during the course of normal, healthy aging. Also strong correlations between structural volumes and overall head size are well established (Barnes et al., 2010). Even though less pronounced, significant gender-specific differences between groups with identical brain sizes were reported in the literature (Luders et al., 2009; Barnes et al., 2010). The variation of structural volumes was calculated by fitting linear regressors with respect to age, gender and brain volume based on the processed healthy control subjects. The results are illustrated for the hippocampus in Figure 5.1.

The division of each structural volume by head or brain size is potentially inadequate (Barnes et al., 2010). In this work a multivariate linear regression approach was followed as described in Koikkalainen et al. (2012) to investigate and correct for three nuisance factors, namely: *age*, *gender* and *brain size*. Specifically, a separate linear regressor is fit using the processed healthy control subjects for each individual structure using the above-mentioned predictor variables. Each multi-

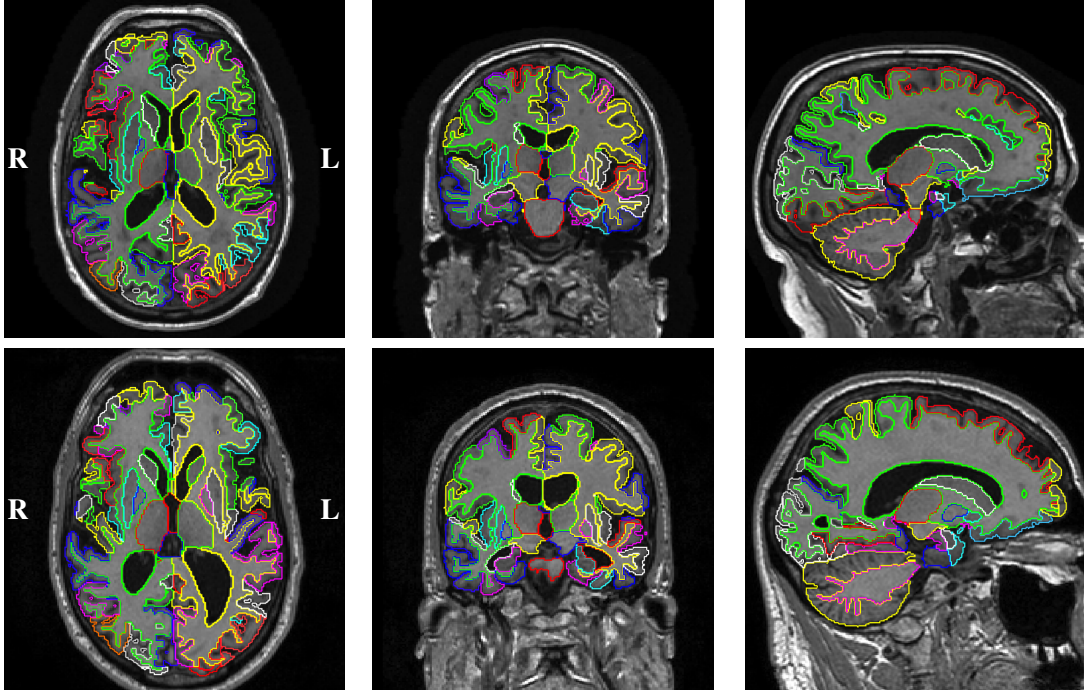


Figure 5.2: Example cross-sectional segmentation results of baseline images of a healthy control subject (top, ADNI_005_S_0553, male, 85 years of age) and a patient diagnosed AD (bottom, ADNI_018_S_0286, male, 66 years of age) in axial (left), coronal (middle) and sagittal (right) view-plane.

variate regressor is defined by the slope for each predictor c_{age} , c_{gender} and c_{size} and an intercept b .

Assuming M distinct features, feature m of subject n , denoted as F_n^m , is corrected as:

$$\tilde{F}_n^m = F_n^m - (c_{\text{age}}^m \text{age}_n + c_{\text{gender}}^m \text{gender}_n + c_{\text{size}}^m \text{size}_n + b^m) \quad (5.1)$$

To correct for head size, the total brain size (sum of all structures) was used as an approximation of the intracranial volume (ICV). This is a commonly used approximation (Buckner et al., 2004; Smith et al., 2007; Barnes et al., 2010).

5.5 Results

5.5.1 Cross-sectional analysis

This cross-sectional analysis investigates the potential of volumes of individual brain ROIs to discriminate different AD disease stages. An example segmentation result of a healthy control subject and a patient diagnosed with AD is illustrated in Figure 5.2.

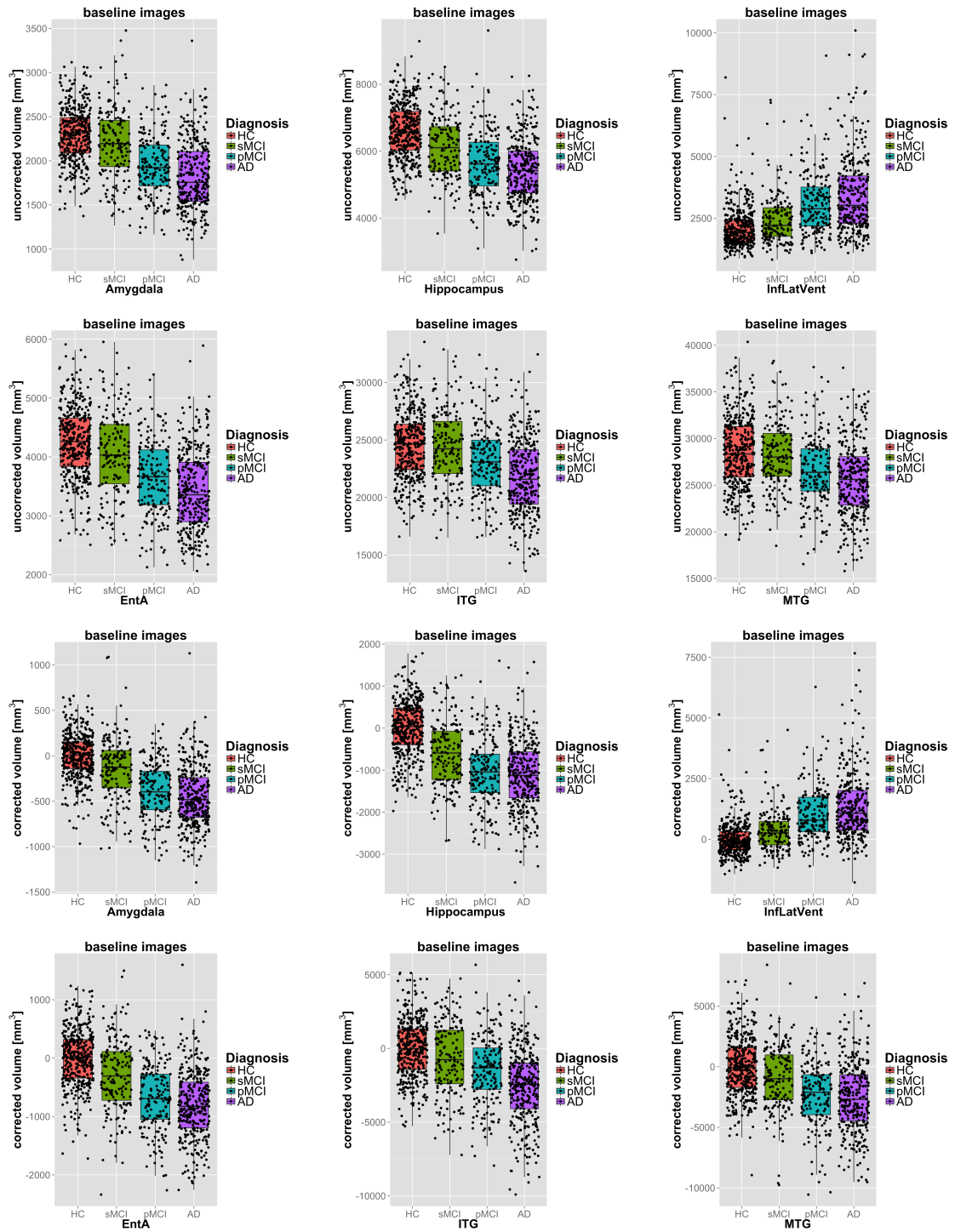


Figure 5.3: Boxplots⁷ of structural volumes at baseline for six selected structures before (top two rows) and after (bottom two rows) correcting for nuisance factors for distinct disease groups. Note that the corrected volumes need to be read with respect to a healthy population of matched age, gender and brain size. Structures were selected based on their performance in classifying the investigated disease groups (*cf.* Table 5.3).

⁷Boxplots were created with the ggplot2 package of R. Data points were jittered along x-axis for better visualisation. A more detailed description of geom_boxplot including the definition of error bars can be found at: http://docs.ggplot2.org/0.9.3/geom_boxplot.html, last accessed: 12 June 2015.

Table 5.2: Balanced classification accuracies in % for separating healthy control and AD subjects (effect sizes in parentheses) after correcting for different nuisance factors (100 runs, 6-fold cross-validation, LDA). Highest effect size in bold.

structures / correction	none	age	brain size	gender	age, brain size, gender
Ventricles	65 (0.70)	65 (0.72)	68 (0.86)	65 (0.71)	71 (0.90)
CorticalGreyMatter	61 (0.53)	62 (0.53)	67 (0.86)	62 (0.67)	67 (0.90)
amygdala	75 (1.35)	76 (1.40)	79 (1.48)	77 (1.42)	80 (1.56)
hippocampus	75 (1.33)	75 (1.38)	77 (1.45)	76 (1.40)	78 (1.52)
EntA	73 (1.28)	74 (1.33)	75 (1.39)	76 (1.41)	78 (1.51)
InfLatVent	72 (1.10)	72 (1.16)	73 (1.19)	72 (1.14)	77 (1.27)
ITG	66 (0.89)	66 (0.89)	71 (1.13)	71 (1.07)	71 (1.17)
MTG	64 (0.74)	63 (0.73)	69 (0.97)	66 (0.86)	69 (0.97)

The distribution of the uncorrected volumes of six selected structures is shown in Figure 5.3 for the four disease groups (HC, sMCI, pMCI and AD). The reduced grey matter volume of structures in the medial temporal lobe and the increased ventricular volume in patients with AD is apparent.

In the following Section 5.5.1.1 the effect of correcting the structural volumes for patient age, gender and brain size is evaluated. In Section 5.5.1.2 these corrected volumes will be employed to investigate their potential for the classification relevant disease stages in AD.

5.5.1.1 Benefit of correcting for nuisance factors

The effect of correcting for the nuisance factors patient age, gender and head size is summarised in Table 5.2. The conducted correction substantially increased classification accuracies and effect sizes obtained on individual structures. The benefit of the performed correction for all investigated structures, including effect sizes, can be found in Appendix E in Table E.1 and Table E.2. A visualisation of the distributions of hippocampal volume in Figure 5.4 confirms that group separation is improved (higher peaks) after the correction procedure. This is in agreement with the results presented in Koikkalainen et al. (2012). Correcting for brain size had the biggest effect. In the conducted experiment correcting for gender had a stronger impact than correcting for age. Correcting for all three nuisance factors, age, gender and brain size, was most beneficial in terms of both classification accuracy and effect size. Note that experiments were performed to regress out magnetic field strength as small but significant differences were observed between volume measurements of the healthy control subjects scanned at 1.5T and 3T. However, preliminary results suggested that correcting for field strength has no significant benefit. This was not further pursued and no explicit correction for field strength differences applied in the following.

A visual illustration of the volumes before and after correction is provided in Figure 5.1 and

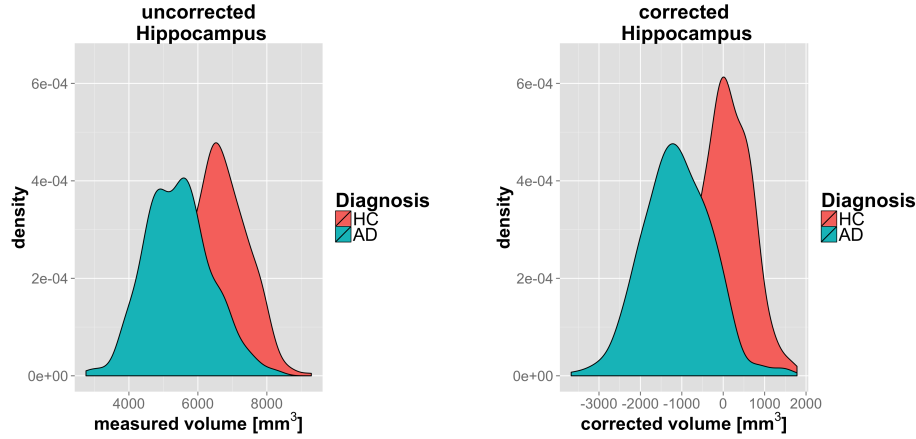


Figure 5.4: Density plot of the hippocampal volumes before (left) and after (right) correction for age, gender and brain size. Note that the corrected volumes need to be interpreted with respect to the healthy reference group and have therefore zero-mean for HCs.

Figure 5.3. It can be seen that the healthy control population has zero mean after the correction and the overall dependence on nuisance factors is clearly reduced.

For the cross-sectional analysis the corrected volumes are used. This means that feature values are no longer actual volumes, but rather volume differences with respect to a healthy population of matched age, gender and brain size. Note that the three independent variables patient age, gender and head size were not corrected.

5.5.1.2 Disease classification using structural volumes

An overview over the results of the conducted classification experiments is given in Table 5.3. The classification results for all structures can be found in Appendix E in Table E.1 (uncorrected volumes, AD vs. HC), Table E.2 (AD vs. HC) and Table E.3 (pMCI vs. sMCI).

The volume of the amygdala allows the best group separation between AD and HC with an effect size (Cohen's d) of 1.56 and a bACC of 80% (SENS: 76%, SPEC: 84%). Other structures in the MTL region such as the hippocampus (bACC: 78%) or the entorhinal area (bACC: 78%) were similarly discriminative ($d > 1.5$). Established findings that overall brain tissue, and cortical grey matter in particular, are of smaller size in patients with AD were confirmed. By combining all considered structural volumes as features in a SVM (bACC: 89%, SENS: 86%, SPEC: 92%) or RF (bACC: 86%, SENS: 83%, SPEC: 90%) accuracy was further improved. As indicated in Table 5.3 numerous structural differences were highly significant even after Bonferroni correction

Table 5.3: Classification results in % (6-fold cross-validation, LDA 100 runs, RF/SVM 20 runs) obtained separating AD from CN (top) and stable from progressive MCI (bottom). Individual structures are sorted by effect size. The 20 structures with highest effect size are listed explicitly. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses. Features were corrected for age, gender and brain size.

AD patients (N = 322, Positives^P) vs. Healthy Controls (N = 404, Negatives^N) (baseline analysis, [†]volumes corrected for age/gender/brain size)									
structure	ACC (bACC)	SENS	SPEC	mean (SD) [mm ³] ^{P,†}	mean (SD) [mm ³] ^{N,†}	effect size (d)	p-value	sig. (corr.)	
RandomForest (all features)	87 (86)	83	90						
SVM (all features)	90 (89)	86	92						
BrainTissue	72 (71)	63	78	-17942.8 (22718.7)	0 (17403.6)	0.900	<0.00001	++	(++)
CorticalGreyMatter	68 (67)	63	72	-23635.9 (28907.3)	0 (24078.1)	0.898	<0.00001	++	(++)
Ventricles	72 (71)	63	79	17757.3 (22639.6)	0 (17293.9)	0.895	<0.00001	++	(++)
WhiteMatter	57 (56)	51	61	7242.2 (29309.8)	0 (28686.1)	0.250	0.00086	++	(o)
DeepGreyMatter	52 (52)	50	54	-1549.1 (13310.2)	0 (11142.2)	0.127	0.08834		(o)
Brain	54 (55)	56	53	1199158.5 (132290.9)	1212115.6 (118623.9)	0.104	0.16526		(o)
Amygdala	80 (80)	76	84	-452.0 (332.0)	0 (250.5)	1.561	<0.00001	++	(++)
Hippocampus	78 (78)	75	80	-1115.4 (817.7)	0 (660.7)	1.519	<0.00001	++	(++)
EntA	78 (78)	76	80	-801.3 (583.3)	0 (485.3)	1.509	<0.00001	++	(++)
LeftHippocampus	79 (78)	76	81	-588.1 (423.1)	0 (364.6)	1.502	<0.00001	++	(++)
RightAmygdala	80 (80)	77	83	-232.2 (182.6)	0 (139.4)	1.452	<0.00001	++	(++)
LeftAmygdala	78 (78)	75	81	-219.8 (179.7)	0 (135.5)	1.403	<0.00001	++	(++)
RightHippocampus	76 (75)	71	79	-527.3 (488.0)	0 (339.1)	1.280	<0.00001	++	(++)
InfLatVent	78 (77)	65	89	1330.8 (1367.4)	0 (702.2)	1.267	<0.00001	++	(++)
LeftInfLatVent	77 (76)	66	86	649.5 (677.1)	0 (360.6)	1.237	<0.00001	++	(++)
ITG	71 (71)	69	73	-2588.0 (2506.9)	0 (1954.5)	1.168	<0.00001	++	(++)
RightInfLatVent	75 (73)	59	87	681.3 (814.9)	0 (391.3)	1.106	<0.00001	++	(++)
PHG	68 (68)	64	72	-739.8 (832.6)	0 (702.4)	0.970	<0.00001	++	(++)
MTG	69 (69)	69	70	-2534.0 (2853.0)	0 (2411.3)	0.969	<0.00001	++	(++)
AnG	68 (67)	65	70	-1880.5 (2345.6)	0 (1973.2)	0.876	<0.00001	++	(++)
TMP	66 (65)	64	67	-1521.7 (1967.2)	0 (1539.3)	0.873	<0.00001	++	(++)
LateralVentricle	71 (70)	62	78	15972.2 (21196.2)	0 (16214.4)	0.859	<0.00001	++	(++)
LeftLateralVentricle	70 (69)	61	76	8584.7 (11586.9)	0 (8816.1)	0.847	<0.00001	++	(++)
RightLateralVentricle	70 (69)	59	78	7387.5 (10387.1)	0 (7883.2)	0.814	<0.00001	++	(++)
STG	65 (65)	64	66	-1100.3 (1610.7)	0 (1478.6)	0.715	<0.00001	++	(++)
3rdVentricle	67 (66)	63	70	434.3 (699.9)	0 (591.1)	0.677	<0.00001	++	(++)

progressive MCI (N = 177, Positives^P) vs. stable MCI (N = 166, Negatives^N) (baseline analysis, [†]volumes corrected for age/gender/brain size)									
structure	ACC (bACC)	SENS	SPEC	mean (SD) [mm ³] ^{P,†}	mean (SD) [mm ³] ^{N,†}	effect size (d)	p-value	sig. (corr.)	
RandomForest (all features)	68 (68)	72	64						
SVM (all features)	67 (67)	70	64						
BrainTissue	60 (60)	51	69	-12652.2 (20144.6)	-4372.1 (20741.3)	0.405	0.00021	++	(+)
Ventricles	60 (60)	51	70	12483.5 (20045.3)	4427.7 (20676.8)	0.396	0.00029	++	(+)
CorticalGreyMatter	56 (56)	55	56	-16743.9 (27348.8)	-6929.4 (26948.6)	0.361	0.00091	++	(o)
Brain	51 (51)	52	49	1222424.9 (131084.2)	1238635.5 (118830.2)	0.129	0.23200		(o)
WhiteMatter	48 (48)	47	49	5966.3 (27919.1)	4607.3 (26735.5)	0.050	0.64592		(o)
DeepGreyMatter	47 (47)	46	48	-1874.6 (11838.0)	-2050.0 (11988.8)	0.015	0.89167		(o)
Amygdala	65 (65)	63	68	-387.4 (308.7)	-149.4 (352.2)	0.720	<0.00001	++	(++)
LeftAmygdala	62 (62)	61	64	-191.6 (161.2)	-70.4 (189.7)	0.690	<0.00001	++	(++)
InfLatVent	64 (64)	54	74	1034.0 (1097.9)	358.6 (903.1)	0.670	<0.00001	++	(++)
LeftInfLatVent	65 (65)	56	75	480.0 (538.2)	156.3 (451.1)	0.650	<0.00001	++	(++)
RightAmygdala	65 (65)	63	67	-195.8 (176.7)	-79.0 (184.1)	0.648	<0.00001	++	(++)
EntA	61 (61)	61	60	-678.3 (565.7)	-294.0 (622.8)	0.647	<0.00001	++	(++)
RightInfLatVent	64 (64)	54	74	554.0 (689.1)	202.3 (527.5)	0.571	<0.00001	++	(++)
Hippocampus	61 (61)	63	60	-1042.8 (817.1)	-604.7 (786.1)	0.546	<0.00001	++	(++)
RightHippocampus	63 (63)	61	64	-514.7 (442.9)	-287.7 (408.3)	0.532	<0.00001	++	(++)
MTG	60 (60)	59	60	-2305.3 (2698.1)	-925.8 (2708.9)	0.510	<0.00001	++	(++)
AnG	60 (60)	61	59	-1579.0 (2223.0)	-495.7 (2084.0)	0.502	<0.00001	++	(++)
LeftHippocampus	61 (61)	62	60	-528.1 (429.6)	-317.0 (425.8)	0.493	<0.00001	++	(++)
CSF	59 (59)	54	64	156.4 (533.9)	-61.6 (491.3)	0.424	0.00010	++	(+)
3rdVentricle	59 (59)	55	64	376.2 (661.3)	113.9 (639.9)	0.403	0.00023	++	(+)
LateralVentricle	60 (61)	51	70	11080.2 (18700.0)	3986.3 (19536.4)	0.371	0.00066	++	(o)
RightLateralVentricle	60 (60)	51	69	5078.1 (9151.1)	1656.3 (9510.1)	0.367	0.00077	++	(o)
LeftLateralVentricle	59 (59)	51	68	6002.1 (10336.4)	2330.0 (10722.7)	0.349	0.00136		(o)
STG	57 (57)	57	56	-1048.8 (1509.9)	-513.8 (1583.2)	0.346	0.00149		(o)
ITG	57 (57)	57	56	-1510.5 (2300.7)	-691.2 (2482.3)	0.343	0.00165		(o)
TMP	58 (58)	58	58	-875.3 (1703.2)	-310.1 (1766.7)	0.326	0.00275		(+)

for multiple comparisons.

For the classification of pMCI vs. sMCI the structural volumes of the amygdala (bACC: 65%, SENS: 63%, SPEC: 68%) and inferior lateral ventricles (bACC: 64%, SENS: 54%, SPEC: 74%) were most discriminative. Combining all extracted structural volumes into a single RF classifier further increased classification accuracies to bACC: 68%, SENS: 72%, SPEC: 64%.

5.5.2 Longitudinal analysis

This longitudinal analysis investigates the volume change of individual ROIs with respect to disease stage. Atrophy was measured for the m12 or m24 follow-up images with respect to their corresponding baseline images. The calculated atrophy rates and the resulting sample sizes are presented in Section 5.5.2.1. Their potential as features to discriminate disease stages is investigated in Section 5.5.2.2. An example segmentation result obtained on a subject diagnosed with AD using MALP-EM4D is shown in Figure 5.5.

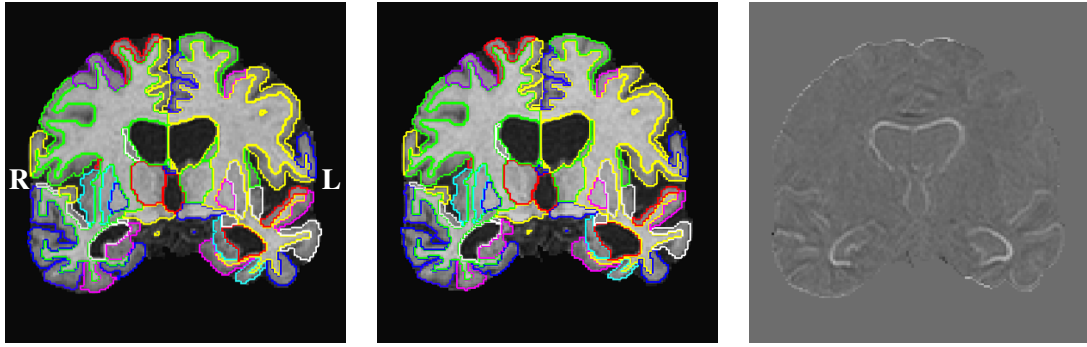


Figure 5.5: Example longitudinal segmentation results of baseline (left) and month 24 (middle) follow-up images of a patient diagnosed AD (ADNI_018_S_0286) in axial view-plane. Substantial hippocampal atrophy (measured: -7.81%) and ventricular enlargement (16.5%) is apparent in the difference image after affine registration (right).

5.5.2.1 Atrophy rates and sample sizes

Atrophy rates for selected structures can be found in Table 5.5 for AD/HC and in Table 5.6 for pMCI/sMCI. Corresponding atrophy rates for all investigated structures can be found in Appendix E. The distribution of the volume change of six selected structures is shown in Figure 5.6 for the four investigated disease groups (HC, sMCI, pMCI and AD).

In AD patients grey matter structures such as the hippocampus (HC: -1.1%, AD: -4.8%) or the

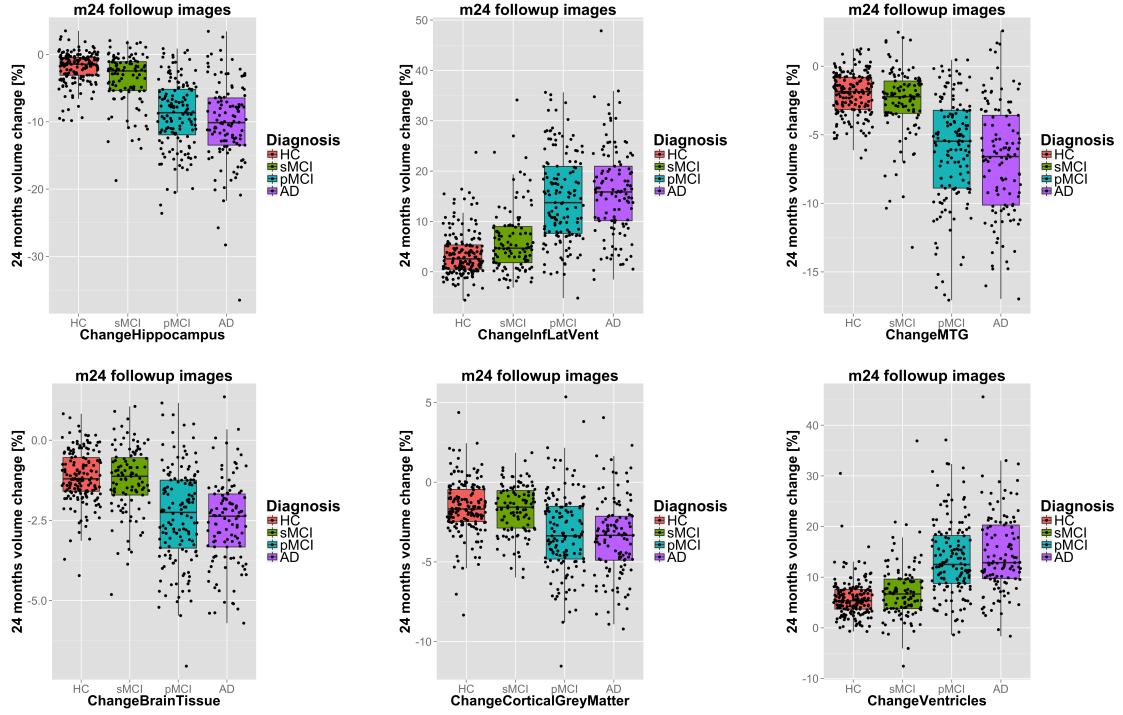


Figure 5.6: Boxplots of volume changes for selected brain structures (top) and surrogate structures (bottom) from baseline to month 24 follow-up image for different disease groups. Features selected based on their performance in classifying the investigated disease groups (*cf.* Table 5.5 and Table 5.6).

medial temporal gyrus (HC: -1.1%, AD: -3.8%) are subject to significant volume loss between baseline and m12. In contrast to this ventricles expand substantially faster in AD patients (HC: 2.8%, AD: 7.2%). Overall sMCI patients show atrophy patterns similar to HC, while MCI patients converting to AD are, in terms of atrophy, very similar to AD patients. Atrophy rates measured between the baseline and m24 follow-up image are in the order of twice the rates measured between baseline and the m12 follow-up image. The results indicate that atrophy in the entorhinal area is slightly higher in pMCI patients than in patients with AD. At m12 an atrophy rate of -4.0% (± 3.7) for pMCI and -3.7% (± 4.1) for AD patients was measured, at m24 for pMCI -7.4% (± 4.7) and -7.0% (± 5.5) for AD patients respectively.

Based on the atrophy rates sample sizes were calculated as outlined in Section 4.1.3.4. The sample sizes were thus calculated to detect a 25% change in atrophy rate with 80% power at a 5% significance level. The atrophy rates and sample sizes for selected structures are summarised in Table 5.4.

The lowest sample sizes (corrected for normal aging) were computed for the inferior lateral

Table 5.4: Mean volume change of selected structures in % with corresponding sample sizes for different disease groups. Standard deviation in parentheses. Measurements based on volume change from baseline to 12 months (top table) or 24 months (bottom table) follow-up visit. Corrected sample sizes were computed on the excess change over normal aging.

bl→m12		hippocampus	inf. lat. vent.	lat. vent.	med. temp. gyr.	brain tissue	ventricles	white matter	cort. GM	deep GM
Atrophy rates	Normal	-1.1 (1.7)	1.8 (3.3)	3.0 (3.2)	-1.1 (1.5)	-0.5 (0.8)	2.8 (3.0)	-0.2 (0.7)	-0.8 (1.4)	-0.8 (1.3)
	sMCI	-1.7 (2.2)	2.6 (3.8)	3.6 (3.3)	-1.4 (1.8)	-0.6 (0.8)	3.4 (3.1)	-0.4 (0.7)	-0.8 (1.3)	-0.7 (1.3)
	pMCI	-4.1 (3.2)	6.7 (5.6)	7.1 (4.6)	-3.1 (2.4)	-1.2 (1.0)	6.8 (4.3)	-0.6 (0.9)	-1.8 (1.7)	-1.3 (1.2)
	AD	-4.8 (3.7)	7.5 (5.5)	7.6 (4.9)	-3.8 (2.7)	-1.3 (1.1)	7.2 (4.6)	-0.8 (1.0)	-1.9 (2.3)	-1.4 (1.6)
Sample sizes	sMCI (uncor.)	412	548	217	384	383	215	966	674	739
	sMCI (CN-cor.)	3130	5190	7499	8823	17269	6337	2834	18432936	1429741
	pMCI (uncor.)	155	175	104	143	168	100	488	235	218
	pMCI (CN-cor.)	284	321	309	351	535	285	864	813	1349
	AD (uncor.)	148	134	104	133	182	101	413	366	336
	AD (CN-cor.)	244	228	285	270	493	265	640	640	1133

bl→m24		hippocampus	inf. lat. vent.	lat. vent.	med. temp. gyr.	brain tissue	ventricles	white matter	cort. GM	deep GM
Atrophy rates	Normal	-2.0 (2.3)	3.5 (4.5)	6.3 (4.0)	-2 (1.6)	-1.1 (0.8)	5.8 (3.7)	-0.6 (0.8)	-1.6 (1.7)	-1.3 (1.3)
	sMCI	-3.7 (3.8)	6.2 (6.5)	7.5 (6.0)	-2.6 (2.6)	-1.2 (1.0)	7.1 (5.6)	-0.8 (1.0)	-1.7 (1.5)	-1.2 (1.0)
	pMCI	-8.9 (5.1)	14.6 (8.9)	14.4 (7.9)	-6.1 (3.9)	-2.3 (1.5)	13.7 (7.4)	-1.6 (1.3)	-3.2 (2.5)	-2.1 (1.4)
	AD	-10.2 (6.2)	15.9 (8.8)	15.4 (8.5)	-6.8 (4.1)	-2.6 (1.3)	14.7 (7.8)	-1.8 (1.4)	-3.5 (2.4)	-2.3 (1.8)
Sample sizes	sMCI (uncor.)	264	274	158	246	173	158	370	215	176
	sMCI (CN-cor.)	1166	1446	6179	4588	23154	5166	4764	102542	23243
	pMCI (uncor.)	82	92	76	105	99	73	163	155	107
	pMCI (CN-cor.)	136	160	241	233	372	220	412	618	782
	AD (uncor.)	93	76	75	92	67	71	145	123	162
	AD (CN-cor.)	142	126	217	185	216	195	316	420	980

ventricles with 228 for bl→m12 and 126 for bl→m24. However, the measured atrophy of other structures such as the hippocampus or the medial temporal gyrus provided sample sizes at a similar level.

5.5.2.2 Disease classification using structural volume change

The potential of structural atrophy rates to separate the clinical groups AD vs. HC and pMCI vs. sMCI was investigated in the following. P-values, effect sizes and classification accuracies were calculated to quantify group separation. The results are summarised in Table 5.5 (AD vs. HC) and Table 5.6 (pMCI vs. sMCI). As features the atrophy rates calculated between the baseline and m12 follow-up image were used, or m24 follow-up image respectively. An overview over all considered features and their individual classification performance can be found in Appendix E.

The most discriminative structure for AD vs. HC separation was the hippocampus ($d_{m12}=1.40$, $d_{m24}=1.88$). Based on m12 atrophy a $bACC_{m12}$ of 78% (SENS: 67%, SPEC: 88%) was calculated, for m24 atrophy a balanced accuracy of 85% respectively. Combining all derived atrophy features in a RF classifier substantially increased classification results to $bACC_{m12}$: 84%; $bACC_{m24}$: 88%.

Hippocampal atrophy was also a very good feature for classifying progressive and stable MCI patients: $bACC_{m12}$: 67%; $bACC_{m24}$: 74%. However, the highest individual classification accuracy was obtained at m12 for the medial temporal gyrus ($bACC_{m12}$: 70%, SENS: 65%, SPEC: 75%) and

Table 5.5: Classification results in % (6-fold cross-validation, LDA 100 runs, RF/SVM 20 runs) for separating AD from CN based on volume change from baseline to m12 (top) or m24 (bottom). Individual structures are sorted by effect size. The 20 structures with highest effect size are listed explicitly. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses.

AD patients (N = 195, Positives^P) vs. Healthy Controls (N = 290, Negatives^N) (longitudinal analysis, bl→m12)									
structure	ACC (bACC)	SENS	SPEC	mean (SD) [%] ^P	mean (SD) [%] ^N	effect size (d)	p-value	sig. (corr.)	
RandomForest (all features)	85 (84)	78	90						
SVM (all features)	84 (82)	71	93						
Ventricles	75 (74)	64	83	7.2 (4.6)	2.8 (3.0)	1.202	<0.00001	++ (++)	
BrainTissue	70 (70)	67	72	-1.3 (1.1)	-0.5 (0.8)	0.862	<0.00001	++ (++)	
WhiteMatter	73 (73)	68	77	-0.8 (1.0)	-0.2 (0.7)	0.736	<0.00001	++ (++)	
CorticalGreyMatter	65 (64)	62	67	-1.9 (2.3)	-0.8 (1.4)	0.585	<0.00001	++ (++)	
Brain	64 (64)	63	65	-0.9 (1.0)	-0.4 (0.7)	0.574	<0.00001	++ (++)	
DeepGreyMatter	66 (64)	57	71	-1.4 (1.7)	-0.8 (1.3)	0.455	<0.00001	++ (++)	
Hippocampus	80 (78)	67	88	-4.8 (3.7)	-1.1 (1.7)	1.400	<0.00001	++ (++)	
InfLatVent	79 (77)	69	86	7.5 (5.5)	1.8 (3.3)	1.334	<0.00001	++ (++)	
LeftHippocampus	80 (78)	70	87	-4.9 (4.2)	-1.1 (1.7)	1.287	<0.00001	++ (++)	
MTG	76 (74)	67	82	-3.8 (2.8)	-1.1 (1.5)	1.274	<0.00001	++ (++)	
RightInfLatVent	77 (75)	65	84	7.4 (6.4)	1.8 (3.3)	1.185	<0.00001	++ (++)	
RightLateralVentricle	75 (73)	63	83	7.5 (5.1)	2.8 (3.1)	1.170	<0.00001	++ (++)	
LeftInfLatVent	78 (76)	70	83	7.5 (5.8)	1.7 (4.4)	1.166	<0.00001	++ (++)	
LateralVentricle	74 (72)	62	83	7.6 (4.9)	3.0 (3.2)	1.160	<0.00001	++ (++)	
RightHippocampus	78 (75)	62	89	-4.8 (4.5)	-1.1 (2.1)	1.138	<0.00001	++ (++)	
LeftLateralVentricle	74 (72)	63	81	7.6 (4.9)	3.1 (3.4)	1.085	<0.00001	++ (++)	
PP	77 (76)	70	81	-6.7 (5.5)	-2.0 (3.5)	1.075	<0.00001	++ (++)	
CerebralWhiteMatter	76 (74)	67	82	-0.9 (0.9)	-0.2 (0.6)	0.914	<0.00001	++ (++)	
PT	70 (69)	62	76	-5.8 (5.3)	-2.1 (3.1)	0.898	<0.00001	++ (++)	
ITG	72 (71)	67	75	-2.7 (2.7)	-1.0 (1.3)	0.866	<0.00001	++ (++)	
FuG	69 (68)	63	73	-1.7 (1.7)	-0.5 (1.1)	0.852	<0.00001	++ (++)	
Amygdala	76 (74)	66	83	-3.3 (5.0)	-0.5 (2.0)	0.787	<0.00001	++ (++)	
EntA	69 (69)	65	72	-3.7 (4.1)	-1.2 (2.8)	0.747	<0.00001	++ (++)	
TMP	72 (71)	67	74	-4.3 (4.6)	-1.8 (2.8)	0.700	<0.00001	++ (++)	
LeftAmygdala	73 (70)	58	82	-3.6 (6.2)	-0.5 (2.5)	0.699	<0.00001	++ (++)	
SMG	64 (64)	61	67	-2.9 (3.5)	-1.1 (2.1)	0.674	<0.00001	++ (++)	

AD patients (N = 117, Positives^P) vs. Healthy Controls (N = 168, Negatives^N) (longitudinal analysis, bl→m24)									
structure	ACC (bACC)	SENS	SPEC	mean (SD) [%] ^P	mean (SD) [%] ^N	effect size (d)	p-value	sig. (corr.)	
RandomForest (all features)	89 (88)	81	94						
SVM (all features)	89 (88)	80	95						
Ventricles	83 (81)	70	92	14.7 (7.9)	5.8 (3.7)	1.531	<0.00001	++ (++)	
BrainTissue	80 (78)	71	86	-2.6 (1.3)	-1.1 (0.8)	1.353	<0.00001	++ (++)	
WhiteMatter	75 (74)	65	82	-1.8 (1.4)	-0.6 (0.8)	1.121	<0.00001	++ (++)	
CorticalGreyMatter	73 (73)	69	77	-3.5 (2.4)	-1.6 (1.7)	0.928	<0.00001	++ (++)	
Brain	68 (67)	65	70	-1.5 (1.0)	-0.8 (0.7)	0.847	<0.00001	++ (++)	
DeepGreyMatter	67 (66)	63	69	-2.3 (1.8)	-1.3 (1.3)	0.604	<0.00001	++ (++)	
Hippocampus	87 (85)	77	94	-10.2 (6.2)	-2.0 (2.4)	1.880	<0.00001	++ (++)	
InfLatVent	84 (83)	76	90	15.9 (8.8)	3.5 (4.5)	1.871	<0.00001	++ (++)	
RightInfLatVent	83 (82)	73	90	16.1 (9.7)	3.5 (4.9)	1.728	<0.00001	++ (++)	
LeftHippocampus	86 (84)	77	92	-9.9 (6.6)	-1.9 (2.7)	1.701	<0.00001	++ (++)	
LeftInfLatVent	80 (78)	69	88	15.5 (9.5)	3.5 (5.2)	1.652	<0.00001	++ (++)	
MTG	83 (81)	70	92	-6.8 (4.1)	-2.0 (1.6)	1.651	<0.00001	++ (++)	
RightHippocampus	84 (81)	69	94	-10.5 (7.6)	-2.0 (2.6)	1.617	<0.00001	++ (++)	
ITG	83 (81)	71	91	-5.3 (3.3)	-1.6 (1.4)	1.603	<0.00001	++ (++)	
FuG	85 (83)	72	93	-3.7 (2.7)	-0.7 (0.9)	1.588	<0.00001	++ (++)	
LeftLateralVentricle	82 (80)	71	89	15.6 (8.2)	6.5 (4.4)	1.470	<0.00001	++ (++)	
PP	80 (79)	73	85	-10.9 (6.1)	-3.9 (3.5)	1.466	<0.00001	++ (++)	
LateralVentricle	80 (78)	65	90	15.4 (8.5)	6.3 (4.0)	1.454	<0.00001	++ (++)	
TMP	84 (82)	75	90	-8.2 (5.1)	-2.8 (2.5)	1.430	<0.00001	++ (++)	
RightLateralVentricle	80 (77)	63	92	15.1 (9.3)	6.2 (4.0)	1.331	<0.00001	++ (++)	
PT	78 (77)	70	83	-10.0 (6.2)	-3.8 (3.5)	1.276	<0.00001	++ (++)	
EntA	80 (79)	72	86	-7.0 (5.5)	-2.2 (3.1)	1.151	<0.00001	++ (++)	
CerebralWhiteMatter	75 (74)	65	83	-1.9 (1.3)	-0.7 (0.9)	1.138	<0.00001	++ (++)	
Amygdala	81 (79)	70	89	-7.3 (6.7)	-1.7 (3.7)	1.076	<0.00001	++ (++)	
RightAmygdala	79 (77)	69	85	-7.6 (7.0)	-1.8 (4.2)	1.050	<0.00001	++ (++)	
SMG	71 (70)	61	78	-5.7 (4.6)	-2.2 (2.6)	0.968	<0.00001	++ (++)	

Table 5.6: Classification results in % (6-fold cross-validation, LDA 100 runs, RF/SVM 20 runs) for separating progressive from stable MCI based on volume change from baseline to m12 (top) or m24 (bottom). Individual structures are sorted by effect size. The 20 structures with highest effect size are listed explicitly. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses.

progressive MCI (N = 168, Positives^P) vs. stable MCI (N = 149, Negatives^N) (longitudinal analysis, bl→m12)									
structure	ACC (bACC)	SENS	SPEC	mean (SD) [%] ^P	mean (SD) [%] ^N	effect size (d)	p-value	sig. (corr.)	
RandomForest (all features)	74 (73)	77	70						
SVM (all features)	74 (74)	72	75						
Ventricles	68 (69)	63	74	6.8 (4.3)	3.4 (3.1)	0.890	<0.00001	++ (++)	
BrainTissue	63 (63)	64	63	-1.2 (1.0)	-0.6 (0.8)	0.652	<0.00001	++ (++)	
CorticalGreyMatter	64 (64)	67	61	-1.8 (1.7)	-0.8 (1.3)	0.608	<0.00001	++ (++)	
Brain	58 (58)	62	54	-0.8 (0.8)	-0.4 (0.6)	0.495	0.00002	++ (+)	
DeepGreyMatter	64 (64)	62	66	-1.3 (1.2)	-0.7 (1.3)	0.428	0.00017	++ (+)	
WhiteMatter	63 (63)	61	64	-0.6 (0.9)	-0.4 (0.7)	0.310	0.00615	+ (o)	
Hippocampus	67 (67)	59	76	-4.1 (3.2)	-1.7 (2.2)	0.867	<0.00001	++ (++)	
LateralVentricle	67 (68)	61	75	7.1 (4.6)	3.6 (3.4)	0.866	<0.00001	++ (++)	
InfLatVent	69 (69)	63	75	6.7 (5.6)	2.6 (3.8)	0.845	<0.00001	++ (++)	
LeftHippocampus	67 (68)	60	75	-4.2 (3.3)	-1.7 (2.5)	0.845	<0.00001	++ (++)	
LeftLateralVentricle	66 (67)	60	73	7.2 (4.8)	3.6 (3.5)	0.836	<0.00001	++ (++)	
RightLateralVentricle	67 (67)	62	72	7.0 (4.6)	3.6 (3.6)	0.828	<0.00001	++ (++)	
MTG	70 (70)	65	75	-3.1 (2.4)	-1.4 (1.8)	0.807	<0.00001	++ (++)	
ITG	67 (67)	64	70	-2.5 (2.0)	-1.1 (1.6)	0.793	<0.00001	++ (++)	
LeftInfLatVent	68 (69)	62	76	6.7 (6.2)	2.4 (4.3)	0.783	<0.00001	++ (++)	
EntA	67 (67)	67	68	-4.0 (3.7)	-1.4 (3.3)	0.749	<0.00001	++ (++)	
PT	68 (68)	63	73	-5.5 (4.5)	-2.6 (3.3)	0.727	<0.00001	++ (++)	
RightInfLatVent	64 (64)	58	71	6.7 (6.2)	2.7 (4.4)	0.726	<0.00001	++ (++)	
FuG	66 (66)	63	69	-1.5 (1.6)	-0.5 (1.1)	0.694	<0.00001	++ (++)	
PP	66 (67)	64	70	-5.8 (4.6)	-2.9 (3.9)	0.692	<0.00001	++ (++)	
RightHippocampus	65 (65)	57	73	-4.0 (3.9)	-1.7 (2.8)	0.677	<0.00001	++ (++)	
3rdVentricle	64 (64)	61	67	4.7 (3.8)	2.5 (3.7)	0.590	<0.00001	++ (++)	
Amygdala	65 (66)	58	73	-3.2 (4.1)	-1.2 (2.8)	0.577	<0.00001	++ (++)	
RightAmygdala	64 (64)	56	73	-3.7 (5.5)	-1.1 (3.3)	0.567	<0.00001	++ (++)	
TTG	61 (62)	61	62	-4.4 (4.2)	-2.2 (3.5)	0.558	<0.00001	++ (++)	
TMP	64 (64)	64	64	-3.9 (3.7)	-2.2 (3.0)	0.524	<0.00001	++ (++)	

progressive MCI (N = 140, Positives^P) vs. stable MCI (N = 107, Negatives^N) (longitudinal analysis, bl→m24)									
structure	ACC (bACC)	SENS	SPEC	mean (SD) [%] ^P	mean (SD) [%] ^N	effect size (d)	p-value	sig. (corr.)	
RandomForest (all features)	79 (78)	82	74						
SVM (all features)	76 (76)	76	76						
Ventricles	71 (72)	65	80	13.7 (7.4)	7.1 (5.6)	0.989	<0.00001	++ (++)	
BrainTissue	68 (69)	62	76	-2.3 (1.5)	-1.2 (1.0)	0.845	<0.00001	++ (++)	
DeepGreyMatter	66 (67)	64	69	-2.1 (1.4)	-1.2 (1.0)	0.714	<0.00001	++ (++)	
CorticalGreyMatter	65 (66)	61	71	-3.2 (2.5)	-1.7 (1.6)	0.707	<0.00001	++ (++)	
Brain	63 (63)	60	67	-1.5 (1.1)	-0.8 (0.7)	0.707	<0.00001	++ (++)	
WhiteMatter	64 (65)	55	74	-1.6 (1.3)	-0.8 (1.0)	0.661	<0.00001	++ (++)	
FuG	75 (76)	69	83	-3.1 (2.1)	-1.0 (1.4)	1.141	<0.00001	++ (++)	
Hippocampus	73 (74)	68	80	-8.9 (5.1)	-3.7 (3.8)	1.118	<0.00001	++ (++)	
LeftHippocampus	74 (74)	72	77	-9.3 (5.6)	-3.8 (4.2)	1.083	<0.00001	++ (++)	
EntA	71 (72)	66	77	-7.4 (4.7)	-2.8 (3.6)	1.070	<0.00001	++ (++)	
InfLatVent	70 (72)	62	81	14.6 (8.9)	6.2 (6.6)	1.053	<0.00001	++ (++)	
MTG	72 (74)	64	83	-6.1 (3.9)	-2.6 (2.6)	1.016	<0.00001	++ (++)	
LeftInfLatVent	68 (69)	60	79	14.7 (9.4)	6.4 (6.8)	0.987	<0.00001	++ (++)	
PT	68 (68)	62	75	-9.7 (5.6)	-4.7 (4.2)	0.971	<0.00001	++ (++)	
LateralVentricle	70 (71)	64	78	14.4 (7.9)	7.5 (6.0)	0.960	<0.00001	++ (++)	
ITG	72 (73)	66	80	-4.7 (2.9)	-2.1 (2.2)	0.958	<0.00001	++ (++)	
LeftLateralVentricle	69 (70)	62	77	14.6 (8.2)	7.6 (6.1)	0.956	<0.00001	++ (++)	
RightInfLatVent	66 (68)	60	76	14.3 (10.1)	5.9 (7.4)	0.923	<0.00001	++ (++)	
RightHippocampus	71 (72)	65	79	-8.6 (5.7)	-3.7 (4.6)	0.920	<0.00001	++ (++)	
RightLateralVentricle	69 (70)	61	79	14.1 (8.1)	7.5 (6.3)	0.907	<0.00001	++ (++)	
Amygdala	67 (68)	61	76	-6.4 (5.2)	-2.9 (3.5)	0.770	<0.00001	++ (++)	
TTG	65 (66)	62	70	-8.0 (5.3)	-4.4 (4.1)	0.752	<0.00001	++ (++)	
PHG	66 (68)	58	77	-3.5 (2.8)	-1.5 (2.4)	0.746	<0.00001	++ (++)	
3rdVentricle	64 (65)	61	68	9.1 (6.2)	4.7 (5.5)	0.740	<0.00001	++ (++)	
PP	68 (69)	65	72	-10.0 (5.7)	-5.9 (5.5)	0.728	<0.00001	++ (++)	
TMP	68 (69)	64	74	-6.9 (4.5)	-3.9 (3.6)	0.721	<0.00001	++ (++)	

the inferior lateral ventricles (bACC_{m12}: 69%, SENS: 63%, SPEC: 75%). Over 24 months, atrophy in the fusiform gyrus was most informative to classify MCI subjects (bACC_{m24}: 76%). Exploiting all available longitudinal features in a RF classifier further increased MCI classification accuracy to bACC_{m12}: 73%; bACC_{m24}: 78%.

Ventricular enlargement is more discriminative than a reduction in brain tissue for both AD vs. HC and pMCI vs. sMCI classification.

5.6 Discussion

In Section 4.2 MALP-EM4D was compared to other state-of-the-art approaches (Wolz et al., 2010b; Leung et al., 2010a) on a *common, identical dataset*. In this chapter a set of 1069 baseline, 802 m12 and 532 m24 follow-up images from the ADNI-1/-GO/-2 cohort was analysed. Given this large number of images the performed analysis should generalise to other datasets.

Table 5.7: Overview over selected articles that use features from T1w MR images from the ADNI cohort. Table adapted from Falahati et al. (2014). CTH: cortical thickness, ENR: elastic net regression, HV: hippocampus volume, LLE: locally linear embedding, MBL: manifold-based learning, MIL: multiple instance learning, OPLS: orthogonal partial least square to latent structure, SR: sparse regression, TBM: tensor-based morphometry.

Article	Dataset (field strength)	Features	Classifier	AD vs. HC				pMCI vs. sMCI			
				bACC	SENS	SPEC	N _{AD} /N _{HC}	bACC	SENS	SPEC	N _{pMCI} /N _{sMCI}
MALP-EM	ADNI1/Go/2 (1.5/3T)	ROI volumes	RF	86	83	90	322/404	68	72	64	177/166
MALP-EM	ADNI1/Go/2 (1.5/3T)	ROI volumes	SVM	89	86	92	322/404	67	70	64	177/166
Coupé et al. (2012b)	ADNI1 (1.5T)	ROI volumes and grading features	LDA	90.5	87	94	198/231	73.5	73	74	167/238
Chincarini et al. (2011)	ADNI1 (1.5T)	intensity and textural features from 9 ROIs	SVM	91.5	89	94	144/189	68.5	72	65	136/166
Guerrero et al. (2014)	ADNI1 (1.5T)	learned ROIs via SR + MBL	SVM	85.5	86	85	106/175	71	75	67	116/114
Liu et al. (2013)	ADNI1 (1.5T)	ROI volumes, CTH	ENR + LLE	89.5	86	93	86/137	68	80	56	97/93
Tong et al. (2014)	ADNI1 (1.5T)	local intensity patches	MIL	89	85	93	198/231	70	67	73	167/238
Wee et al. (2013)	ADNI1 (1.5T)	correlative and ROI-based morphological features	SVM	92.4	90.4	94.3	198/200	74.0	63.5	84.4	89/111
Westman et al. (2013)	ADNI1 (1.5T)	ROI volumes, CTH, curvature and folding features	OPLS	91.5	90	93	187/225	71.2	75.9	66.5	87/200
Wolz et al. (2011)	ADNI1 (1.5T)	HV, CTH, TBM, MBL	LDA	89	93	85	198/231	68	67	69	167/238

A selection of articles is summarised in Table 5.7 to put the calculated classification results into the context of those reported in the literature. When comparing results to other studies it must be noted that these usually use other ADNI subsets. A further potential confounding factor is the definition of the sMCI and pMCI disease groups. The stratification of MCI subjects into sMCI and pMCI respectively was defined in Section 5.1. However, this definition is not identical throughout published studies.

The presented cross-sectional classification results (*cf.* Section 5.5.1.2) are very similar to those presented in Wolz et al. (2011), which are based on the ADNI-1 cohort. In this work the authors

classified AD vs. HC (bACC: 89%, SENS: 93%, SPEC: 85%) and pMCI vs. sMCI (bACC: 68%, SENS: 67%, SPEC: 69%) based on a multitude of features including more abstract features from TBM and manifold-learning based methods. Other studies report even higher classification results of up to bACC: 92.4% for AD vs. HC and bACC: 74.0% for pMCI vs. sMCI classification (Wee et al., 2013). In contrast to the presented analysis, the articles shown in Table 5.7 analysed the 1.5T images of ADNI-1 only. Also usually more complex features such as cortical features (Wolz et al., 2011; Westman et al., 2013; Liu et al., 2013; Wee et al., 2013), textural features (Chincarini et al., 2011), manifold-based features (Wolz et al., 2011; Liu et al., 2013) or grading based features (Coupé et al., 2012b) are employed. In summary, the results presented in this chapter are comparable to the state-of-the-art. This is encouraging as a large and heterogeneous dataset was analysed acquired at both 1.5T and 3T using structural volumes, which are trivial to interpret, only.

Furthermore, structures that were found to be most discriminative agree well with those highlighted in Fennema-Notestine et al. (2009). Examples are the hippocampus, amygdala, entorhinal area and regions within the temporal gyrus in general.

Using MALP-EM4D and hippocampal atrophy a $bACC_{m12}$ of 78% and $bACC_{m24}$ of 85% was calculated for classifying AD vs. HC. These results are similar to those obtained on a different ADNI subset with a method dedicated to hippocampal atrophy measurement (Wolz et al., 2010b): $bACC_{m12}$ of 82% (SENS: 81%, SPEC: 83%); $bACC_{m24}$ of 86% (SENS: 85%, SPEC: 87%).

The results for AD vs. HC classification using all longitudinal features are similar to those using all structural volumes at baseline. However, classification accuracies substantially above 90% were not expected due to potential bias in the study data caused by for example variations in scanner type, field strength or clinical centres. This is in contrast to the pMCI vs. sMCI classification. Using longitudinal information available at the month 24 follow-up visit increased classification accuracy for the MCI groups from 68% at baseline to 78% using a RF classifier. This confirms the discriminative value of increased structural atrophy in MCI patients that progress to AD. In contrast to patients with AD, atrophy in MCI patients has not yet manifested itself in substantially reduced structural volumes at baseline.

In a large meta-analysis by Barnes et al. (2009b) the authors conclude that the annualised hippocampus atrophy of healthy elderly people is -1.4% and for patients with AD -4.6%. The rate of hippocampal volume change that was calculated in this chapter is comparable with -1.1% (± 1.7) for healthy controls and -4.8% (± 3.7) for AD subjects. For brain tissue a mean reduction of -0.5%

(± 0.8) and -1.3% (± 1.1) from baseline to m12 follow-up was observed for HC and AD groups. These results are in line with previously published annual brain atrophy rates of around -0.6% for controls and -1.5% for AD patients respectively (Leung et al., 2010b). These comparisons suggest that MALP-EM4D yields informative and realistic atrophy measurements on individual structures while providing a comprehensive overview of structural change throughout the whole brain.

The presented sample sizes and atrophy rates are comparable to the results calculated on the ADNI-1 subset of 309 subjects in Section 4.2. This is encouraging as it indicates that the results from the smaller subset generalise quite well and the method is rather robust towards variations in field strengths. Nevertheless, as briefly discussed in Section 5.4, significant differences were found between the healthy subjects scanned at 1.5T and 3T respectively. Including field strength into the regression model did not consistently improve results and was not further investigated in this study. The influence of the field strength on both the segmentation of individual images and the quantification of atrophy needs to be further explored. Another field for future work is to determine the ideal combination strategy for the individually measured atrophy rates to maximise their discriminative potential.

For the four structures hippocampus, amygdala, inferior lateral ventricle and lateral ventricle their counterparts in the left/right brain hemisphere were analysed separately. The results suggest that structural change in the left hippocampus is slightly more discriminative than change in the right hippocampus. For the other investigated structures there is no consistent trend. Overall features of corresponding structures in both brain hemispheres perform similar for all investigated structures and combining them seems, in summary, beneficial.

The use of either SVM or RF classifiers provided substantial improvements over the results obtained using individual features only. Overall both approaches performed similar with slight advantages for either of them in individual experiments.

5.7 Conclusion

Using the ADNI cohort the presented results confirm that MALP-EM (*cf.* Chapter 3) has the potential to support the analysis and classification of patient groups on a per structure basis. Calculated accuracies of up to 89% for AD vs. HC classification at baseline, using multiple structures as features, are in line with the state-of-the-art. The conducted experiments confirmed a benefit of

correcting for nuisance factors such as age, gender and brain size. Further, m12 and m24 atrophy rates were calculated for those subjects with respective imaging information available. Using MALP-EM4D (*cf.* Chapter 4) atrophy rates and corresponding sample sizes were determined for a multitude of brain structures simultaneously. A clear difference was observed between subjects converting from the MCI stage to AD and those that do not. The conducted experiments confirm the methods' potential to analyse the whole brain rather than individual structures only. A particular strength is that the methods allow both the accurate analysis of single images (MALP-EM) as well as image series (MALP-EM4D). Thus the presented methodology has strong potential to support both cross-sectional and longitudinal studies that include MR imaging of the brain.

6 Automated brain morphometry in patients with traumatic brain injury

Abstract: *In the previous chapter MR brain images acquired from a large cohort of subjects with and without dementia were analysed. On this data it was confirmed that the methodology developed in this thesis has the potential to support automated brain morphometry by providing informative, disease-related measures based on either single time point MR images (structural volumes) or longitudinal MR images (structural atrophy). The main characteristic of patients with AD is ongoing brain atrophy and atrophy within the structures of the MTL in particular. In contrast to AD, brain injury in TBI is caused by a sudden event and can be very heterogeneous in its manifestation. Typical consequences of a head injury such as brain midline shift or the presence of contusions are not characteristic for AD. In this chapter, an analysis is presented of a cohort of T1w MR images prospectively acquired from patients who sustained mild to severe TBI. The goal is to investigate the applicability of the methodology developed in Chapter 3 (MALP-EM) and Chapter 4 (MALP-EM4D) to extract meaningful ROIs from MR brain images of patients with substantially altered brain structure. As segmentation quality can not be quantified directly in the absence of reference segmentations, the potential value of volume and asymmetry measures as a predictor of outcome is investigated instead. These features are quantified from brain MRI at the acute/subacute stage¹ of the injury (median: 19 days after injury), to predict the disability outcome of 67 patients at the chronic disease stage (median: 229 days after injury). Further, differences in structural atrophy are investigated between groups of varying disease outcome severity. Patients with unfavourable outcome prognosis show increased white matter atrophy.*

This chapter is organised as follows: First, an overview is given over the study data in Section

¹Following the definition used in Newcombe et al. (2015), the images in this study were taken either in the acute or subacute phase. For readability both stages will be referred to as 'acute'.

6.1. The applied methods are briefly described in Section 6.2. Section 6.3 summarises the extracted imaging features and describes the classification approach. In Section 6.4.1, a cross-sectional analysis at the acute stage of the injury explores whether individual structural volumes have potential to predict injury outcome. Further, a longitudinal analysis is performed in Section 6.4.2 and structural atrophy rates are calculated between images acquired in the acute and chronic disease stage. Group differences are investigated between patient groups of distinct outcome categories. In Section 6.6 the findings are discussed and segmentation results are illustrated that were obtained on brain MR images with disease related changes such as for example subdural haematomas, substantial structural deformation or atrophy. This chapter concludes with a brief summary in Section 6.6.

6.1 Materials

The imaging data was acquired at Turku University Hospital, Finland in the course of the TBI-care project (<http://www.tbicare.eu>, last accessed 29 June 2015). For the T1w MR images an MPRAGE sequence was acquired on a Siemens Verio 3T system with the following parameters: TR 2300 ms, TE 2.98 ms, TI 900 ms, flip angle 9° , matrix size $256 \times 249 \times 176$ and an isotropic voxel size of $1.0\text{mm} \times 1.0\text{mm} \times 1.0\text{mm}$, sagittal slices, using Prescan Normalizer, 2D distortion correction and a standard 12 channel head coil.

Over the course of the project a total of 141 subjects with mild to severe TBI have had MR images taken both at the acute stage of the injury (baseline) and in the chronic phase (follow-up) of the disease. In total 120 subjects were processed for which both baseline and follow-up images were available when the analysis was started in December 2013. After visual review, six subjects were excluded for the following reasons:

- TBI031, TBI073: Images seem to be from the same subject, but labelled with different IDs
- TBI116: The processed baseline image was acquired very late (3 months after the injury)
- TBI152: The baseline and follow-up images are identical
- TBI157: Too short follow-up interval (≈ 1 month) and poor image quality
- TBI184: Poor image quality, brain extraction not sufficiently accurate

In the following the remaining 114 patients are considered, which are summarised in Table 6.1.

Table 6.1: Overview of **all processed MR images** with patient gender, patient age, scan time relative to injury, GCS, MCS and injury severity.

GOSe	low disability 7 & 8	moderate disability 5 & 6	severe disability 3 & 4
# of subjects	69	32	13
gender (# male / # female)	47/22	21/11	4/9
years of age (median [min; max])	40 [18; 82]	51 [19; 83]	74 [33; 86]
days since injury, acute scan (median [min; max])	14 [1; 50]	22 [1; 51]	22 [4; 51]
days since injury, chronic scan (median [min; max])	230 [151; 399]	228 [177; 429]	251 [180; 422]
GCS (median [min; max])	15 [3; 15]	15 [3; 15]	14 [2; 15]
Marshall score (median [min; max])	1 [1; 5]	2 [1; 5]	5 [2; 5]
injury severity [†] (median [min; max])	2 [1; 4]	2.5 [2; 5]	3 [2; 5]

†: 1: very mild, 2: mild, 3: moderate, 4: severe, 5: very severe

6.1.1 Clinical information

In addition to MR imaging data the clinical variables age, gender, GCS, MCS, GOSe and injury severity were available. After the head injury the GCS is potentially assessed several times, *e.g.* at the injury site, at hospital admission, in the intensive care unit. The GCS might not have been recorded for each patient at each time point. Thus a pragmatic approach was followed and the GCS score chosen that was recorded first. This is usually either at the injury site or at admission to the hospital. The MCS groups 5 and 6 were pooled together (*cf.* Table C.1), which means that the MCS scores did not distinguish between evacuated and non-evacuated mass lesions. Further the GOSe groups 3 & 4, 5 & 6 and 7 & 8 are summarised into three patient groups with severe, moderate and low disability outcome respectively. This was necessary to obtain reasonable group sizes and corresponds to using the GOS five point scale instead of the eight point GOSe (*cf.* Table D.1). The GOSe score was assessed on the day when the follow-up MR image was acquired. Injury severity was classified based on either the GCS or the duration of PTA depending on which one gave the higher severity (Schoenberger et al., 2009; Arlinghaus et al., 2011). A widely used approach for classifying the duration of PTA was employed where less than 24 hours is mild, 1 to 7 days moderate, 1 to 4 weeks severe and more than four weeks very severe (Schoenberger et al., 2009; Arlinghaus et al., 2011).

Next to the imaging information the dataset includes a multitude of clinical data such as metabolomics. However, this data will not be further exploited in the following analyses.

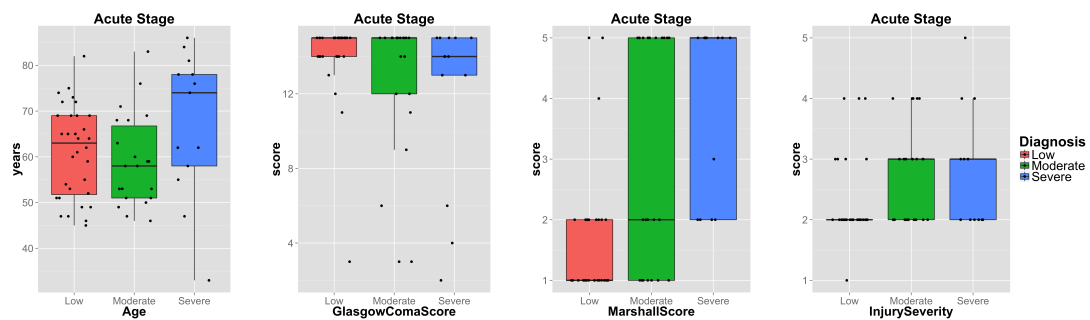


Figure 6.1: Boxplots of age, GCS, MCS and injury severity of patients with low, moderate or severe disability outcome. Shown are only subjects of the age-matched dataset.

6.1.2 Age-matching of patient groups

The study groups which are summarised in Table 6.1 show a significant mismatch in age ($p < 0.001$ for all groups with respect to the severe outcome group). It is important to account for these substantial differences when studying changes caused by the disease to minimise age-related effects. Due to the limited size of the study cohort this is a challenging endeavour. The group with severe disability outcome has the fewest samples and subjects are significantly older in age. In order to not further reduce the size of this group, all subjects were removed from the low and moderate disability outcome groups that are younger than 45 years of age. The age difference between the groups was no longer significant ($p > 0.05$) after this correction. This approach improved the age-match between study groups, however, reduced the number of study subjects from 114 to 67. An overview over the remaining subjects that will be studied in the following is provided in Table 6.2. The distribution of age, GCS and MCS is illustrated with respect to the three outcome groups in Figure 6.1.

Table 6.2: Overview of the **MR images used for the analysis** with patient gender, patient age, scan time relative to injury, GCS, MCS and injury severity.

GOSe	low disability 7 & 8	moderate disability 5 & 6	severe disability 3 & 4
# of subjects	32	22	13
gender (# male / # female)	21/11	14/8	4/9
years of age (median [min; max])	63 [45; 82]	58 [46; 83]	74 [33; 86]
days since injury, acute scan (median [min; max])	15 [1; 50]	23 [2; 51]	22 [4; 51]
days since injury, chronic scan (median [min; max])	225 [151; 276]	227 [177; 429]	251 [180; 422]
GCS (median [min; max])	15 [3; 15]	15 [3; 15]	14 [2; 15]
Marshall score (median [min; max])	1 [1; 5]	2 [1; 5]	5 [2; 5]
injury severity [†] (median [min; max])	2 [1; 4]	3 [2; 4]	3 [2; 5]

[†]: 1: very mild, 2: mild, 3: moderate, 4: severe, 5: very severe

6.2 Methods

All available images were preprocessed by first correcting for intensity inhomogeneities with the N4 bias correction algorithm (Tustison et al., 2010). Images were subsequently brain extracted using pinfram (Heckemann et al., 2015), an iterative, atlas-based method that was developed with particular focus on robustness. Each image was then segmented individually using MALP-EM as described in Chapter 3. Longitudinal changes were measured with MALP-EM4D as described in Chapter 4. As brain atlases the 30 manually annotated NMM brain atlases were employed (*cf.* Section 2.2.5). In the context of TBI, gross structural changes can be expected between both imaging time points. Thus the weighted differential bias field correction procedure was used which was introduced in Section 4.2.2.4. This is in contrast to the processing of the ADNI dataset in Chapter 5. Otherwise, the applied methods are, except for the additional N4 bias correction, identical to the one described and applied in Chapter 5.

All brain masks and segmentation results were visually reviewed to ensure reasonable accuracy, allowing for pathology. A typical segmentation result is visualised in Figure 6.2.

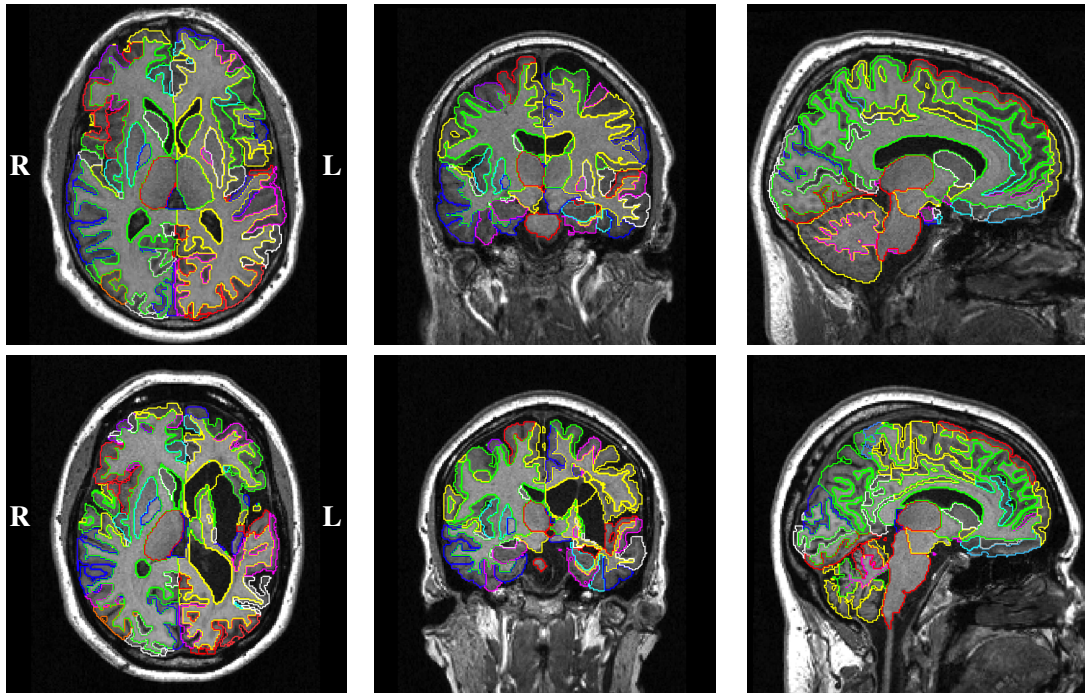


Figure 6.2: Example cross-sectional segmentation results of images at the acute stage of a TBI. Top: TBI010_2.8, male, 21 years of age, mild TBI, GCS: 15, GOSe: 8, MCS: 1, fall accident, image acquired 2 days after injury. Bottom: TBI038, female, 47 years of age, mild TBI, GCS: 15, GOSe: 4, MCS: 5, fall, image acquired 4 days after injury with clear sequelae of intra-cerebral haematoma.

6.3 Features and classification

Group differences were investigated between GOS groups with a focus on the comparison of patients with low and severe disability outcome. Classification experiments were performed to quantify the accuracy of predicting the GOS outcome category when using imaging information available at the acute disease stage.

All 67 subjects were analysed cross-sectionally at the *acute* stage of the TBI and longitudinally employing the follow-up image acquired in the *chronic* phase of the disease. As features all available non-cortical structural volumes were used, however, corresponding volumes in the left and right brain hemisphere were merged ($28/2+7=21$ non-cortical features). Note that the seven non-cortical structures 3rd ventricles, 4th ventricle, brain stem, CSF, cerebellar vermal lobule I-V, cerebellar vermal lobule VI-VII and cerebellar vermal lobule VIII-X have no symmetric counterpart. Individual structures were summarised into ventricles, cortical GM, deep GM, WM, brain tissue and brain volume (brain tissue including ventricles/CSF) (6 features). The optic chiasm was excluded from the analysis due to its very small size. Due to the limited number of study subjects and the heterogeneity of the injury, cortical structures were investigated as surrogate structure (cortical GM) rather than individually.

In the *cross-sectional* analysis at the acute stage structural asymmetry was quantified as the AAI (*cf.* Equation 3.17 in Chapter 3). The AAI was calculated for the 14 non-cortical structures appearing in both brain hemispheres and the six surrogate structures. Additionally the AAI of all individual non-cortical, cortical and all brain structures were added up. Note that the sum of for example all cortical AAI is different to the AAI of the cortical GM surrogate structure. The segmentations at the acute stage were calculated with MALP-EM and not MALP-EM4D as the latter exploits information of later scanning time points, which is not available at baseline. Further, the five non-imaging features age, gender, GCS, MCS and injury severity were considered.

For the *longitudinal* analysis structural volumes of all 67 subjects were extracted based on the respective MALP-EM4D segmentations. Atrophy rates were calculated as described in Section 4.2.3.2. The volume change was measured for the six surrogate structures and for the individual 21 non-cortical structures. This yields, considering age and gender, 29 features for the longitudinal analysis. Even though the analysis was limited to changes in structural volume, other longitudinal alterations such as changes in brain symmetry could be investigated.

For classification 100 runs of a 6-fold CV were chosen using LDA for individual features and SVM or RF classifiers when combining multiple features. The same MATLAB classification framework was used as in Chapter 5 that relies on `classify` (LDA), `TreeBagger` (RF, 100 trees) and `libSVM` (SVM, (Chang and Lin, 2011)). Note that in contrast to the analysis on the ADNI dataset 100 runs were chosen for the multi-feature classifiers. The motivation for this is two-fold: a) due to the smaller group sizes this is computationally more feasible, b) variations in classification results were observed when using only 20 CV repetitions. This might be caused by the small group sizes. As in Chapter 5 the features were normalised individually based on the respective training set to the range 0 to 1 for the SVM classification. Both the LDA classifier, which was used for single-feature classification only, and the RF classifier do not require feature normalisation. No explicit correction for age, gender or head size was applied. This is further discussed in Section 6.5.

Significance levels were quantified as p-values of two-sided, unpaired Student's t-tests. Further, effect sizes were calculated as Cohen's d by dividing the differences of the sample means (absolute value) by their pooled standard deviation (Cohen, 1988; Fennema-Notestine et al., 2009; Wang et al., 2013b).

6.4 Results

6.4.1 Cross-sectional analysis

Individual brain ROIs were extracted from the *acute* T1w MR images. From these ROIs features (volume, asymmetry) were derived and their potential investigated to discriminate TBI patients according to their *outcome* severity. Example segmentation results of a TBI patient with low disability outcome and a patient with severe disability outcome are shown in Figure 6.2.

The subject shown in the top row sustained a mild TBI with a favourable disease outcome (GOSe: 8) and had no visible intracranial pathological changes on CT (MCS: 1). The subject shown in the bottom row sustained a mild TBI with a more unfavourable disease outcome (GOSe: 4) and substantial pathological changes (MCS: 5). Before the actual TBI event the patient suffered a spontaneous intra-cerebral haematoma due to an untreated hypertension.

The distribution of structural volumes for nine selected ROIs with respect to the three considered outcome categories low, moderate and severe disability is shown in Figure 6.3. Several subcortical structures were identified that are of particularly small size in subjects with severe disability

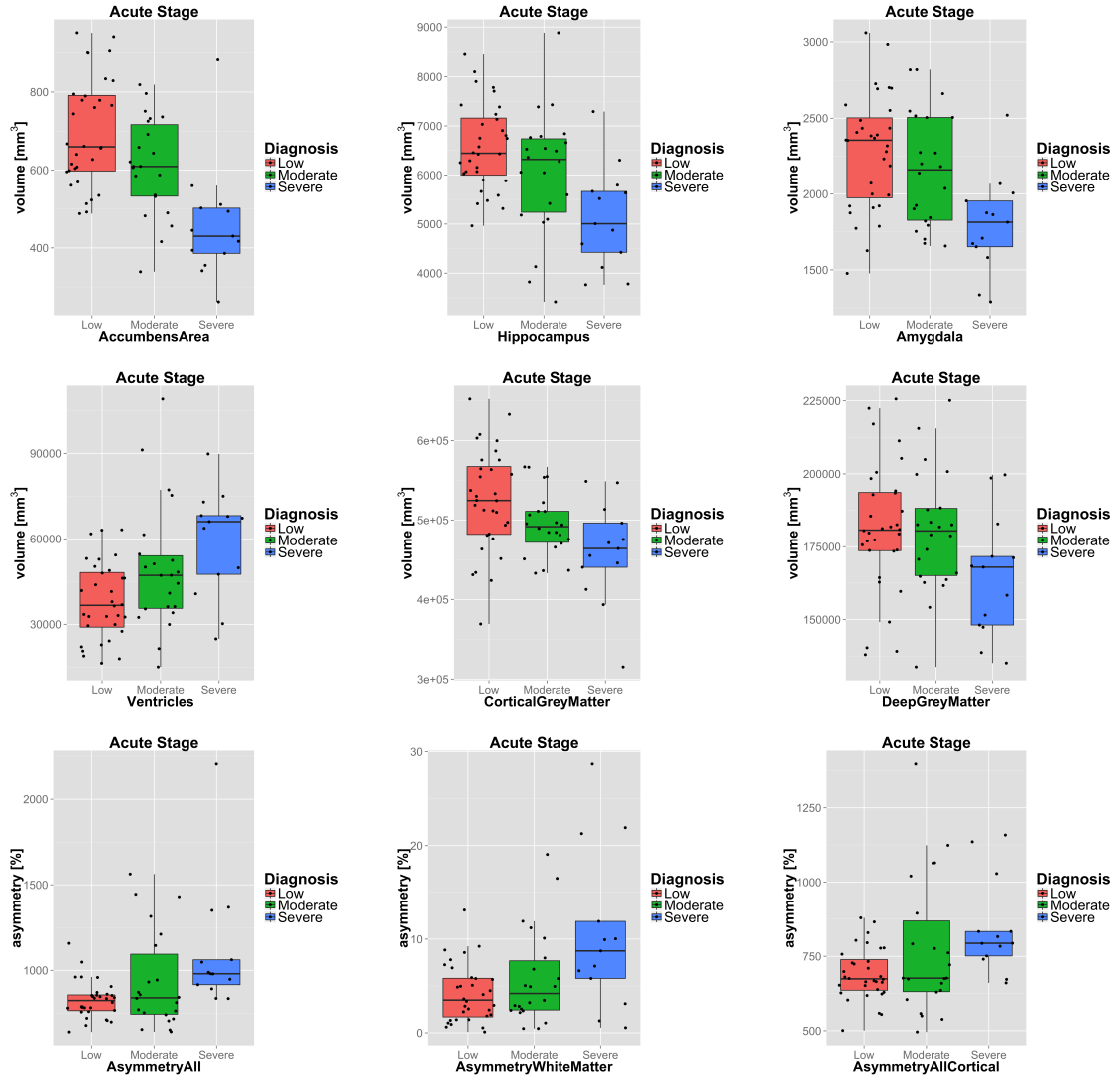


Figure 6.3: Boxplots of selected structural volumes (first row), surrogate structures (second row) and asymmetry indices (third row) with respect to the investigated disease outcome groups. Features selected based on their performance in classifying severe disability vs. low disability outcome (*cf.* Table 6.3).

Table 6.3: Classification results in % (6-fold cross-validation, LDA/RF/SVM 100 runs) obtained separating TBI patients with a severe disability from patients with low disability outcome based on structural volumes and asymmetry at the acute stage of the injury. Individual structures are sorted by effect size. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses.

Severe disability (N = 13, Positives ^P) vs. Low disability (N = 32, Negatives ^N) (acute stage)									
structure	ACC (bACC)	SENS	SPEC	mean (SD) [mm ³] ^P	mean (SD) [mm ³] ^N	effect size (d)	p-value	sig. (corr.)	
RandomForest (all features)	85 (79)	65	93						
SVM (all features)	87 (84)	76	92						
RandomForest (imaging only)	85 (79)	65	93						
SVM (imaging only)	83 (77)	62	92						
RandomForest (non-imaging only)	82 (77)	67	88						
SVM (non-imaging only)	82 (77)	62	91						
RandomForest (age and gender only)	69 (60)	39	81						
SVM (age and gender only)	70 (50)	1	99						
MarshallScore	84 (80)	69	91	3.9 (1.4)	1.6 (1.1)	1.907	<0.00001	++ (++)	
InjurySeverity	71 (66)	54	78	2.8 (1.0)	2.3 (0.7)	0.724	0.03308	+	(+)
Gender (female = 0, male = 1)	67 (67)	69	66	0.3 (0.5)	0.7 (0.5)	0.723	0.03331	+	(+)
GlasgowComaScale	72 (59)	29	89	11.9 (4.6)	14.1 (2.2)	0.700	0.03912	+	(+)
Age	55 (54)	54	55	67.2 (15.9)	61.1 (10.1)	0.514	0.12518	o	(o)
Ventricles	72 (71)	68	74	58766.8 (18808.4)	38131.9 (13556.0)	1.357	0.00017	++ (++)	
CorticalGreyMatter	70 (70)	70	71	459934.2 (63395.6)	524318.2 (64687.5)	1.001	0.00398	+	(+)
DeepGreyMatter	74 (74)	74	75	164564.8 (20726.5)	181380.2 (22209.8)	0.771	0.02373	+	(+)
BrainTissue	60 (60)	60	60	1130166.1 (102464.8)	1198466.6 (135087.2)	0.539	0.10886	o	(o)
Brain	58 (59)	62	57	1191022.6 (110493.8)	1238799.0 (138060.0)	0.365	0.27348	o	(o)
WhiteMatter	50 (46)	37	55	505667.1 (82222.8)	492768.2 (71084.9)	0.173	0.60062	o	(o)
AccumbensArea	82 (85)	91	79	460.0 (150.5)	696.3 (138.7)	1.664	<0.00001	++ (++)	
Hippocampus	80 (81)	83	79	5136.2 (1030.6)	6572.2 (873.5)	1.561	0.00002	++ (++)	
Amygdala	73 (76)	83	68	1795.5 (321.1)	2280.9 (380.7)	1.329	0.00022	++ (++)	
LateralVentricle	75 (74)	73	76	51162.3 (17123.4)	32704.8 (12459.3)	1.326	0.00022	++ (++)	
InfLatVent	81 (75)	62	89	3131.4 (1360.4)	1913.3 (676.5)	1.324	0.00023	++ (++)	
ThalamusProper	73 (74)	76	72	11584.6 (1556.0)	13832.8 (1845.3)	1.271	0.00037	++ (++)	
BasalForebrain	74 (75)	77	72	548.7 (232.3)	840.4 (233.9)	1.250	0.00045	++ (++)	
CerebellarVermalLobulesI-V	75 (74)	72	76	3471.9 (535.9)	4242.3 (699.1)	1.171	0.00092	++ (++)	
3rdVentricle	71 (71)	69	72	2288.6 (814.2)	1552.5 (617.3)	1.086	0.00194	+	(+)
Putamen	76 (76)	77	75	6066.7 (2492.2)	7849.7 (1338.7)	1.025	0.00326	+	(+)
BrainStem	69 (69)	71	68	17417.2 (2275.5)	19651.1 (2739.4)	0.853	0.01291	+	(+)
CerebellumWhiteMatter	73 (72)	69	74	36782.2 (7917.1)	30605.9 (8306.2)	0.753	0.02697	+	(+)
VentralDC	56 (55)	54	56	8247.1 (1062.4)	8961.7 (1114.9)	0.649	0.05477	o	(o)
CerebellarVermalLobulesVIII-X	57 (62)	74	51	2659.5 (304.5)	2902.0 (420.4)	0.619	0.06650	o	(o)
CerebellumExterior	64 (64)	63	64	94910.8 (12670.3)	101795.8 (13850.8)	0.509	0.12919	o	(o)
4thVentricle	60 (55)	45	66	2184.5 (673.2)	1961.2 (526.9)	0.391	0.24153	o	(o)
Caudate	53 (47)	35	60	7388.5 (3739.3)	6613.9 (1066.8)	0.356	0.28451	o	(o)
Pallidum	52 (51)	48	54	2801.2 (796.4)	2990.8 (504.4)	0.316	0.34230	o	(o)
CerebellarVermalLobulesVI-VII	54 (54)	54	54	2076.6 (384.4)	2150.0 (299.1)	0.226	0.49648	o	(o)
CSF	47 (45)	42	48	2025.5 (719.9)	2140.6 (538.2)	0.194	0.55923	o	(o)
CerebralWhiteMatter	43 (38)	28	49	468884.8 (76690.5)	462162.3 (67838.9)	0.095	0.77302	o	(o)
AsymmetryAllCortical [†]	76 (74)	69	79	846.2 (160.7)	690.8 (88.0)	1.374	0.00014	++ (++)	
AsymmetryAll [†]	81 (73)	54	91	1108.3 (369.6)	826.6 (107.9)	1.306	0.00027	++ (++)	
AsymmetryWhiteMatter	75 (69)	55	82	10.5 (8.5)	4.2 (3.1)	1.220	0.00059	++ (++)	
AsymmetryCerebralWhiteMatter	76 (68)	48	88	10.2 (9.5)	3.8 (2.8)	1.150	0.00110	+	(+)
AsymmetryAmygdala	82 (78)	68	87	16.9 (12.5)	7.7 (5.9)	1.119	0.00145	+	(+)
AsymmetryBrain	75 (72)	64	79	4.0 (3.3)	1.6 (1.5)	1.086	0.00194	+	(+)
AsymmetryBrainTissue	81 (73)	54	92	5.2 (5.2)	1.9 (1.7)	1.068	0.00226	+	(+)
AsymmetryCorticalGreyMatter	79 (72)	55	89	4.9 (4.6)	2.0 (2.3)	0.938	0.00663	+	(+)
AsymmetryAllNonCortical [†]	79 (71)	50	91	262.0 (245.1)	135.8 (46.7)	0.932	0.00695	+	(+)
AsymmetryCerebellumWhiteMatter	69 (69)	69	69	19.2 (13.9)	10.5 (7.8)	0.887	0.00997	+	(+)
AsymmetryPutamen	75 (63)	34	92	24.2 (44.3)	4.2 (4.8)	0.844	0.01384	+	(+)
AsymmetryAccumbensArea	66 (62)	53	71	23.4 (18.8)	12.7 (10.2)	0.805	0.01850	+	(+)
AsymmetryCaudate	71 (63)	46	81	17.3 (23.1)	7.2 (6.0)	0.759	0.02585	+	(+)
AsymmetryDeepGreyMatter	72 (66)	52	81	3.9 (4.2)	2.0 (1.9)	0.706	0.03743	+	(+)
AsymmetryThalamusProper	75 (60)	27	94	14.2 (35.4)	2.2 (2.0)	0.642	0.05762	o	(o)
AsymmetryVentricles	68 (59)	39	79	27.8 (32.4)	15.4 (13.5)	0.600	0.07507	o	(o)
AsymmetryHippocampus	63 (57)	43	71	16.2 (16.2)	9.9 (7.3)	0.595	0.07721	o	(o)
AsymmetryPallidum	68 (56)	28	84	15.9 (32.6)	5.5 (5.4)	0.580	0.08469	o	(o)
AsymmetryLateralVentricle	66 (57)	38	77	31.5 (35.6)	18.2 (15.6)	0.574	0.08797	o	(o)
AsymmetryVentralDC	65 (56)	36	76	9.5 (17.3)	4.2 (2.7)	0.556	0.09791	o	(o)
AsymmetryBasalForebrain	63 (55)	37	74	37.3 (40.4)	24.4 (17.1)	0.498	0.13745	o	(o)
AsymmetryCerebellumExterior	68 (64)	54	74	4.8 (3.1)	3.5 (3.7)	0.373	0.26322	o	(o)
AsymmetryInfLatVent	40 (37)	29	45	21.5 (20.5)	21.6 (17.7)	0.008	0.98119	o	(o)

[†]: Sum of the AAls of the individual structures.

outcome as compared to those patients with a low disability outcome (*cf.* Table 6.3). The four structures with the largest effect size are the accumbens (Cohen's $d = 1.66$), hippocampus ($d = 1.56$),

amygdala ($d = 1.33$) and the thalamus ($d = 1.27$). These differences are significant after Bonferroni correction for multiple comparisons. Overall, a larger ventricular volume ($d = 1.36$) and lower cortical GM volume ($d = 1.0$) was observed among patients with unfavourable outcome as compared to those with a good outcome prognosis. Asymmetry throughout the whole brain and in particular within the cortex and WM was significantly higher in patients with severe disability outcome than in the low disability outcome group. The group with moderate outcome disability had overall a wider spread (variance) of the measured features and was thus less well separated from both the low and severe outcome groups.

Further it was investigated whether the disability outcome of a patient can be predicted using only features that are available at the acute stage of the injury. The classification results for severe vs. moderate and moderate vs. low disability outcome classification can be found in Appendix F in Table F.1. In the following the focus will be on the separation of low vs. severe disability outcome groups. Results of this experiment are summarised in Table 6.3.

The volume of the accumbens provided the best group separation with the highest bACC of 85% (SENS: 91%, SPEC: 79%). Next to the accumbens, several structures were predictive for the disease outcome including hippocampus (bACC: 81%), amygdala (bACC: 76%) and thalamus (bACC: 75%).

Available features were combined into multi-feature classifiers. Using a SVM, imaging information only yielded similar results (bACC: 77 %) to those obtained with non-imaging information only (bACC: 77 %). Clinical variables and measurements from imaging contained complementary information. Their combination increased classification accuracy to bACC: 84 % (SENS: 76%, SPEC: 92%).

6.4.2 Longitudinal analysis

Further a longitudinal analysis was performed to investigate the volume change of individual ROIs between the acute stage of the injury and the follow-up visit in the chronic phase of the disease. Example segmentation results obtained on two subjects with low and moderate disability outcome are shown in Figure 6.4.

The correlation of structural atrophy rates with patient groups of distinct disease outcome was investigated. Specifically, classification accuracies, effect sizes and p-values were calculated to quantify group separation. The distribution of the volume change of six selected structures is shown

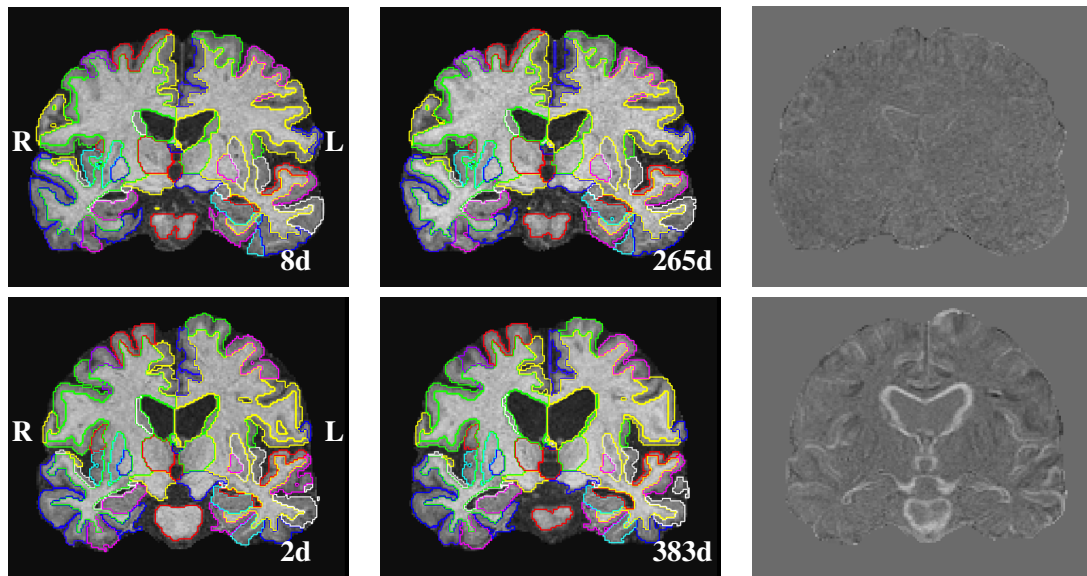


Figure 6.4: Example longitudinal segmentation results of images at the acute and chronic stage of a TBI. Top: TBI061, male, 69 years of age, GCS: 14, GOSe: 8, MCS: 1, fall accident, acute/chronic image acquired 8/265 days after injury. Bottom: TBI142, male, 51 years of age, GCS: 3, GOSe: 6, MCS: 5, transport accident, acute/chronic image acquired 2/383 days after injury, diffuse axonal injury. The difference image of subject TBI142 illustrates the clear ventricular enlargement (measured: 47%), hippocampal atrophy (-6.2%) and reduction of brain stem volume (-13%).

in Figure 6.5 for the three considered outcome groups. Quantitative results for moderate vs. low and severe vs. moderate outcome disability can be found in Appendix F in Table F.2. In the following the focus is on the comparison of severe and low disability outcome for which quantitative results are summarised in Table 6.4.

Substantial ventricular expansion (11.6%) was observed in patients with poor disease outcome. In contrast brain tissue and in particular cerebral WM (-3.1%), brain stem (-3.1%) and thalamus (-4.2%) showed increased atrophy.

The clearest group separation was calculated for atrophy of the cerebral WM ($d = 1.29$, $p \approx 10^{-3}$). Volumetric change of individual ROIs, such as the thalamus, brain stem and cerebellum WM were significantly different ($d > 0.8$, $p < 0.05$) between patients with low and severe outcome disability. Significant group differences were found for cortical GM atrophy and ventricular enlargement with effect sizes above 0.7. Cerebral WM is, however, the only individual structure which remained statistically different between the groups after correcting for multiple comparisons.

The most discriminative structure in terms of accuracy was atrophy of the whole brain (bACC: 77 %, SENS: 70%, SPEC: 84%) and the brain stem (bACC: 76 %, SENS: 70%, SPEC: 81%).

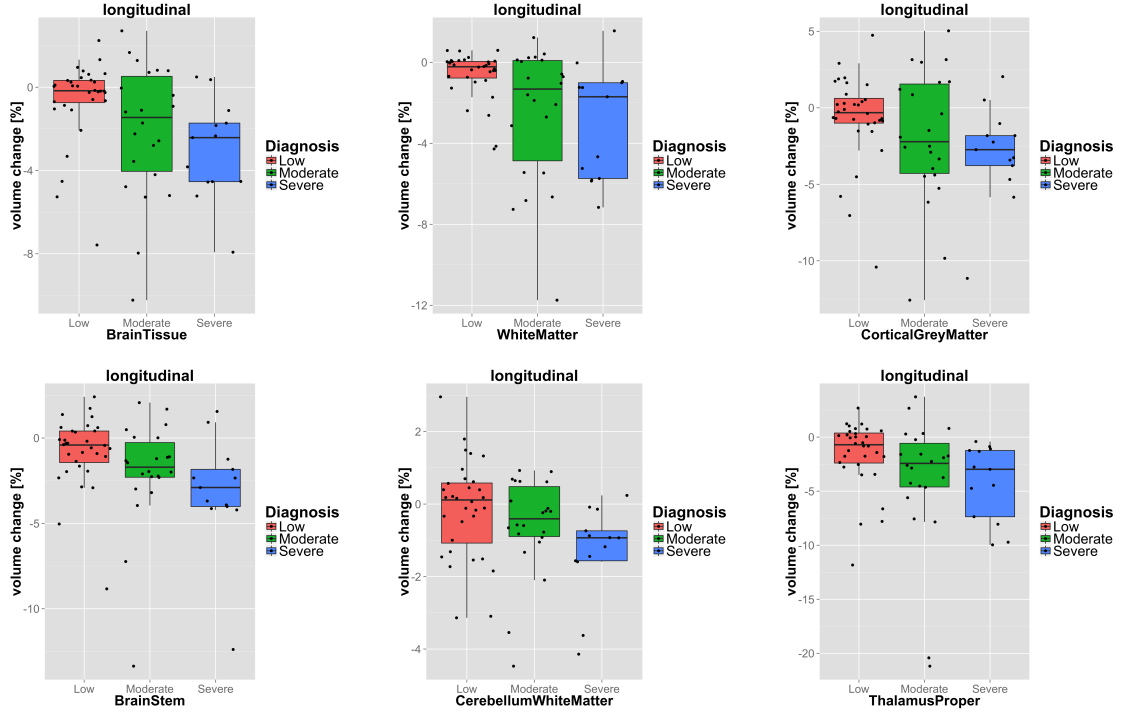


Figure 6.5: Boxplots of change rates of selected ROIs with respect to the investigated disease outcome groups. Features selected based on their performance in classifying severe disability vs. low disability outcome (*cf.* Table 6.4).

A combination of all measured atrophy rates in a multi-feature classifiers did not further improve classification results, *e.g.* using a RF classifier resulted in bACC: 69 %.

6.5 Discussion

The potential presence of pathologies such as haemorrhage lesions, contusions or a substantial midline shift pose particular challenges for the analysis of brain MRI in TBI. Three example image series with a high degree of injury are visualised with overlaid segmentation results in Figure 6.6. Clinical information and a brief description of the images is provided in the figure caption.

These examples show clear segmentation inaccuracies observed for example for the hippocampus, putamen or the caudate (white arrows). The inclusion of haemorrhage lesions in both WM and cortical GM ROIs (blue arrows) is a limitation of the proposed segmentation framework. This is expected as atlas-based approaches are restricted to the labelling of structures that are represented in the reference atlases. Assuming that explicit segmentations of pathologies (*e.g.* lesions) are available (*e.g.* via dedicated lesion segmentation), this allows to determine the location of le-

Table 6.4: Classification results in % (6-fold cross-validation, LDA/RF/SVM 100 runs) obtained separating TBI patients with a severe disability from patients with low disability outcome based on structural volume changes between the acute and chronic disease stage. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses.

Severe disability (N = 13, Positives ^P) vs. Low disability (N = 32, Negatives ^N) (longitudinal analysis, bl→follow-up)									
structure	ACC (bACC)	SENS	SPEC	mean (SD) [%] ^P	mean (SD) [%] ^N	effect size (d)	p-value	sig. (corr.)	
RandomForest (all features)	77 (69)	52	86						
SVM (all features)	75 (64)	36	91						
Gender (female = 0, male = 1)	67 (67)	69	66	0.3 (0.5)	0.7 (0.5)	0.723	0.03331	+ (o)	
Age	55 (55)	54	55	67.2 (15.9)	61.1 (10.1)	0.514	0.12518	o (o)	
WhiteMatter	75 (67)	47	87	-3.0 (2.8)	-0.6 (1.2)	1.323	0.00023	++ (+)	
BrainTissue	79 (75)	66	84	-3.0 (2.4)	-0.6 (2.0)	1.130	0.00133	+ (+)	
Brain	80 (77)	70	84	-2.0 (1.7)	-0.4 (1.6)	1.020	0.00339	+ (o)	
DeepGreyMatter	70 (69)	64	73	-2.7 (2.2)	-0.5 (2.5)	0.896	0.00929	+ (o)	
CorticalGreyMatter	78 (75)	66	84	-3.0 (3.2)	-0.7 (2.9)	0.780	0.02223	+ (o)	
Ventricles	68 (63)	50	76	11.6 (12.0)	4.2 (9.7)	0.711	0.03627	+ (o)	
CerebralWhiteMatter	75 (67)	47	87	-3.1 (3.0)	-0.7 (1.2)	1.293	0.00030	++ (+)	
CerebellumExterior	70 (67)	60	74	-3.1 (3.1)	-0.3 (2.9)	0.933	0.00691	+ (o)	
BrainStem	78 (76)	70	81	-3.1 (3.3)	-0.8 (2.1)	0.922	0.00758	+ (o)	
CerebellumWhiteMatter	72 (72)	73	72	-1.3 (1.3)	-0.1 (1.3)	0.884	0.01019	+ (o)	
ThalamusProper	72 (66)	52	80	-4.2 (3.5)	-1.6 (3.1)	0.824	0.01605	+ (o)	
LateralVentricle	69 (63)	51	76	12.8 (12.0)	4.7 (10.7)	0.725	0.03289	+ (o)	
Amygdala	57 (47)	26	69	-2.6 (6.7)	0.0 (2.5)	0.622	0.06518	o (o)	
Hippocampus	73 (65)	46	83	-4.8 (10.6)	-1.1 (2.9)	0.614	0.06895	o (o)	
CerebellarVermalLobulesVI-VII	68 (65)	58	72	-2.1 (3.9)	-0.1 (3.1)	0.613	0.06917	o (o)	
CSF	66 (61)	49	72	6.2 (6.5)	3.2 (6.0)	0.500	0.13556	o (o)	
Putamen	58 (51)	34	67	-0.1 (0.8)	0.1 (0.4)	0.497	0.13840	o (o)	
VentralDC	76 (74)	68	79	-2.5 (1.8)	-1.2 (3.0)	0.480	0.15191	o (o)	
4thVentricle	61 (61)	61	61	0.3 (7.5)	-2.2 (4.5)	0.468	0.16234	o (o)	
InfLatVent	69 (64)	52	76	7.7 (19.6)	3.0 (8.0)	0.380	0.25411	o (o)	
CerebellarVermalLobulesI-V	59 (57)	52	62	-0.9 (1.6)	-0.5 (1.0)	0.355	0.28602	o (o)	
CerebellarVermalLobulesVIII-X	65 (64)	62	66	-0.3 (1.9)	0.4 (2.3)	0.331	0.32032	o (o)	
3rdVentricle	56 (53)	45	60	5.2 (8.4)	2.4 (9.2)	0.316	0.34251	o (o)	
Caudate	39 (37)	30	43	0.1 (12.6)	-1.1 (6.6)	0.132	0.69072	o (o)	
AccumbensArea	49 (47)	43	52	-0.2 (3.0)	-0.5 (1.7)	0.118	0.72056	o (o)	
BasalForebrain	44 (41)	34	48	1.2 (16.6)	2.7 (14.6)	0.096	0.77201	o (o)	
Pallidum	56 (44)	15	72	0.2 (0.8)	0.2 (1.1)	0.023	0.94465	o (o)	

sions relative to anatomical structure. However, in the present approach this can lead to a substantial bias as it artificially increases a structure's volume (*e.g.* cortical GM, blue arrows). The inclusion or exclusion of pathology in the brain mask (orange arrow) can have an even more severe effect. It can lead, as shown in the top row of Figure 6.6, to an overestimation of the reduction of brain tissue or WM volume. The explicit segmentation of pathology or structures that are not available in the atlases is highly desirable in this context and should be pursued in future research. Another limitation is related to the refinement of spatial priors introduced in Chapter 3. The refinement of priors compensates inaccuracies in the atlas alignment based on the intensity profiles within the MR image. However, priors between labels with similar intensity profile are not corrected. In extreme cases this can lead to mislabeling of these structures (*e.g.* CSF $\leftrightarrow$ background). An example can be observed in the segmentation results of TBI047 (middle row) or TBI114 (bottom row) where parts of the inferior lateral ventricle are incorrectly labelled as background (*cf.* red arrow).

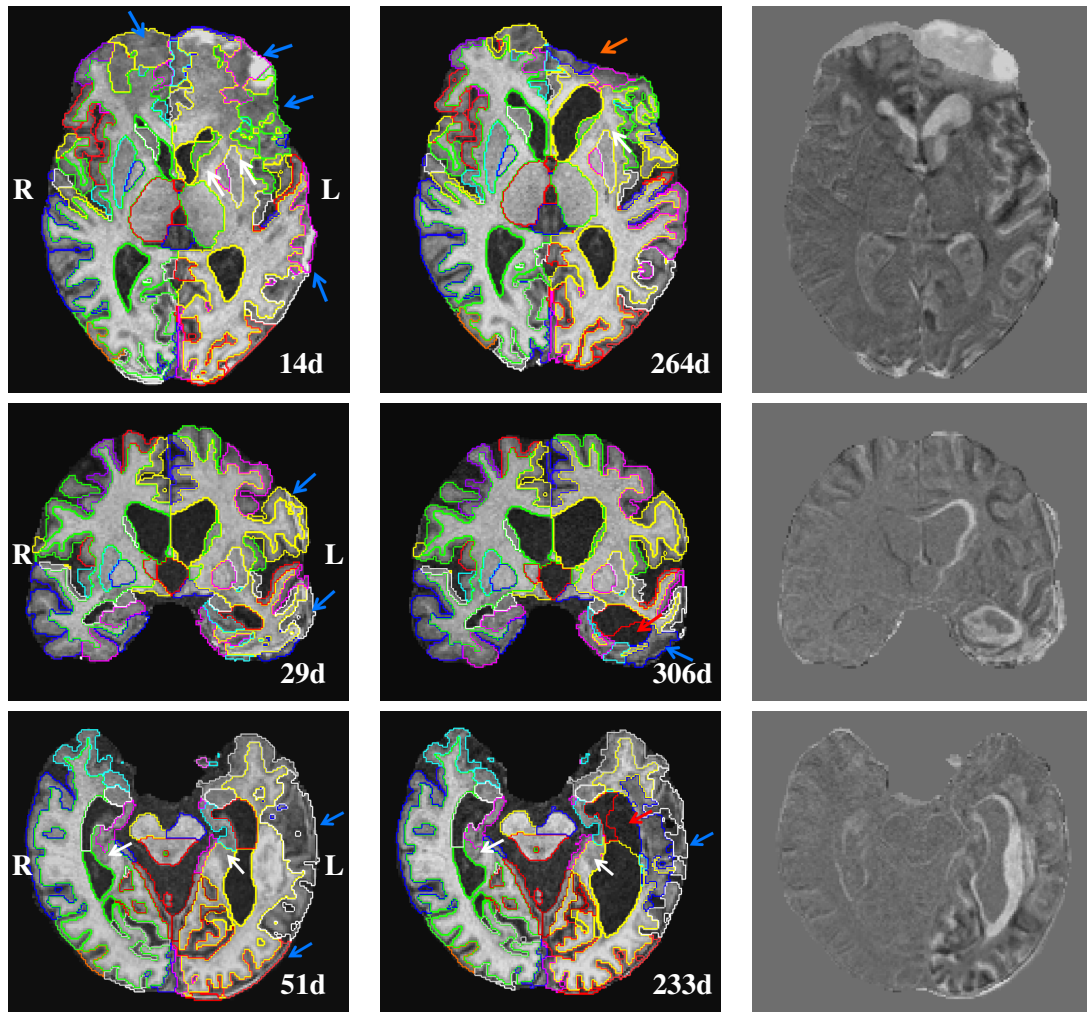


Figure 6.6: Example longitudinal segmentation results of images at the acute and chronic stage of a TBI. **Top row, TBI150:** male, 68 years of age, GCS: 15, GOSe: 5, MCS: 5, unclear accident, acute/chronic image acquired 14/264 days after injury, craniotomy and evacuation of an acute SDH, intracranial haematoma induced brain destruction². **Middle row, TBI047:** male, 62 years of age, GCS: 13, GOSe: 3, MCS: 5, fall accident, acute/chronic image acquired 29/306 days after injury, craniotomy and evacuation of an acute subdural haemorrhage, broad gliosis on left temporal lobe. **Bottom row, TBI114,** male, 69 years of age, GCS: 14, GOSe: 5, MCS: 5, cycling accident, acute/chronic image acquired 51/233 days after injury, no craniotomy, traumatic subarachnoid haemorrhage, intracranial haematoma with surrounding oedema in left temporal lobe, central and cortical atrophy.

²these are incomplete reproductions of the radiological reports

In this chapter, the segmentation quality could not be evaluated quantitatively. Manual reference segmentations of the images were not available and a quantitative quality assessment through experts (*cf.* Chapter 3) is prohibitive. Thus a different approach was used by assessing the methods

indirectly. Instead of quantifying the actual segmentation quality, the disease-relevance of extracted measurements in the context of classification was analysed. Specifically, several classification experiments were conducted to investigate whether measures of structural volume, asymmetry and volume change are meaningful features in patients with TBI. In this setup both the assessment of segmentation accuracy and the validation of measured features as predictors of disease outcome are coupled a priori. In Chapter 3 and Chapter 4 the employed segmentation methodology was evaluated and accuracy assessed quantitatively. Further, a non-expert visually reviewed the segmentations calculated in this chapter to rule out gross failures. *Cf.* Figure 6.2 and Figure 6.4 for included subjects with segmentations considered successful, however, affected by pathology. With this limitation in mind, the assumption is made that anatomical structures throughout the study cohort were segmented successfully. The goal of the performed analysis was therefore to show the potential of derived quantitative measures as valid biomarkers that are predictive for the outcome of a head trauma.

Quantified features extracted from acute MR imaging allowed a specific prediction whether a patient will have a negative outcome diagnosis (SPEC: 94%), when classified against the low disability outcome group. In this experiment a high sensitivity was observed using volumes of structural ROIs (*e.g.* SENS accumbens: 91%, SENS hippocampus: 83%) and high specificity using symmetry-based features (*e.g.* SPEC AAI of all structures: 91%). This suggests that brain symmetry is only a “necessary” criterion for a favourable disease outcome while asymmetry is a “sufficient” criterion for unfavourable outcome. This is in agreement with the findings presented in Chapter 3 on another dataset. However, it must be noted that there is a trade-off between SENS and SPEC, which should be further investigated using for example ROC curves (*cf.* Chapter 3, Figure 3.11).

In Strangman et al. (2010) the authors reported that structural volumes correlate with the potential of patients to recover within a memory rehabilitation program. In this study, structural volumes were extracted from 50 TBI patients several years after the injury in the chronic stage. The authors raise the question whether structural volumes extracted at the acute stage have similar potential. The results of the presented study in this chapter are encouraging: structural volumes (*e.g.* hippocampus, thalamus) were identified that are predictive for the outcome of the disease. These findings agree with those presented in Strangman et al. (2010). The influence of brain capacity/reserve on the ability to recover from a TBI needs to be further investigated.

The conducted longitudinal analysis revealed increased atrophy in the WM, brainstem and tha-

lamus which was also shown in other studies (Warner et al., 2010b; Newcombe et al., 2015). In contrast to Warner et al. (2010b), significant changes for the cerebellum but not for the amygdala were found in this chapter. However, when comparing to other studies the substantial heterogeneity of TBI studies must be considered: In comparison to the study presented in this thesis, Warner et al. (2010b) analysed 25 patients with DAI and 22 age-matched controls where patients had a lower GCS of 6.2 ± 4.5 (mean $\pm$ SD) and were much younger (26.8 ± 11.3 years).

In the conducted analyses, the significance of many findings could not be confirmed after correcting for multiple comparisons. However, the performed Bonferroni correction is very conservative and the calculated effect sizes indicate that significance levels could be increased on a larger cohort. In summary, the experiments confirm that the developed algorithms can be valuable when automatically analysing cohorts with images covering a wide range of significantly altered brain anatomy.

A simple approach was chosen to account for the uneven age distribution between patients of the investigated outcome categories. This resulted in a substantial reduction of the number of study subjects ($N = 131 \rightarrow N = 67$). Even though no subjects were excluded from the severe disability outcome group ($N = 13$), this is a limitation of the conducted experiments as the number of subjects (or samples) per group is small. In the future the applicability of more sophisticated methods to adjust for selection bias, such as inverse-probability weighting (Haneuse et al., 2009), should be explored.

In addition to the performed age matching, experiments were carried out to employ the regression models trained on the healthy control subjects from ADNI (*cf.* Chapter 5) to account for differences in age, gender and brain size. Preliminary results that are not included here suggest that correction for brain size is not beneficial. One possible explanation is that the calculated brain size estimate may be distorted by TBI-related brain pathology, such as lesions or contusions. Correcting for either age and gender did not substantially alter the results presented in Table 6.3. Correcting for both age *and* gender using a multi-variate regression approach reduced the discriminative value of the investigated structures. It should be noted that the ADNI subjects are overall older than the TBI subjects so that the trained model needs to be extrapolated in order to be applicable to the TBI dataset. One hypothesis is therefore that the regression model trained on ADNI data is not straightforward applicable, due to confounding cohort differences in patient age but also acquisition protocol. These effects require a deeper analysis so that no explicit correction for age or gender was performed.

No clear benefit was observed of combining all available measurements within multi-feature classifiers except when predicting severe and low disability outcome at baseline based on structural volumes. Classifiers were not explicitly tuned. Given the size of the study cohort and the large number of features a lack of generalisation of the classifiers/overfitting might be a problem that should be further investigated in the future.

6.6 Conclusion

In this chapter 67 subjects were analysed from a recently acquired cohort of mild to moderate TBI patients. The conducted analyses demonstrate that the methodology developed in Chapters 3 and 4 allows the extraction of meaningful biomarkers from MR brain images such as the volume or volumetric change of individual ROIs. It was confirmed that imaging information can add predictive value when performing an outcome prognosis at the acute stage of the injury. Structural volumes, measured from acute MR images, of the accumbens, hippocampus, amygdala and thalamus were related to the disease outcome. Both WM and brain atrophy was increased in patients with unfavourable outcome prognosis. Overall, the employed methodology has potential to support automated brain morphometry in patients with TBI.

7 Patch-based Evaluation of Image Segmentation

This chapter is based on:

C. Ledig, W. Shi, W. Bai, and D. Rueckert, “Patch-based Evaluation of Image Segmentation”, Computer Vision and Pattern Recognition (CVPR), pp. 3065-3072, 2014.

Abstract: *The quantification of similarity between image segmentations is a complex yet important task. The ideal similarity measure should be unbiased to segmentations of different volume and complexity, and be able to quantify and visualise segmentation bias. Similarity measures based on overlap, e.g. Dice score, or surface distances, e.g. Hausdorff distance, clearly do not satisfy all of these properties. To address this problem, Patch-based Evaluation of Image Segmentation (PEIS) is introduced, a general method to assess segmentation quality. The method is based on finding patch correspondences and the associated patch displacements, which allow the estimation of segmentation bias. Both the agreement of the segmentation boundary and the conservation of the segmentation shape is quantified. Furthermore, the segmentation complexity within patches is assessed to weight the contribution of local segmentation similarity to the global score. PEIS is evaluated on both synthetic data and two medical imaging datasets. On synthetic data evidence is provided that PEIS, in comparison to the Dice score, produces more comparable scores across segmentations of varying size, has increased sensitivity and estimates segmentation bias accurately. On cardiac MR images, it is demonstrated that PEIS can evaluate the performance of a segmentation method independent of the complexity of the segmentation under consideration. On brain MR images, five different automatic hippocampus segmentation techniques are compared using PEIS. Finally, the segmentation bias is visualised on a selection of the cases.*

The main goal of this thesis is the development of automatic whole-brain segmentation methods. These methods have been evaluated using established similarity measures such as the Dice coefficient (Dice, 1945) (DC). While DC is a widely used measure to compare competing segmentation approaches (Landman and Warfield, 2012; Asman et al., 2013a) it has some limitations such as the strong dependence of the observed value range on the size and shape of the segmentation under investigation. The aim of this chapter is to present an alternative similarity measure that overcomes some of the limitations of existing techniques and can provide relevant additional information when evaluating image segmentations.

7.1 Introduction

The validation of automatic segmentation methods usually relies on the comparison to reference segmentations, ideally annotated by an expert observer. However, evaluation is inherently difficult as similarity features are hard to define and potentially subjective. Furthermore manual expert segmentations cannot be considered as gold standard as they are subject to both significant inter- and intrasubject variability (Chalana and Kim, 1997; Gerig et al., 2001; Warfield et al., 2004). Warfield et al. (2004) presented a popular method, STAPLE (*cf.* Section 2.1.3.6), to estimate a probable ground truth given several manual expert segmentations. However, assuming a reliable ground truth or reference segmentation is available, the problem of comparison remains. The importance of quantifying segmentation accuracy in the context of biological imaging was formulated almost 40 years ago (Yasnoff et al., 1977; Udupa et al., 2006). However, while the development of segmentation algorithms is an active research area, the progress in developing evaluation methods is more limited (Zhang, 1996; Udupa et al., 2006; Chang et al., 2009).

The measurement of similarity between two different sets or image segmentations is thus an ongoing challenge (Dice, 1945; Tversky, 1977; Unnikrishnan et al., 2007; Cárdenes et al., 2009). The most popular approaches to quantify segmentation similarity in medical imaging are based on overlap (Dice, 1945; Crum et al., 2006), with the Dice coefficient being the most renowned representative. With the so called *ratio model* Tversky (1977) presented a framework that generalised the widely used Dice (Dice, 1945) and Jaccard (Jaccard, 1912) similarity index. A thorough examination of the Dice and Jaccard index is presented in Chang et al. (2009). Generalised overlap measures focusing on fuzzy multi-label segmentations are described in Crum et al. (2006). In Cárdenes et al.

(2009), the authors modified the ratio model by weighting wrongly segmented pixels in 2D, or voxels in 3D respectively, based on their spatial distance from the reference label. Another family of measures is based on the surface distance between segmentations (Chalana and Kim, 1997; Aspert et al., 2002; Pichon et al., 2004; Cárdenes et al., 2009; Heimann et al., 2009). Examples are the Yasnoff discrepancy (Yasnoff et al., 1977) or the Hausdorff distance (Huttenlocher et al., 1993), which is highly sensitive to errors (Pichon et al., 2004). Taking the distance between misclassified voxels and ground truth into account is important (Pichon et al., 2004; Crum et al., 2006; Cárdenes et al., 2009).

Next to this, other methods have been proposed, often with focus on a particular application. In Gerig et al. (2001) the authors have developed a tool that assesses object segmentations with both overlap and distance-based measures. By visualising segmentation differences the tool was employed to increase intrarater reliability of manually obtained hippocampus segmentations (Gerig et al., 2001). Pichon et al. (2004) propose to use statistics based on surface distances to estimate segmentation bias. Another method employs confidence maps of multiple reference segmentations to quantify segmentations in 2D (Restif, 2007). Juneja et al. (2013) introduced the validation index which is based on multiple expert segmentations to support radiotherapy planning. Recently a method based on genetic programming was proposed (Vojodi et al., 2013) that combines single measures for colour image segmentation.

A good overview over different segmentation evaluation approaches is given in Zhang (1996); Udupa et al. (2006); Zhang et al. (2008).

7.1.1 Contribution and overview

In the following, inspired by Unnikrishnan et al. (2007), the desired attributes of a similarity measure for image segmentation are postulated:

- *Comparable scores*: Scores calculated on segmentations of varying size and boundary complexity should be comparable. This is desirable as it renders the comparison of automatic segmentation methods less sensitive towards the evaluation dataset or characteristics of the reference segmentation under consideration.
- *Shape conservation*: The measure should assess the preservation of the shape of segmentation boundaries. This is desirable as varying definitions of segmentation protocols or the

presence of partial volume effects in low-resolution images can produce different segmentation boundaries that are, however, similar in shape.

- *Boundary conformity*: It should measure the agreement of segmentation boundaries. This is desirable as a good segmentation not only preserves shape but also matches the segmentation boundaries.
- *Segmentation bias*: The measure should allow the detection and visualisation of segmentation bias, such as over- or under-segmentation. This is desirable as it provides an insight into the systematic differences between segmentations. Visualisation of segmentation bias can also support the development and tuning of automatic segmentation methods.
- *Similarity*: The ability to quantify *visual* similarity rather than pure overlap. This is desirable as especially segmentations of structures with a high surface to volume ratio, such as vessel segmentations, can be highly similar while having imperfect overlap.

Based on the ratio model (Tversky, 1977), the Patch-based Evaluation of Image Segmentation (PEIS) is proposed. PEIS is formulated for binary segmentations. While PEIS can be generalised to multi-label segmentations, this is beyond the scope of this chapter and left for future work.

For each voxel a patch displacement vector is found that locally transforms the reference segmentation into the test segmentation. This voxel-wise displacement allows the estimation and visualisation of segmentation bias. It furthermore allows the quantification of how well a test segmentation conserves shape and matches segmentation boundaries. The presented approach is inspired by patch-based methods that have been used successfully in image synthesis (Liang et al., 2001) or image denoising (Kervrann and Boulanger, 2006). More recently patch-based approaches have also enjoyed increasing popularity in the medical imaging community and have been applied to image segmentation (Coupé et al., 2011; Asman and Landman, 2013). Secondly, it is proposed to employ the local segmentation complexity within patches to control the contribution of local segmentation similarity to the global score. This allows improving sensitivity in regions that are non-trivial to segment. The approach extends pure overlap measures and allows for a quantitative assessment of segmentation bias. PEIS is evaluated on both synthetic and real medical imaging data: PEIS is compared to the Dice score on synthetic data and to both Dice score and the average surface distance (ASD), as used in Heimann et al. (2009), on real datasets from cardiac and brain MR images. Visual evidence is provided that PEIS allows the visualisation of segmentation bias.

The remaining chapter is organised as follows: The original formulation of the ratio model is reviewed and PEIS is presented in Section 7.2. PEIS is then evaluated on synthetic and real datasets in Section 7.3. The findings will be discussed in Section 7.4. Finally conclusions are formulated and future work is outlined in Section 7.5.

7.2 Method

7.2.1 Framework: The ratio model (RM)

Given a binary reference segmentation R , the goal is to measure the similarity of a test segmentation $S = \{s_{i=1\dots N}\}$ to $R = \{r_{i=1\dots N}\}$. In the following it is assumed that $s_i, r_i \in \{0, 1\}, \forall i \in \Omega \subset \mathbb{N}^d$. Here $d = 2, 3$ is the image dimension and i serially numbers the N voxels in Ω . Ω is the subset of $\mathbb{N}^d$ where not both s_i and r_i are zero.

The ratio model (RM) framework, originally described by Tversky (1977), can be stated as:

$$RM(S, R) = \frac{\theta f(S \cap R)}{\theta f(S \cap R) + \alpha f(S \setminus R) + \beta f(R \setminus S)} \quad (7.1)$$

In this chapter a similar notation to the one in Tversky (1977) is used, where the scale f describes a family of similarity measures based on the global parameters θ, α, β . In the standard formulation $f(S)$ is the cardinality of the set S . For $\theta = 1$ and $\alpha = \beta = \frac{1}{2}$ this model reduces to the widely used Dice coefficient (Dice, 1945). For $\theta = 1$ and $\alpha = \beta = 1$ the formulation yields the Jaccard coefficient respectively (Jaccard, 1912). More details on the characteristics of this formulation can be found in Tversky (1977).

In the context of classification, one can consider $f(S \cap R)$ as true positive (TP), $f(S \setminus R)$ as false positive (FP) and $f(R \setminus S)$ as false negative (FN) fraction of the test segmentation S and the reference segmentation R . The false positives and false negatives are summarised as false labels (FL = FP + FN). Thus the non-directional case, $\alpha = \beta$, is modelled where the denominator reduces to $\theta TP + \alpha FL$. Here false positives and false negatives are equally penalised. In the following a novel way of modelling the scale f is presented.

7.2.2 Patch-based ratio model (PBRM)

Standard overlap measures are restricted to the evaluation of the voxel-wise agreement of segmentations. Here, a novel patch-based approach is proposed to relax this constraint by finding a voxel-wise displacement that transforms the reference segmentation R *locally* into the test segmentation S . A patch extraction operator is given by $\mathcal{N}_p(\cdot)$. $\mathcal{N}_p(R(i))$ extracts a patch of the reference segmentation at position i , $\mathcal{N}_p(S(j))$ a patch of the test segmentation at position j accordingly.

To compare two patches the sum of absolute differences is calculated between $\mathcal{N}_p(R(i))$ and $\mathcal{N}_p(S(j))$. This difference is defined as:

$$\mathcal{D}(i, j) = |\mathcal{N}_p(R(i)) - \mathcal{N}_p(S(j))| \quad (7.2)$$

At a certain position i an optimal patch displacement $\Delta(i) = (\Delta(i)_x, \Delta(i)_y, \Delta(i)_z)$ is derived that locally minimises $\mathcal{D}(i, j)$. This is equivalent to maximising the agreement of a fixed patch in R , $\mathcal{N}_p(R(i))$, with a moving test patch in S , $\mathcal{N}_p(S(j))$. The optimisation is described in Section 7.2.3. In the following, i' denotes the index that yields the best locally matching test patch $\mathcal{N}_p(S(i'))$ for a given reference patch $\mathcal{N}_p(R(i))$. Thus i' corresponds to i displaced by the optimal displacement $\Delta(i)$. For a certain displacement $\Delta(i)$, the spatial patch overlap between the reference patch and the displaced test patch is defined as $\mathcal{A}(\Delta(i))$. Spatial overlaps, $\mathcal{A}(\Delta)$, are illustrated for different displacements Δ in yellow in Figure 7.1. $\mathcal{A}(\Delta)$ only depends on Δ and is thus independent of the segmentation content within the patches.

Assuming that the optimal displacement $\Delta(i)$ and thus the corresponding i' was found, a voxel-wise segmentation similarity index $\gamma \in [0; 1]$ is defined as $\gamma(i) = 1.0 - \frac{\mathcal{D}(i, i')}{|\mathcal{N}_p(\cdot)|}$ and a spatial similarity index $\tau \in [0; 1]$ as $\tau(i) = \frac{\mathcal{A}(\Delta(i))}{|\mathcal{N}_p(\cdot)|}$. These indices are normalised by the number of voxels within a patch, $|\mathcal{N}_p(\cdot)|$, and thus the maximal possible patch dissimilarity. By defining τ based on the amount of spatial overlap of the corresponding patches, it is ensured that both γ and τ take on comparable values. This formulation thus allows the averaging of γ and τ to a combined similarity score $\eta(i) = 0.5 \times (\gamma(i) + \tau(i))$.

In the notation of the ratio model (Tversky, 1977), $\sum_i \eta(i)$ denotes the true positive fraction, $f(S \cap R)$, while $\sum_i (1.0 - \eta(i))$ summarises the falsely segmented fraction, $f(S \setminus R)$ and $f(R \setminus S)$.

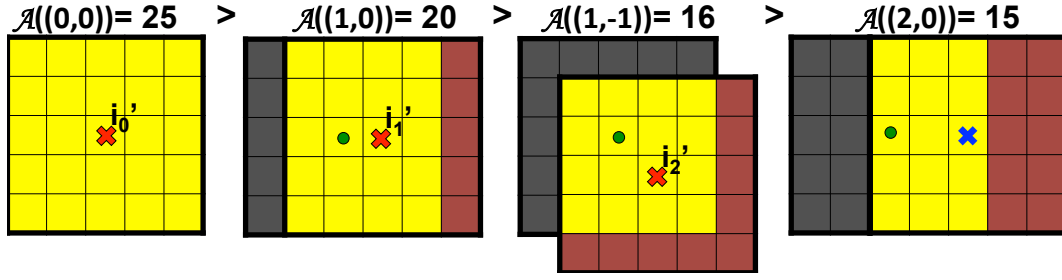


Figure 7.1: The *spatial* overlaps $\mathcal{A}$ between the *supports* of a reference patch (grey) and a moving test patch (dark red) is shown in yellow. For increasing patch displacements $\Delta = (\Delta_x, \Delta_y)$ the spatial overlap decreases. If two displacements have the same L1-norm (both examples shown on the right with $\|(1, -1)\|_1 = \|(2, 0)\|_1 = 2$) the proposed method favours a diagonal displacement as $\mathcal{A}(1, -1) > \mathcal{A}(2, 0)$. The reference centre is indicated as circle. The centres of the test patches are indicated as crosses and correspond to the displacements shown in Figure 7.2.

The patch-based ratio model (PBRM) is formulated as:

$$PBRM(S, R) = \frac{\theta \sum_i \eta(i)}{\theta \sum_i \eta(i) + \alpha \sum_i (1.0 - \eta(i))} \quad (7.3)$$

7.2.3 Optimisation of local patch correspondences

To ensure that a *locally* optimal patch correspondence i' is found it is assumed that $\mathcal{D}(i, j)$ monotonically decreases with increasing distance d_{ij} from the reference position i . Here d_{ij} denotes the L1-distance between the position of two voxels i and j in image space. To obtain i' , solutions i'_k are calculated for different levels of distance k that minimise $\mathcal{D}(i, j)$ for all j with L1-distance d_{ij} equal to k ($\forall j : d_{ij} = k$). At a certain level k only test patches are considered that are centred at a voxel j with L1-distance k from i , $d_{ij} = k$. Specifically, at a certain level k the following optimisation problem is solved to calculate i'_k for a given position i in the reference segmentation:

$$\begin{aligned} i'_k = \underset{j}{\operatorname{argmin}} \quad & \mathcal{D}(i, j) \\ \text{s. t.} \quad & d_{ij} = k \end{aligned} \quad (7.4)$$

The optimisation process for an example reference patch is shown in Figure 7.2. In the case that there are multiple possible solutions that minimise $\mathcal{D}(i, \cdot)$, the first found displacement is chosen

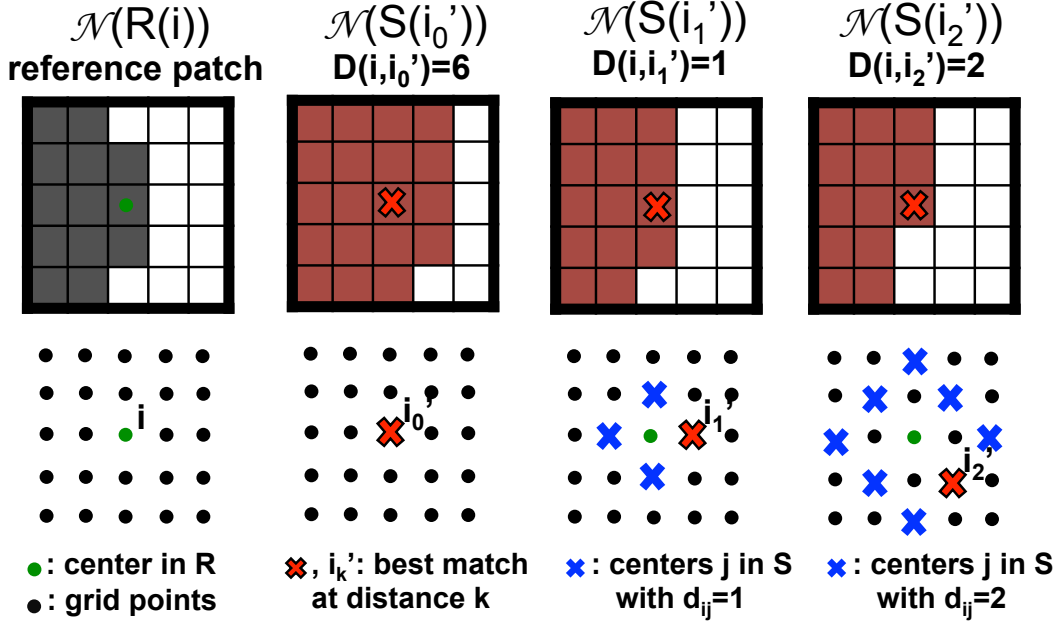


Figure 7.2: Calculation of optimal patch centers i'_k at different distance-levels k . An example of a reference patch is shown in grey, while the best matching patches at the three corresponding distance-levels ($k = 0, 1, 2$) are shown in dark red.

that yields the maximal spatial patch overlap $\mathcal{A}(\Delta)$ which is illustrated in Figure 7.1. This will, for a certain distance-level k , favour diagonal over purely horizontal or vertical displacements (*cf.* Figure 7.1).

Then the *locally* optimal voxel-specific displacement $\Delta(i)$, determined by i' , is calculated by solving the following optimisation problem:

$$\begin{aligned}
 i' = \underset{i'_k}{\operatorname{argmin}} \quad & \mathcal{D}(i, i'_k) \\
 \text{s. t.} \quad & \mathcal{D}(i, i'_0) \geq \mathcal{D}(i, i'_1) \geq \dots > \mathcal{D}(i, i'_k)
 \end{aligned} \tag{7.5}$$

The monotonicity constraint ensures that the best matching patch is found locally. The proposed strategy also renders the search very efficient, as it provides an efficient stopping criterion for the calculation of i'_k . The calculation can be stopped once $\mathcal{D}(i, i'_{k+1}) > \mathcal{D}(i, i'_k)$. With this formulation the need for a fixed search neighbourhood is avoided. This is usually required to allow for a tractable computation. Hence, the presented approach can find arbitrary large displacements.

7.2.4 Patch-based Evaluation of Image Segmentation (PEIS)

Determining the parameters θ, α globally yields a measure to which correctly or falsely segmented voxels equally contribute. For large structures with a low surface to volume ratio this is problematic as the overall measure is dominated by the large number of true positives in the interior of the structure. This problem is addressed by introducing spatially varying parameters $\theta(i), \alpha(i)$ into the model. The following formulation presents a data-driven way to increase the measure's sensitivity in the relevant regions, *e.g.* regions close to the boundary, when comparing structures.

The following describes the complete patch-based spatially varying ratio model:

$$\text{PEIS}(S, R) = \frac{\sum_i \theta(i) \eta(i)}{\sum_i \theta(i) \eta(i) + \sum_i \alpha(i) (1 - \eta(i))} \quad (7.6)$$

There are different ways to model $\theta(i)$ and $\alpha(i)$. In this work it was chosen to determine $\theta(i)$ and $\alpha(i)$ based on the complexity of the reference patch under consideration. This complexity is quantified by counting the number of facets $n_{\text{facets}}(i)$ that separate adjacent voxels with different labels in the reference patch $\mathcal{N}_p(R(i))$ with centre i . Then $\theta(i)$ and $\alpha(i)$ can be calculated as:

$$\theta(i) = \frac{n_{\text{facets}}(i)}{n_{\text{facets}}^{\max}} \quad \alpha(i) = 1.0 - \theta(i) \quad (7.7)$$

Here $n_{\text{facets}}^{\max}$ is chosen in the order of the number of voxels within a patch of maximal complexity. For a given patch width or patch size p_w the maximal number of facets is chosen to be $n_{\text{facets}}^{\max} = 4 \times (p_w - 1)$ (2D) or $n_{\text{facets}}^{\max} = 4 \times (p_w - 1) \times p_w$ (3D). This corresponds to the number of facets contained in a patch separated by a diagonal line (2D) or diagonal plane (3D). Example patches of varying patch complexity are shown in Figure 7.3. Furthermore, $n_{\text{facets}}^{\max}$ serves as a threshold for $n_{\text{facets}}(i)$ to ensure $0 \leq \theta(i) \leq 1$.

With this formulation the contribution of correct labels $\theta(i)$ is higher in patches that are difficult to segment (high complexity) than in those that have little complexity or are even homogeneous ($n_{\text{facets}}(i) = 0$). In contrary, false labels of “easy to segment” or quite homogeneous regions have a high weight $\alpha(i)$ and thus degrade the measure more seriously than false labels in “difficult to segment” patches.

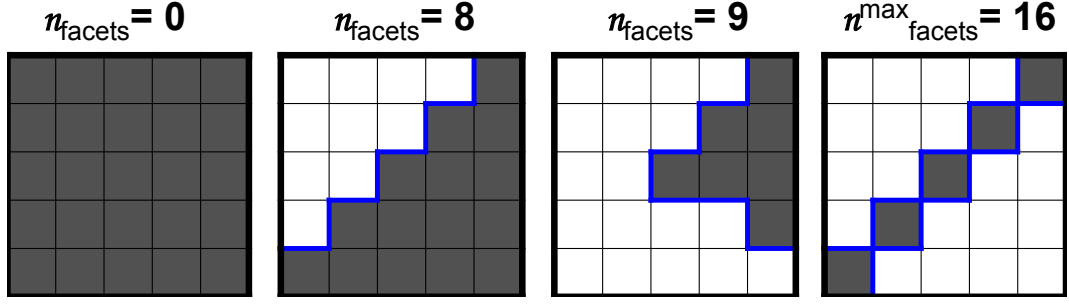


Figure 7.3: Example reference patches of varying complexity. Foreground segmentation is shown in grey and corresponding facets in blue. The patch shown on the right has maximal complexity of $n_{\text{facets}}^{\text{max}} = 4 \times (p_w - 1) = 16$ for a 2D patch with width $p_w = 5$.

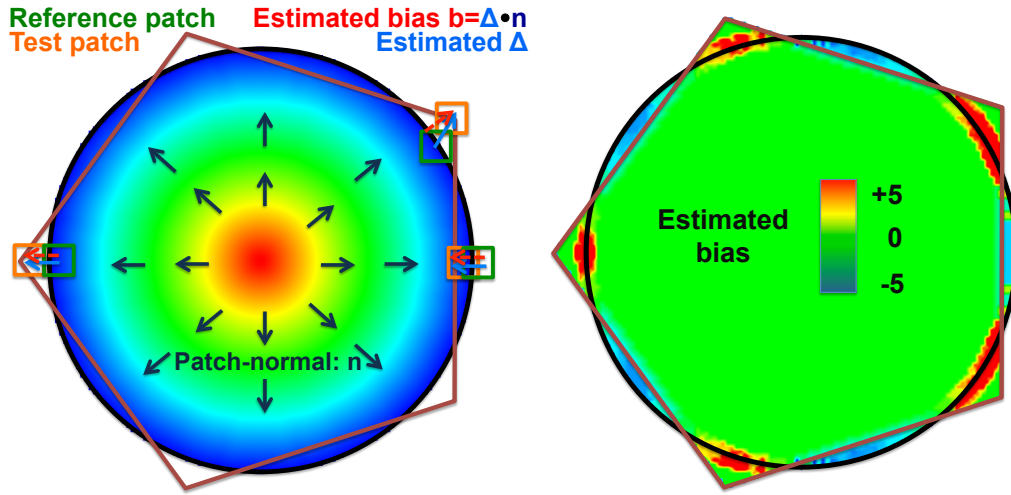


Figure 7.4: left: Example comparison of a polygon (test segmentation, outlined in dark red) to a circle (reference, outlined in black). Patch-correspondences of fixed reference patches (green) and moving test patches (orange). Vectors represent patch-normals $\mathbf{n}$ (dark blue), optimal patch displacements Δ (light blue) and the projections of Δ on $\mathbf{n}$ (red) which represents the estimated segmentation bias. The distance transform of the reference segmentation is shown in a rainbow colour scheme. right: Estimated voxel-wise segmentation bias $b(i) = \mathbf{n}(i) \cdot \Delta(i)$ within the domain Ω .

7.2.5 Quantification of segmentation bias

In addition to the similarity score, the presented approach also provides a displacement vector $\Delta(i) = (\Delta(i)_x, \Delta(i)_y, \Delta(i)_z)$ quantifying the local spatial difference between test and reference segmentation. In the following, a single measure is derived that quantifies the segmentation bias based on Δ . Specifically, the interest is if a segmentation is too generous (“over-segmentation”) or too conservative (“under-segmentation”).

First, for a patch with centre i a vector is calculated that is referred to as patch-normal $\mathbf{n}(i)$. This patch-normal is calculated based on the Euclidean distance transform $\mathcal{T}(R)$ of the zero level-set in the reference segmentation. Specifically, $\mathbf{n}(i)$ is computed as the average first derivative of $\mathcal{T}(R)$ at segmented voxels in the reference patch $\mathcal{N}_p(R(i))$. This can be formalised as:

$$\mathbf{n}(i) = -\frac{1}{Z} \sum_{j \in \mathcal{N}_p(R(i))} \delta(R(j)) \nabla \mathcal{T}(R)(j) \quad (7.8)$$

Here j indexes the voxels in the reference patch $\mathcal{N}_p(R(i))$ and $\delta(\cdot)$ indicates if $R(j)$ is segmented. Z is a factor that normalises $\mathbf{n}(i)$ to unit length. This defines $\mathbf{n}(i)$ as a patch-normal representing the most significant direction of the transition “from the segmented to the unsegmented region”.

Together with the calculated patch displacement $\Delta(i)$, the patch-normal $\mathbf{n}(i)$ allows the quantification of over- or under-segmentation by calculating the scalar product $b(i) = \mathbf{n}(i) \cdot \Delta(i)$. If a certain patch is displaced along $\mathbf{n}(i)$ then $b(i) > 0$ indicates over-segmentation. $b(i) < 0$ indicates under-segmentation accordingly. Figure 7.4 illustrates the bias estimation based on the example of comparing a five sided regular polygon to a reference circle.

Furthermore, the weighted mean μ_b and the standard deviation σ_b is calculated to quantify systematic segmentation bias. As weights the $\theta(i)$ are employed to focus the bias calculation on patches that contain edges. The translational bias is estimated for all directions, x, y, z , based on the components of $\Delta(i)$ and patch complexities $(\theta(i)_x, \theta(i)_y, \theta(i)_z)$ calculated based on facets perpendicular to the corresponding directions. Table 7.1 summarises the proposed quantities.

	Calculation	Description
PEIS	Equation 7.6	similarity
μ_b	$\frac{1}{\sum \theta(i)} \sum_i \theta(i) (\mathbf{n}(i) \cdot \Delta(i))$	segmentation bias
σ_b	$\left[\frac{1}{\sum \theta(i)} \sum_i \theta(i) (b(i) - \mu_b)^2 \right]^{\frac{1}{2}}$	shape conservation
$\mu_{d \in \{x, y, z\}}$	$\frac{1}{\sum \theta(i)_d} \sum_i \theta(i)_d \Delta(i)_d$	translational bias (μ)
$\sigma_{d \in \{x, y, z\}}$	$\left[\frac{1}{\sum \theta(i)_d} \sum_i \theta(i)_d (\Delta(i)_d - \mu_d)^2 \right]^{\frac{1}{2}}$	translational bias (σ)

Table 7.1: Overview over all quantities available through PEIS and their description.

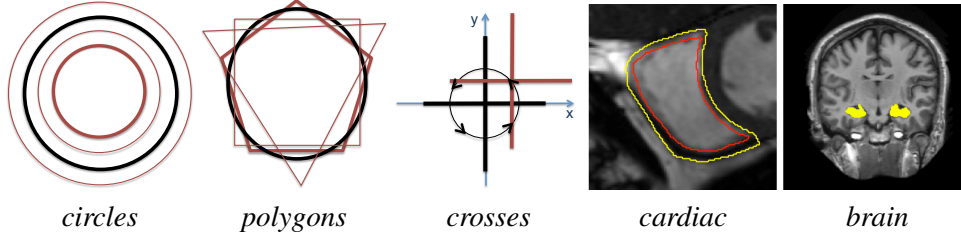


Figure 7.5: Overview over employed datasets. Synthetic datasets with reference segmentation outline (black) and example test segmentation outlines (dark red). Medical imaging datasets with overlaid ground truth reference segmentation. Cardiac MR image with outlined right ventricle myocardium (yellow) and blood pool (red), and a MR brain scan with overlaid hippocampus segmentation (yellow).

7.3 Experiments and results

7.3.1 Materials

The presented measure PEIS was evaluated on both synthetic and real data. Geometric objects were created synthetically to show that PEIS yields, in contrast to the Dice score, *comparable scores* on structures of different size and has the potential to quantify *shape conservation*.

- *Circles*: Two-dimensional circles of varying radiuses ($C_{r_S}^S$) were compared to two reference circles with fixed radius ($C_{r_R}^R$, $r_R = 15$ or $r_R = 80$). In each experiment the radiuses of the test segmentations, r_S , varied from $r_R - 10$ to $r_R + 10$.
- *Polygons*: Two-dimensional regular polygons with a varying number of sides were compared to a reference circle with fixed radius ($r_R = 50$). The polygons approximate the circle and have a fixed area of $A_p = \pi r_R^2$ and thus the same area as the circle. The number of polygon sides varied from 3 to 50.

To evaluate PEIS on vessel-like structures with high surface to volume ratio a dataset of synthetic *crosses* was created. This dataset was used to confirm that PEIS accurately estimates translational bias and compares shape similarity rather than pure overlap.

- *Crosses*: Three-dimensional crosses with varying centre ($c_x = 10\cos(t)$, $c_y = 10\sin(t)$, $c_z = 1$) parameterised by $t \in [0, 2\pi]$ were compared to a reference cross centred at the origin. All crosses are one voxel thick and have a length of 100.

Furthermore, the characteristics of PEIS were investigated on real medical imaging data. Automatic segmentations were compared to available manual ground truth segmentations. On a *cardiac* dataset

(Petitjean et al., 2012) it was shown that PEIS provides *comparable scores* between segmentations of different volume and topology. On a *brain* imaging dataset (Landman and Warfield, 2012) it was shown that PEIS can be used to rank automatic segmentation methods and allows the visualisation of *segmentation bias*. The employed datasets from the medical imaging domain are:

- *Cardiac data*: Automatic segmentations were calculated based on 16 training subjects used in a recent Right Ventricle (RV) Segmentation Challenge (Petitjean et al., 2012, 2015). The Dice score, average surface distance (ASD) and PEIS score were calculated for the myocardium and the blood pool. The reference segmentations were obtained manually by experts (Petitjean et al., 2015).
- *Brain data*: Automatic segmentations were calculated on 20 subjects from the NMM brain atlas used for evaluation in a recent whole-brain segmentation challenge (Landman and Warfield, 2012). The Dice score and PEIS were compared on the hippocampus segmentations calculated by the five best performing methods in the challenge. The reference segmentations were obtained manually by experts (*cf.* Section 2.2.5). The focus was on the hippocampus as its accurate segmentation has been proven to be of high value for the diagnosis and outcome prediction of dementia patients, in AD in particular.

An overview over the datasets is provided in Figure 7.5.

7.3.2 Results on synthetic data

On synthetically generated data, it is observed that PEIS, in contrast to the Dice coefficient, fulfils the desired characteristics: *comparable scores*, *shape conservation*, *boundary conformity*, *segmentation bias* and *similarity*.

In a first experiment the synthetically generated *circles* were used. This experiment, as summarised in Figure 7.6, confirms that the measured segmentation bias μ_b quantifies *boundary conformity* as it represents the true bias for both under- and over-segmentation. The low standard deviation σ_b underlines *shape conservation*. PEIS provides *comparable scores* for different reference radiuses. In contrast to that, Dice scores increase with increasing radiuses. Dice scores are reasonably discriminative only for deviations from a circle with a rather small radius (C_{15}^R), as it tends to degenerate for circles with larger radiuses (C_{80}^R). The vast number of true positives in the

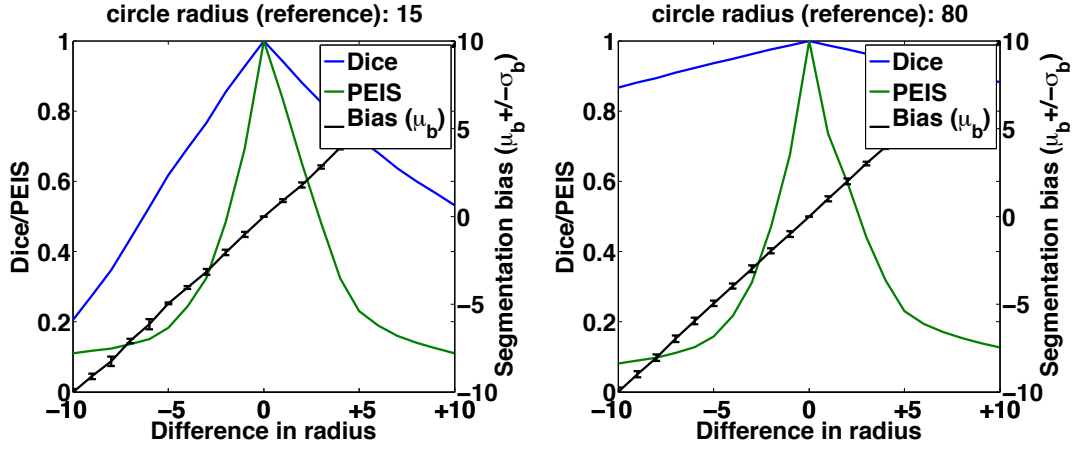


Figure 7.6: *Circles* experiments for reference radiuses $r_R = 15$ (left) and $r_R = 80$ (right): Different similarity scores (colours) and segmentation bias μ_b (black, σ_b as error bars).

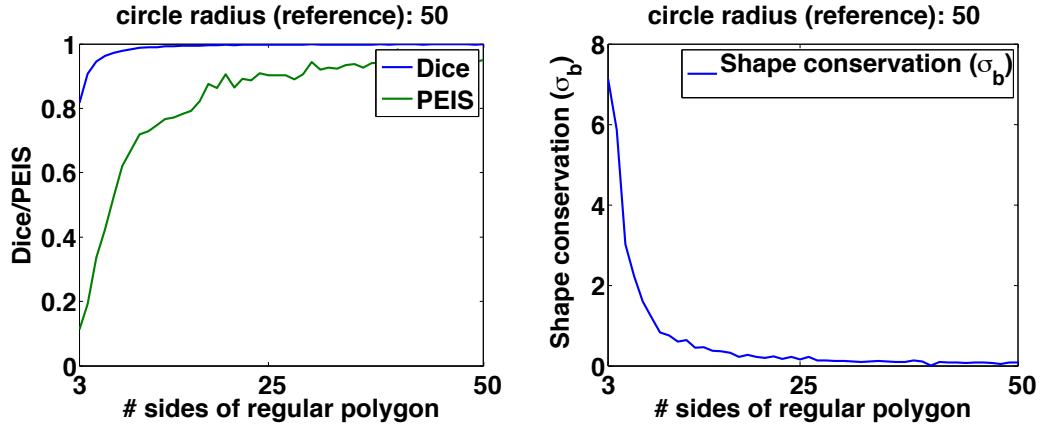


Figure 7.7: *Polygons* experiment: Different similarity scores (left) and estimated shape conservation, σ_b (right).

interior of large structures dominate the overlap measure and the Dice score loses its sensitivity to small deviations from the reference.

In a second experiment, the *polygons* were compared to a reference circle of fixed radius. Figure 7.7 illustrates the calculated similarities. With high Dice scores, *e.g.* 0.91 for the comparison of a square to a circle, the Dice coefficient is unable to detect the significant shape differences between objects of equivalent area/volume. PEIS detects differences over a wide spectrum and is able to quantify the shape differences between segmentations (σ_b). A visualisation of the segmentation bias is illustrated in Figure 7.4, confirming the potential of PEIS to quantify and visualise *segmentation bias*.

In the final experiment on synthetic data the ability of PEIS was evaluated to detect translational

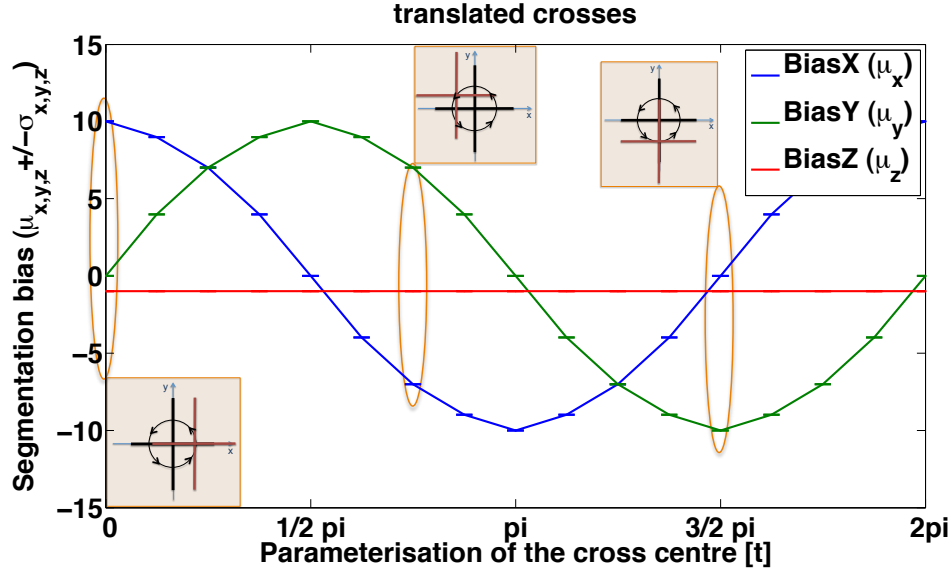


Figure 7.8: *Crosses* experiment: Estimated translational bias for all coordinate directions ($\sigma_{x,y,z}$ as error bars). Crosses were translated by ($c_x = 10\cos(t)$, $c_y = 10\sin(t)$, $c_z = 1$ with $t \in [0, 2\pi]$).

bias in the segmentation: For this, the 3D *crosses* were translated along a circular trajectory. While a Dice score of almost zero, due to neglectable overlap, fails to quantify any similarity, non-zero PEIS scores of 0.21 ± 0.06 (mean $\pm$ SD) confirm that PEIS quantifies similarity rather than pure overlap. PEIS yields *comparable scores* while moving the test cross in a constant distance around the reference cross. As shown in Figure 7.8, PEIS also provides a good estimate for the translational bias (μ_x, μ_y, μ_z) for all coordinate directions with almost zero variance ($\sigma_x = \sigma_y = \sigma_z \approx 0$).

7.3.3 Results on medical imaging data

In this section, PEIS is employed to evaluate segmentations of medical imaging data. Firstly, evidence is obtained that PEIS provides *comparable scores* on *cardiac data*: The measured segmentation accuracy was compared for segmentations of the myocardium, a thin, ring-like structure, and the blood pool, a structure with a rather low surface to volume ratio. Even though both structures share a segmentation boundary (*cf.* Figure 7.5), both the Dice score and the ASD (Heimann et al., 2009) provide similarity scores of significantly different levels. In contrast to this, PEIS remains comparable, as illustrated in Figure 7.9. This is important as it indicates that the proposed measure is indeed independent of the topology and volume of the segmentation under consideration. This result also suggests that PEIS has the potential to compare different segmentation methods over

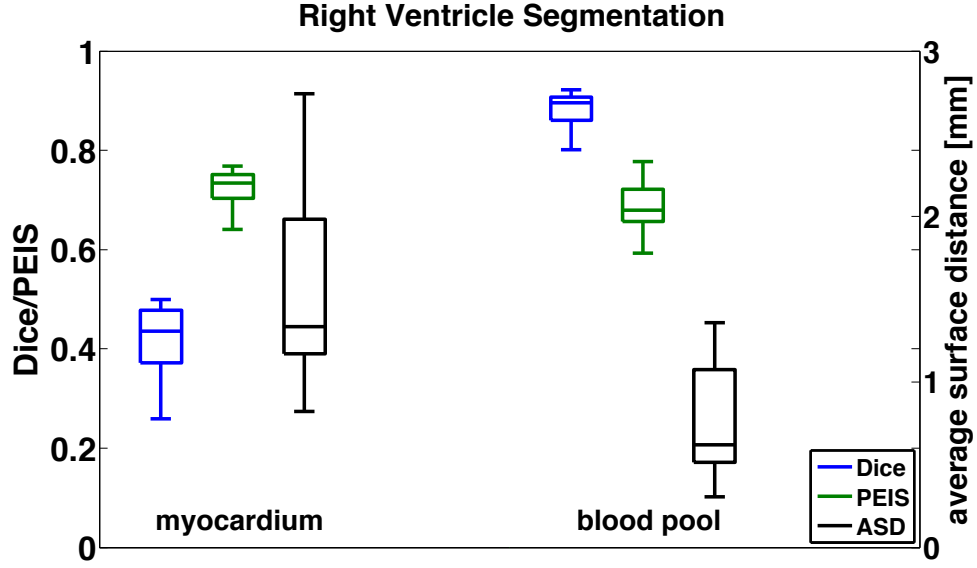


Figure 7.9: Similarity scores calculated on *cardiac data* segmented into myocardium and blood pool. Both the Dice coefficient and the ASD do not yield *comparable scores* for structures of different complexities.

different datasets.

Secondly, PEIS was applied to evaluate the hippocampus segmentations of the five best performing methods in a recent whole-brain segmentation challenge (Landman and Warfield, 2012). The results are summarised in Table 7.2. Here, PEIS produces a similar ranking as the Dice coefficient, which was used in the challenge. More interesting, however, is the additional information provided through PEIS. The best performing method is also the method that best conserves the segmentation shape, lowest σ_b . Furthermore it is possible to obtain the visualisation of the segmentation bias on right hippocampus segmentations of the five methods (Figure 7.10).

These experiments on real data confirm that PEIS allows the quantification and visualisation of *segmentation bias*, yields *comparable scores* and adds information regarding *shape conservation*.

	CIS_JHU	MALP_EM*	PICSL_Joint	NLS	PICSL_BC
Dice	0.851	0.861	0.862	0.866	0.870
PEIS	0.788	0.801	0.799	0.809	0.810
μ_b	-0.007	-0.069	0.011	-0.007	-0.017
σ_b	0.541	0.483	0.492	0.471	0.468

* based on an early version of MALP-EM (Ledig et al., 2012a).

Table 7.2: Similarity scores (Dice coefficient, PEIS) and mean segmentation bias (μ_b) and shape conservation (σ_b) averaged over 20 hippocampus segmentations for different methods.

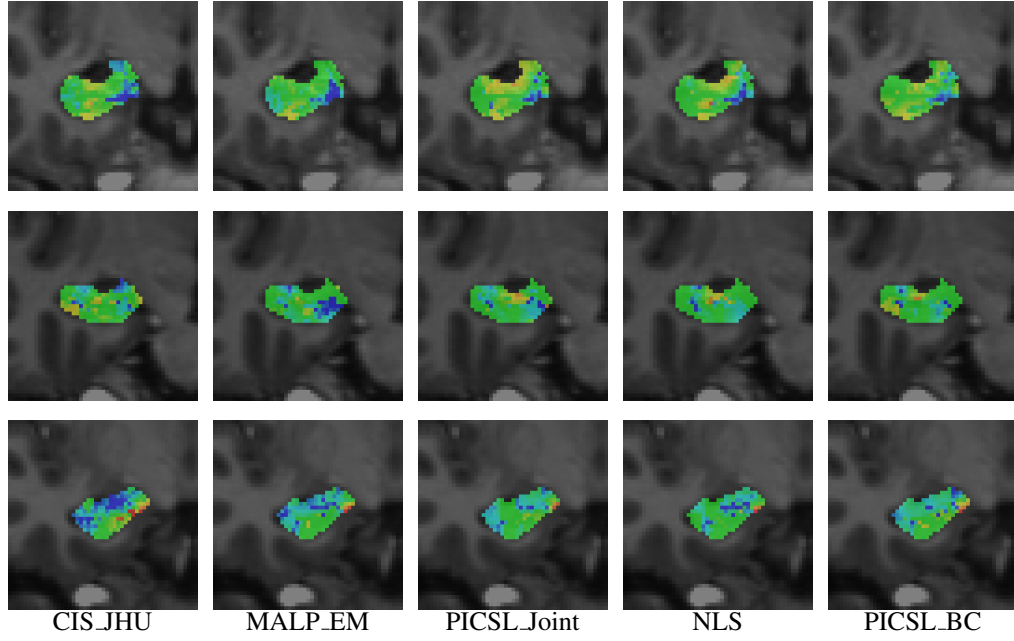


Figure 7.10: Visualisation in coronal view of segmentation bias for the right hippocampus of three random subjects (rows) for five automatic segmentation methods (columns). Colours relate to positive bias or over-segmentation (red) and negative bias or under-segmentation (blue).

7.4 Discussion

It was shown that PEIS yields *comparable scores* on real medical imaging datasets, which sets it apart from many existing approaches such as the Dice coefficient or surface-based distance measures. In the presented formulation, PEIS is not independent of the resolution level of a segmentation as the spatial similarity τ is linked to the voxel-wise patch displacement. *E.g.* in the *circles* experiments (*cf.* Figure 7.6), PEIS will yield lower scores if the data is resampled at a higher resolution. Dependent on the application this may or may not be appropriate. However, the estimated segmentation bias, if calculated in millimetres, is comparable. In general it is difficult to compare or even rank similarity measures as they usually provide complementary information. However, it was shown that PEIS provides a large spectrum of information including *shape conservation* and *bias estimation*. As PEIS finds patches of the reference segmentation in the test segmentation, it provides a directional measure. This differs from the Dice coefficient, which is non-directional. While non-directionality seems a desirable attribute it brings the risk that the measure's scores may be biased towards certain test segmentations. A further discussion of directionality can be found in Tversky (1977). The presented optimisation approach of PEIS is efficient, has no fixed search

window size and identifies a local optimum, which is desired for this application. In the conducted experiments, a patch size of $5 \times 5 \times 5$ voxels was chosen. This is similar to commonly used patch-based segmentation methods, *e.g.* Coupé et al. (2011). In the conducted experiments the runtimes of PEIS and ASD were comparable in terms of computational effort.

7.5 Conclusion and future work

In this chapter the novel approach Patch-based Evaluation of Image Segmentation (PEIS) was introduced for the comparison of binary segmentations. It was demonstrated on both synthetic data and real medical imaging data that PEIS has the potential to fulfil the key characteristics which are important for the evaluation of image segmentation: Providing a similarity score that is both comparable over different applications and informative. The presented measure has been shown to be sensitive to small, but critical shape differences of objects with varying size. Furthermore, PEIS yields a voxel-wise bias estimate which allows the quantification of systematic boundary differences of both shape and translational nature. Especially the possibility to visualise segmentation bias is valuable when comparing segmentations. This can support the development and tuning of automatic segmentation methods. With the incorporation of locally varying weights based on patch complexity, PEIS extends the standard ratio model. This is a step away from pure overlap measures towards measures with a focus on segmentation similarity.

In this work PEIS was evaluated based on the combined quantity $\eta = 0.5(\gamma + \tau)$. However, PEIS can be based on other scales such as the pure segmentation similarity γ or the pure spatial similarity τ . In the future it will be interesting to investigate whether PEIS can be adapted by varying the patch-size to define similarity at different levels of detail. The presented optimisation strategy can potentially be employed in other patch-based applications, such as label fusion (Asman and Landman, 2013; Coupé et al., 2011). It is desirable to extend PEIS to multi-label segmentations. A publicly available implementation has been released (<https://biomedia.doc.ic.ac.uk/software/>, last accessed 17 November 2015). The presented experiments have shown that PEIS is able to recover rigid transformations, which links PEIS to image registration. Potentially not only segmentation bias but also a deformation field relating reference and test segmentation can be recovered.

8 Conclusion

The main contribution of this thesis is the development and implementation of a robust and automatic method (MALP-EM, *cf.* Chapter 3) to segment T1w MR images of the whole brain into its individual anatomical ROIs. While there are numerous established methods that perform this task, MALP-EM has been shown to be particularly accurate (Landman et al., 2012) while being robust in the presence of disease-related atrophy or severe brain deformation due to *e.g.* head trauma (Ledig et al., 2015). Further MALP-EM was extended to be applicable to the analysis of longitudinal image series (MALP-EM4D, *cf.* Chapter 4). In contrast to many other state-of-the-art methods that measure volume change over time of individual structures only, MALP-EM4D allows atrophy measurement of more than a hundred structures simultaneously. During the development of MALP-EM4D particular focus was placed on the avoidance of bias caused by image distortions such as differential bias fields or by asymmetric image processing (*e.g.* registration).

Both MALP-EM and MALP-EM4D were then employed to analyse a large cohort of patients with AD (*cf.* Chapter 5) and a prospectively acquired cohort of patients that sustained a TBI (*cf.* Chapter 6). The methods' potential to support cross-sectional and longitudinal image studies was confirmed on both datasets. Established findings in the AD literature were confirmed such as the potential of volumes of MTL structures to discriminate patients with AD from healthy control subjects. The calculated atrophy rates are comparable to those published in meta-analyses articles and clearly related to the progression of dementia. The ADNI cohort has been extensively studied in the past, which makes it an ideal resource to validate novel techniques. In this thesis, it was shown that a multi-structure analysis of the whole brain has a discriminative power comparable to highly sophisticated machine learning approaches, while allowing for an intuitive interpretation. Clinical interpretability of biomarkers has been a particular goal of this thesis as it can lower the barrier to translate research methodology into clinical practice. The physical meaning of volume or volume change is clear, while the interpretation of features that are derived from more involved

approaches such as sparse representations, manifold learning or patch-correspondences is often less obvious. A strength of the performed study is also the processing of 5074 images acquired from 1674 distinct subjects from ADNI-1/-GO/-2. For each image/subject structural volumes and volume changes of a large number of individual brain structures were quantified. These features can be a valuable resource for future analyses, such as investigating the progression of structural volumes with healthy aging. It was shown that measurements of distinct anatomical structures can provide complementary information that can be exploited in multi-feature classifiers.

In contrast to disease progression in dementia, the disease-related processes that follow a head trauma are studied less extensively. In TBI research it is therefore particularly interesting to analyse processes within the whole brain rather than restricting the focus to selected ROIs only. Quantifying atrophy over the months following the injury allows a better characterisation of the disease progression in TBI. In an analysis of 67 patients with mild to moderate TBI disease-related volume loss of *e.g.* total brain tissue and white matter but also individual structures such as the thalamus or the brain stem was found. Further, the conducted experiments present evidence that reduced hippocampal volume at the acute disease stage might predict poor outcome of a head trauma. The correlation between hippocampal volume at injury and outcome diagnosis needs to be confirmed and further investigated.

Next to the developed segmentation methodology, a novel measurement was proposed to quantify the similarity of image segmentations (PEIS, *cf.* Chapter 7). The PEIS software has been made publicly available. In this context several desirable attributes have been defined that an ideal measure of similarity should possess. The conducted experiments confirmed that PEIS provides a promising framework that has the potential to fulfil these characteristics. When compared to widely used measures such as the Dice coefficient or the average surface distance, PEIS can provide a more complete assessment of similarity with respect to the defined criteria. A particular strength of PEIS is that it allows the visualisation of segmentation bias which can provide valuable information when developing and optimising segmentation methodology.

8.1 Limitations

The presented segmentation methods rely on manually annotated brain atlases of healthy subjects. During the development of both MALP-EM and MALP-EM4D the focus was placed on robustness.

However, in the presence of pathology the registration algorithm can fail to establish an accurate transformation when aligning the atlases to the target image. In these difficult cases an essential assumption in image registration of an existing and smooth one-to-one mapping between the images is often violated. The proposed method was designed to correct obvious labelling errors due to misregistration based on the image intensities. Nevertheless, the framework is not able to explicitly segment pathology that is not represented in the atlases such as lesions, contusions or tumours. This can result in pathologies being included in one or more ROIs (*e.g.* within the WM, Figure 4.5). It can be argued that image segmentations thus need to be reviewed individually to exclude affected ROIs from a further analysis. In this thesis not all segmentation results were explicitly reviewed to ensure that each individual ROI was segmented accurately. Rather, the overall segmentation result was checked and it was investigated whether the whole feature set contains disease-relevant information even though individual measurements might be affected by pathology.

Another limitation originates in the imaging cohorts and the corresponding clinical data used in this thesis. As ADNI-2 is still a continuously growing resource it is difficult to define a standardised evaluation set. In this work some pragmatic choices had been made. Thus the ADNI study subset was defined based on the images that had been available and processed at the time of the analysis. The lists of the processed image identifiers have been published online to enable other researchers to compare their results to those reported here. Another challenge when comparing results across studies are variations in the definition of disease stages, *e.g.* the definition of sMCI/pMCI.

In the TBI analysis substantial age differences between patients of different disease outcome categories were observed. This problem was addressed by defining an ad-hoc rule to exclude subjects so that age differences are no longer significant at the expense of substantially reducing the study cohort. Subsequent studies should investigate and consider the effects of confounding factors such as age, gender, but also smoking or education carefully. Furthermore, the performed volumetric analysis did not account for focal lesions or early oedema that might be present at the acute disease stage. This is problematic as structural volumes measured shortly after the injury event are likely to be distorted by effects due to neuropathology such as brain swelling. Finally, measured changes in structural volume are likely not exclusively attributable to actual modifications of brain anatomy including cell death. Instead, measurements are potentially distorted by changes of pathology such as a reduction of swelling. This is a limitation and needs to be considered when interpreting both cross-sectional biomarkers at the acute time point and longitudinal measurements.

In this thesis the focus was placed on binary classification experiments. This means that patients (samples) were classified into either of two classes presuming that one of both is correct. This is a substantial simplification of realistic clinical scenarios where usually more than two diagnosis (*e.g.* dementia types, disease stages) or multiple outcome scenarios (*e.g.* GOSe scores) are possible. With this in mind the application of a classification framework allowing for multi-class classification is desirable. In dementia diagnosis this could support the differential diagnosis of distinct dementias such as FTD, LBD, VD and AD. In the context of TBI, outcome prediction at a finer level is also of significant interest.

8.2 Future work

In the future the benefits of applying MALP-EM4D to more than two time points should be explored. For example adding information of the m12 image when estimating baseline→m24 atrophy has the potential to reduce measurement bias and thus sample sizes. This also allows the consistent measurement of volumetric changes at several time points leading to a finer temporal resolution of measurements and thus better understanding of disease processes. For example, in another study (Newcombe et al., 2015) a version of MALP-EM4D was employed to process up to five time points and increased WM loss in patients with moderate-to-severe TBI was reported.

In the work towards this thesis, a large set of features quantifying brain anatomy was calculated that can enable research beyond the analyses performed and presented here. For example, the effect of nuisance factors such as age, gender, brain size but also field strength can be further explored using the healthy control subjects from ADNI. Another important field for future research is the differential diagnosis of dementia types. In this context it might be interesting to calculate additional features such as shape descriptors or cortical thickness measurements based on the segmentation estimates.

In addition a WM atlas is currently integrated into both the MALP-EM and MALP-EM4D frameworks to further parcelate the rather large ROIs denoted as 'cerebral white matter' in the NMM atlas. Beyond allowing for a structural analysis of different WM subregions this is a crucial prerequisite for established DWI pipelines. The source code and binaries of both MALP-EM and MALP-EM4D will be made available. In combination with a python script that integrates the complete pipeline, this allows both clinicians and researchers to apply and improve on the work presented in this thesis.

A first version of this updated framework is currently tested at the Cambridge University Hospital, UK. The approximate runtime to process a standard T1w image depends on the computing environment but is estimated at around 10 hours using a single CPU core. The pipeline can be run parallelised so that computation times between 1 and 2 hours are realistic on a desktop computer using 6 cores. This includes all steps relevant for MALP-EM: bias correction, brain extraction, atlas alignment, atlas fusion and label refinement.

The problem of explicitly segmenting pathologies is a further direction of research. Two promising approaches to address this limitation are 1) the targeted segmentation of different pathologies (*e.g.* contusions, lesions) using machine learning (*e.g.* deep learning, RFs) and 2) the detection of outlier voxels that do not match the intensity characteristics of the label. The calculated structural segmentations are then a useful resource to put both pathology or outliers in general into spatial context. Quantifying distinct lesion types and lesion sizes can result in valuable features that allow a deeper investigation of the disease progression in TBI. It is expected that imaging information from multiple MR protocols such as T2w, FLAIR or SWI sequences but also DWI images needs to be incorporated to achieve these goals.

Bibliography

- C. Aguilar, E. Westman, J.-S. Muehlboeck, P. Mecocci, B. Vellas, M. Tsolaki, I. Kloszewska, H. Soininen, S. Lovestone, C. Spenger, A. Simmons, and L.-O. Wahlund. Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment. *Psychiatry Research: Neuroimaging*, 212(2):89–98, 2013.
- 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, and M. W. Weiner. Clinical core of the Alzheimer's disease neuroimaging initiative: Progress and plans. *Alzheimer's & Dementia*, 6(3):239–246, 2010.
- A. Akhondi-Asl and S. K. Warfield. Simultaneous Truth and Performance Level Estimation Through Fusion of Probabilistic Segmentations. *IEEE Transactions on Medical Imaging*, 32(10):1840–1852, 2013.
- H. Akiyama, S. Barger, S. Barnum, B. Bradt, J. Bauer, G. M. Cole, N. R. Cooper, P. Eikelenboom, M. Emmerling, B. L. Fiebich, C. E. Finch, S. Frautschy, W. S. T. Griffin, H. Hampel, M. Hull, G. Landreth, L.-F. Lue, R. Mrazek, I. R. Mackenzie, P. L. McGeer, M. K. O. Banion, J. Pachter, G. Pasinetti, C. P. Salaman, J. Rogers, R. Rydel, Y. Shen, W. Streit, R. Strohmeyer, I. Tooyoma, F. L. van Muiswinkel, R. Veerhuis, D. Walker, S. Webster, B. Wegrzyniak, G. Wenk, and T. Wyss-Coray. Inflammation and Alzheimer's disease. *Neurobiology of Aging*, 21:383–421, 2000.
- 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 & Dementia*, 7(3):270–279, 2011.

- 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.
- A. Alzheimer. Über eine eigenartige Erkrankung der Hirnrinde. *Allgemeine Zeitschrift für Psychiatrie und Psychisch-Gerichtliche Medizin*, 64(11):146–148, 1907.
- S. M. Andersen, S. Z. Rapcsak, and P. M. Beeson. Cost function masking during normalization of brains with focal lesions: Still a necessity? *NeuroImage*, 53(1):78–84, 2010.
- J. Andersson, M. Jenkinson, and S. Smith. Non-linear registration, aka spatial normalisation. *Technical Report TR07JA2, FMRIB Centre, Oxford, UK*, 2007.
- E. D. Angelini, O. Clatz, E. Mandonnet, E. Konukoglu, L. Capelle, and H. Duffau. Glioma Dynamics and Computational Models: A Review of Segmentation, Registration, and In Silico Growth Algorithms and their Clinical Applications. *Current Medical Imaging Reviews*, 3(4):262–276, 2007.
- K. Arlinghaus, A. Shoaib, and T. Price. Neuropsychiatric assessment. In J. M. Silver, T. W. McAllister, and S. C. Yudofsky, editors, *Textbook of Traumatic Brain Injury - 2nd Edition*, pages 55–71. American Psychiatric Publishing, Inc., Washington DC, 2011.
- X. Artaechevarria, A. Munoz Barrutia, and C. d. S. Ortiz. Combination Strategies in Multi-Atlas Image Segmentation: Application to Brain MR Data. *IEEE Transactions on Medical Imaging*, 28(8):1266–1277, 2009.
- J. Ashburner and K. Friston. Multimodal Image Coregistration and Partitioning - A Unified Framework. *NeuroImage*, 6(3):209–217, 1997.
- J. Ashburner and K. Friston. Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, 7(4):254–266, 1999.
- J. Ashburner and K. J. Friston. Voxel-Based Morphometry - The Methods. *NeuroImage*, 11(6):805–821, 2000.
- J. Ashburner and K. J. Friston. Unified segmentation. *NeuroImage*, 26(3):839–851, 2005.

- I. Asikainen, M. Kaste, and S. Sarna. Predicting late outcome for patients with traumatic brain injury referred to a rehabilitation programme: a study of 508 finnish patients 5 years or more after injury. *Brain Injury*, 12(2):95–107, 1998.
- A. Asman, A. Akhondi-Asl, H. Wang, N. J. Tustison, B. B. Avants, S. K. Warfield, and B. A. Landman. MICCAI Challenge Workshop on Segmentation: Algorithms, Theory and Applications (SATA). 2013a. URL https://masi.vuse.vanderbilt.edu/workshop2013/images/1/1b/SATA_2013_Proceedings.pdf.
- A. J. Asman and B. A. Landman. Formulating Spatially Varying Performance in the Statistical Fusion Framework. *IEEE Transactions on Medical Imaging*, 31(6):1326–1336, 2012.
- A. J. Asman and B. A. Landman. Non-local statistical label fusion for multi-atlas segmentation. *Medical Image Analysis*, 17(2):194–208, 2013.
- A. J. Asman, L. B. Chambless, R. C. Thompson, and B. A. Landman. Out-of-atlas likelihood estimation using multi-atlas segmentation. *Medical Physics*, 40(4):043702–01–043702–10, 2013b.
- N. Aspert, D. Santa-Cruz, and T. Ebrahimi. MESH: measuring errors between surfaces using the Hausdorff distance. In *IEEE International Conference on Multimedia and Expo*, volume 1, pages 705–708. 2002.
- B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, 2008.
- M. Bach Cuadra, C. Pollo, A. Bardera, O. Cuisenaire, and J.-P. Thiran. Atlas-Based Segmentation of Pathological Brain {MR} Images using a model of lesion growth. *IEEE Transactions on Medical Imaging*, 23(10):1301–1314, 2004.
- W. Bai, W. Shi, C. Ledig, and D. Rueckert. Multi-atlas segmentation with augmented features for cardiac MR images. *Medical Image Analysis*, 19(1):98–109, 2015.
- R. Barber, A. Gholkar, P. Scheltens, C. Ballard, I. G. McKeith, and J. T. O’Brien. Medial temporal lobe atrophy on MRI in dementia with Lewy bodies. *Neurology*, 52(6):1153, 1999.

- C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing. *ACM Transactions on Graphics*, 28(3):24:1–24:11, 2009a.
- J. Barnes, R. I. Scahill, R. G. Boyes, C. Frost, E. B. Lewis, C. L. Rossor, M. N. Rossor, and N. C. Fox. Differentiating AD from aging using semiautomated measurement of hippocampal atrophy rates. *NeuroImage*, 23(2):574–581, 2004.
- J. Barnes, R. G. Boyes, E. B. Lewis, J. M. Schott, C. Frost, R. I. Scahill, and N. C. Fox. Automatic calculation of hippocampal atrophy rates using a hippocampal template and the boundary shift integral. *Neurobiology of Aging*, 28(11):1657–1663, 2007.
- J. Barnes, J. Foster, R. G. Boyes, T. Pepple, E. K. Moore, J. M. Schott, C. Frost, R. I. Scahill, and N. C. Fox. A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. *NeuroImage*, 40(4):1655–1671, 2008.
- J. Barnes, J. W. Bartlett, L. A. van de Pol, C. T. Loy, R. I. Scahill, C. Frost, P. Thompson, and N. C. Fox. A meta-analysis of hippocampal atrophy rates in Alzheimer’s disease. *Neurobiology of Aging*, 30(11):1711–1723, 2009b.
- J. Barnes, G. R. Ridgway, J. Bartlett, S. M. D. Henley, M. Lehmann, N. Hobbs, M. J. Clarkson, D. G. MacManus, S. Ourselin, and N. C. Fox. Head size, age and gender adjustment in MRI studies: a necessary nuisance? *NeuroImage*, 53(4):1244–1255, 2010.
- S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes. A survey of MRI-based medical image analysis for brain tumor studies. *Physics in Medicine and Biology*, 58(13):R97–R129, 2013.
- H. G. Belanger, R. D. Vanderploeg, G. Curtiss, and D. L. Warden. Recent Neuroimaging Techniques in Mild Traumatic Brain Injury. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 19(1):5–20, 2007.
- H. G. Belanger, T. Kretzmer, R. D. Vanderploeg, and L. M. French. Symptom complaints following combat-related traumatic brain injury: Relationship to traumatic brain injury severity and posttraumatic stress disorder. *Journal of the International Neuropsychological Society*, 16(1):194–199, 2010.

- B. B. Bendlin, M. L. Ries, M. Lazar, A. L. Alexander, R. J. Dempsey, H. A. Rowley, J. E. Sherman, and S. C. Johnson. Longitudinal changes in patients with traumatic brain injury assessed with diffusion-tensor and volumetric imaging. *NeuroImage*, 42(2):503–514, 2008.
- R. R. Benson, S. A. Meda, S. Vasudevan, Z. Kou, K. A. Govindarajan, R. A. Hanks, S. R. Millis, M. Makki, Z. Latif, W. Coplin, J. Meythaler, and E. M. Haacke. Global white matter analysis of diffusion tensor images is predictive of injury severity in traumatic brain injury. *Journal of Neurotrauma*, 24(3):446–59, 2007.
- E. D. Bigler. Quantitative magnetic resonance imaging in traumatic brain injury. *The Journal of head trauma rehabilitation*, 16(2):117–134, 2001.
- E. D. Bigler. Traumatic brain injury, neuroimaging, and neurodegeneration. *Frontiers in Human Neuroscience*, 7(395), 2013.
- D. D. Blatter, E. D. Bigler, S. D. Gale, S. C. Johnson, C. V. Anderson, B. M. Burnett, D. Ryser, S. E. Macnamara, and B. J. Bailey. MR-based brain and cerebrospinal fluid measurement after traumatic brain injury: correlation with neuropsychological outcome. *American Journal of Neuroradiology*, 18(1):1–10, 1997.
- K. Blennow, M. J. de Leon, and H. Zetterberg. Alzheimer’s disease. *The Lancet*, 368(9533):387–403, 2006.
- M. Bobinski, J. Wegiel, H. M. Wisniewski, M. Tarnawski, M. Bobinski, B. Reisberg, M. J. De Leon, and D. C. Miller. Neurofibrillary pathology correlation with hippocampal formation atrophy in Alzheimer disease. *Neurobiology of Aging*, 17(6):909–919, 1996.
- W. Bondareff, C. Q. Mountjoy, M. Roth, and D. L. Hauser. Neurofibrillary degeneration and neuronal loss in Alzheimer’s disease. *Neurobiology of Aging*, 10(6):709–715, 1989.
- L. Bonilha, T. Nesland, C. Rorden, and J. Fridriksson. Asymmetry of the structural brain connectome in healthy older adults. *Frontiers in Psychiatry*, 4(186), 2013.
- V. Bonnelle, T. E. Ham, R. Leech, K. M. Kinnunen, M. A. Mehta, R. J. Greenwood, and D. J. Sharp. Salience network integrity predicts default mode network function after traumatic brain injury. *Proceedings of the National Academy of Sciences*, 109(12):4690–4695, 2012.

- Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004.
- Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.
- H. Braak and E. Braak. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*, 82(4):239–259, 1991.
- L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001.
- M. Brett, A. P. Leff, C. Rorden, and J. Ashburner. Spatial Normalization of Brain Images with Focal Lesions Using Cost Function Masking. *NeuroImage*, 14(2):486–500, 2001.
- K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann. The Balanced Accuracy and Its Posterior Distribution. In *20th International Conference on Pattern Recognition (ICPR)*, pages 3121–3124. 2010.
- E. E. Bron, M. Smits, W. M. van der Flier, H. Vrenken, F. Barkhof, P. Scheltens, J. M. Papma, R. M. E. Steketee, C. Méndez Orellana, R. Meijboom, M. Pinto, J. R. Meireles, C. Garrett, A. J. Bastos-Leite, A. Abdulkadir, O. Ronneberger, N. Amoroso, R. Bellotti, D. Cárdenas-Peña, A. M. Álvarez-Meza, C. V. Dolph, K. M. Iftekharruddin, S. F. Eskildsen, P. Coupé, V. S. Fonov, K. Franke, C. Gaser, C. Ledig, R. Guerrero, T. Tong, K. R. Gray, E. Moradi, J. Tohka, A. Routier, S. Durrleman, A. Sarica, G. Di Fatta, F. Sensi, A. Chincarini, G. M. Smith, Z. V. Stoyanov, L. Sørensen, M. Nielsen, S. Tangaro, P. Inglese, C. Wachinger, M. Reuter, J. C. van Swieten, W. J. Niessen, and S. Klein. Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge. *NeuroImage*, 111:562–579, 2015.
- R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi. Forecasting the global burden of Alzheimer’s disease. *Alzheimer’s & Dementia*, 3(3):186–191, 2007.
- K. R. Brunden, C. Ballatore, A. Crowe, A. B. Smith, V. M.-Y. Lee, and J. Q. Trojanowski. Tau-directed drug discovery for Alzheimer’s disease and related tauopathies: A focus on tau assembly inhibitors. *Experimental Neurology*, 223(2):304–310, 2010.

- A. Buades, B. Coll, and J. Morel. A Review of Image Denoising Algorithms, with a New One. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005.
- R. L. Buckner, D. Head, J. Parker, A. F. Fotenos, D. Marcus, J. C. Morris, and A. Z. Snyder. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *NeuroImage*, 23(2):724–738, 2004.
- A. Büki and J. T. Povlishock. All roads lead to disconnection? - Traumatic axonal injury revisited. *Acta Neurochirurgica*, 148(2):181–194, 2006.
- M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, and M. Bach Cuadra. A review of atlas-based segmentation for magnetic resonance brain images. *Computer Methods and Programs in Biomedicine*, 104(3):e158–e177, 2011.
- R. Cárdenes, R. de Luis-García, and M. Bach Cuadra. A multidimensional segmentation evaluation for medical image data. *Computer methods and programs in biomedicine*, 96(2):108–124, 2009.
- M. J. Cardoso, M. J. Clarkson, G. R. Ridgway, M. Modat, N. C. Fox, and S. Ourselin. LoAd: A locally adaptive cortical segmentation algorithm. *NeuroImage*, 56(3):1386–1397, 2011.
- M. J. Cardoso, R. Wolz, M. Modat, N. C. Fox, D. Rueckert, and S. Ourselin. Geodesic Information Flows. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 7511 of *Lecture Notes in Computer Science*, pages 262–270. 2012.
- M. J. Cardoso, K. Leung, M. Modat, S. Keihaninejad, D. Cash, J. Barnes, N. C. Fox, and S. Ourselin. STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcellation. *Medical Image Analysis*, 17(6): 671–684, 2013a.
- M. J. Cardoso, A. Melbourne, G. S. Kendall, M. Modat, N. J. Robertson, N. Marlow, and S. Ourselin. AdaPT: An adaptive preterm segmentation algorithm for neonatal brain MRI. *NeuroImage*, 65:97–108, 2013b.
- V. Chalana and Y. Kim. A methodology for evaluation of boundary detection algorithms on medical images. *IEEE Transactions on Medical Imaging*, 16(5):642–652, 1997.

- C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.
- H. Chang, A. H. Zhuang, D. J. Valentino, and W. Chu. Performance measure characterization for evaluating neuroimage segmentation algorithms. *NeuroImage*, 47(1):122–135, 2009.
- G. Chételat, B. Landeau, F. Eustache, F. Mézenge, F. Viader, V. de la Sayette, B. Desgranges, and J.-C. Baron. Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: A longitudinal MRI study. *NeuroImage*, 27(4):934–946, 2005.
- A. Chincarini, P. Bosco, P. Calvini, G. Gemme, M. Esposito, C. Olivieri, L. Rei, S. Squarcia, G. Rodriguez, R. Bellotti, P. Cerello, I. De Mitri, A. Retico, and F. Nobili. Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer’s disease. *NeuroImage*, 58(2):469–480, 2011.
- N. Chitphakdithai and J. S. Duncan. Non-rigid Registration with Missing Correspondences in Pre-operative and Postresection Brain Images. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 6361 of *Lecture Notes in Computer Science*, pages 367–374. 2010.
- Y. Cho, J.-K. Seong, Y. Jeong, and S. Y. Shin. Individual subject classification for Alzheimer’s disease based on incremental learning using a spatial frequency representation of cortical thickness data. *NeuroImage*, 59(3):2217–2230, 2012.
- M. Chupin, E. Gérardin, R. Cuingnet, C. Boutet, L. L., S. Lehericy, H. Benali, L. Garnero, and O. Colliot. Fully automatic hippocampus segmentation and classification in Alzheimer’s disease and mild cognitive impairment applied on data from ADNI. *Hippocampus*, 19(6):579–587, 2009a.
- M. Chupin, A. Hammers, R. S. N. Liu, O. Colliot, J. Burdett, E. Bardinet, J. S. Duncan, L. Garnero, and L. Lemieux. Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: Method and validation. *NeuroImage*, 46(3):749–761, 2009b.
- M. Citron. Strategies for disease modification in Alzheimer’s disease. *Nature Reviews Neuroscience*, 5:677–785, 2004.

- J. Cohen. *Statistical Power Analysis for the Behavioral Sciences - Second Edition*. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey, 1988.
- A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multi-modality image registration based on information theory. *International Conference on Information Processing in Medical Imaging (IPMI)*, 3:263–274, 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 Disease, Mild Cognitive Impairment, and Normal Aging by Using Automated Segmentation of the Hippocampus. *Radiology*, 248(1):194–201, 2008.
- O. Commowick, A. Akhondi-Asl, and S. K. Warfield. Estimating A Reference Standard Segmentation With Spatially Varying Performance Parameters: Local MAP STAPLE. *IEEE Transactions on Medical Imaging*, 31(8):1593–1606, 2012.
- C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.
- P. Coupé, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot. An Optimized Blockwise Nonlocal Means Denoising Filter for 3-D Magnetic Resonance Images. *IEEE Transactions on Medical Imaging*, 27(4):425–441, 2008.
- P. Coupé, J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins. Nonlocal Patch-Based Label Fusion for Hippocampus Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 6363 of *Lecture Notes in Computer Science*, pages 129–136. 2010.
- P. Coupé, J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage*, 54(2):940–954, 2011.
- P. Coupé, S. F. Eskildsen, J. V. Manjón, V. S. Fonov, and D. L. Collins. Simultaneous segmentation and grading of anatomical structures for patient’s classification: Application to Alzheimer’s disease. *NeuroImage*, 59(4):3736–3747, 2012a.
- P. Coupé, S. F. Eskildsen, J. V. Manjón, V. S. Fonov, J. C. Pruessner, M. Allard, and D. L. Collins.

- Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease. *NeuroImage: Clinical*, 1(1):141–152, 2012b.
- 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.
- W. R. Crum, O. Camara, and D. L. G. Hill. Generalized Overlap Measures for Evaluation and Validation in Medical Image Analysis. *IEEE Transactions on Medical Imaging*, 25(11):1451–1461, 2006.
- R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehericy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot. 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.
- C. Davatzikos, P. Bhatt, L. M. Shaw, K. N. Batmanghelich, and J. Q. Trojanowski. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiology of Aging*, 32(12):2322.e19–2322.e27, 2011.
- L. De Beaumont, H. Théoret, D. Mongeon, J. Messier, S. Leclerc, S. Tremblay, D. Ellemborg, and M. Lassonde. Brain function decline in healthy retired athletes who sustained their last sports concussion in early adulthood. *Brain*, 132(3):695–708, 2009.
- L. R. Dice. Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3):297–302, 1945.
- 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.
- K. Ding, C. Marquez de la Plata, J. Y. Wang, M. Mumphy, C. Moore, C. Harper, C. J. Madden, R. McColl, A. Whittemore, M. D. Devous, and R. Diaz-Arrastia. Cerebral Atrophy after Traumatic White Matter Injury: Correlation with Acute Neuroimaging and Outcome. *Journal of Neurotrauma*, 25(12):1433–1440, 2008.

- B. Dubois and M. L. Albert. Amnesic MCI or prodromal Alzheimer's disease? *The Lancet Neurology*, 3(4):246–248, 2004.
- B. Dubois, H. H. Feldman, C. Jacova, S. T. DeKosky, P. Barberger-Gateau, J. Cummings, A. Delacourte, D. Galasko, S. Gauthier, G. Jicha, K. Meguro, J. O'Brien, F. Pasquier, P. Robert, M. Rossor, S. Salloway, Y. Stern, P. J. Visser, and P. Scheltens. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *The Lancet Neurology*, 6(8):734–746, 2007.
- J. Eritiaia, S. J. Wood, G. W. Stuart, N. Bridle, P. Dudgeon, P. Maruff, D. Velakoulis, and C. Pantelis. An optimized method for estimating intracranial volume from magnetic resonance images. *Magnetic Resonance in Medicine*, 44(6):973–977, 2000.
- S. F. Eskildsen, P. Coupé, V. Fonov, J. V. Manjón, K. K. Leung, N. Guizard, S. N. Wassef, L. R. Østergaard, and D. L. Collins. BEaST: Brain extraction based on nonlocal segmentation technique. *NeuroImage*, 59(3):2362–2373, 2012.
- S. F. Eskildsen, P. Coupé, D. García-Lorenzo, V. Fonov, J. C. Pruessner, and D. L. Collins. Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. *NeuroImage*, 65(0):511–521, 2013.
- A. C. Evans, D. L. Collins, S. R. Mills, E. D. Brown, R. L. Kelly, and T. M. Peters. 3D statistical neuroanatomical models from 305 MRI volumes. 3:1813–1817, 1993.
- F. Falahati, E. Westman, and A. Simmon. Multivariate Data Analysis and Machine Learning in Alzheimer's Disease with a Focus on Structural Magnetic Resonance Imaging. *Journal of Alzheimer's Disease*, 41(3):685–708, 2014.
- Y. Fan, N. Batmanghelich, C. M. Clark, and C. Davatzikos. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *NeuroImage*, 39(4):1731–1743, 2008.
- C. Fennema-Notestine, I. B. Ozyurt, C. P. Clark, S. Morris, A. Bischoff-Grethe, M. W. Bondi, T. L. Jernigan, B. Fischl, F. Segonne, D. W. Shattuck, R. M. Leahy, D. E. Rex, A. W. Toga, K. H. Zou, and G. G. Brown. Quantitative evaluation of automated skull-stripping methods applied

- to contemporary and legacy images: Effects of diagnosis, bias correction, and slice location. *Human Brain Mapping*, 27(2):99–113, 2006.
- C. Fennema-Notestine, D. J. Hagler, L. K. McEvoy, A. S. Fleisher, E. H. Wu, D. S. Karow, and A. M. Dale. Structural MRI biomarkers for preclinical and mild Alzheimer’s disease. *Human Brain Mapping*, 30(10):3238–3253, 2009.
- C. P. Ferri, M. Prince, C. Brayne, H. Brodaty, L. Fratiglioni, M. Ganguli, K. Hall, K. Hasegawa, H. Hendrie, Y. Huang, A. Jorm, C. Mathers, P. R. Menezes, E. Rimmer, and M. Scazufca. Global prevalence of dementia: a Delphi consensus study. *The Lancet*, 366(9503):2112–2117, 2006.
- B. Fischl and A. M. Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, 97(20):11050–11055, 2000.
- B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale. Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain. *Neuron*, 33(3):341–355, 2002.
- B. Fischl, A. van der Kouwe, C. Destrieux, E. Halgren, F. Ségonne, D. H. Salat, E. Busa, L. J. Seidman, J. Goldstein, D. Kennedy, V. Caviness, N. Makris, B. Rosen, and A. M. Dale. Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex*, 14(1):11–22, 2004.
- 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.
- N. C. Fox, E. K. Warrington, P. A. Freeborough, P. Hartikainen, A. M. Kennedy, J. M. Stevens, and M. N. Rossor. Presymptomatic hippocampal atrophy in Alzheimer’s disease. *Brain*, 119(6):2001–2007, 1996.
- N. C. Fox, S. Cousens, R. Scahill, R. J. Harvey, and M. N. Rossor. Using serial registered brain magnetic resonance imaging to measure disease progression in Alzheimer disease: Power calculations and estimates of sample size to detect treatment effects. *Archives of Neurology*, 57(3):339–344, 2000.

- N. C. Fox, G. R. Ridgway, and J. M. Schott. Algorithms, atrophy and Alzheimer's disease: Cautionary tales for clinical trials. *NeuroImage*, 57(1):15–18, 2011.
- P. Freeborough and N. C. Fox. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Transactions on Medical Imaging*, 16(5):623–629, 1997.
- P. A. Freeborough and N. C. Fox. Modeling Brain Deformations in Alzheimer Disease by Fluid Registration of Serial 3D MR Images. *Journal of Computer Assisted Tomography*, 22(5):838–843, 1998.
- 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.
- Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Computational Learning Theory*, volume 904 of *Lecture Notes in Computer Science*, pages 23–37. Springer Berlin Heidelberg, 1995.
- G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, and P. M. Thompson. The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology*, 6(2):67–77, 2010.
- A. M. Galaburda, J. Corsiglia, G. D. Rosen, and G. F. Sherman. Planum temporale asymmetry, reappraisal since Geschwind and Levitsky. *Neuropsychologia*, 25(6):853–868, 1987.
- S. D. Gale, L. Baxter, N. Roundy, and S. C. Johnson. Traumatic brain injury and grey matter concentration: a preliminary voxel based morphometry study. *Journal of Neurology, Neurosurgery & Psychiatry*, 76(7):984–988, 2005.
- E. Gerardin, G. Chételat, M. Chupin, R. Cuingnet, B. Desgranges, H.-S. Kim, M. Niethammer, B. Dubois, S. Lehericy, L. Garnero, F. Eustache, and O. Colliot. Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. *NeuroImage*, 47(4):1476–1486, 2009.
- S. Gerber, T. Tasdizen, P. T. Fletcher, S. Joshi, and R. Whitaker. Manifold modeling for brain population analysis. *Medical Image Analysis*, 14(5):643–653, 2010. Special Issue on the 12th Interna-

- tional Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2009.
- G. Gerig, M. Jomier, and M. Chakos. Valmet: A New Validation Tool for Assessing and Improving 3D Object Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 2208 of *Lecture Notes in Computer Science*, pages 516–523. 2001.
- T. E. Golde, L. S. Schneider, and E. H. Koo. Anti-A β therapeutics in alzheimer’s disease: The need for a paradigm shift. *Neuron*, 69(2):203–213, 2011.
- A. F. Goldszal, C. Davatzikos, D. L. Pham, M. X. H. Yan, R. N. Bryan, and S. M. Resnick. An image-processing system for qualitative and quantitative volumetric analysis of brain images. *Journal of Computer Assisted Tomography*, 22(5):827–837, 1998.
- 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.
- D. I. Graham, T. K. McIntosh, W. L. Maxwell, and J. A. R. Nicoll. Recent Advances in Neurotrauma. *Journal of neuropathology and experimental neurology*, 59(8):641–651, 2000.
- K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, D. Rueckert, and Alzheimer’s Disease Neuroimaging Initiative. Random forest-based similarity measures for multi-modal classification of Alzheimer’s disease. *NeuroImage*, 65:167–175, 2013.
- D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact Maximum A Posteriori Estimation for Binary Images. *Journal of the Royal Statistical Society. Series B (Methodological)*, 51(2):271–279, 1989.
- W. S. Griffin, J. G. Sheng, M. C. Royston, S. M. Gentleman, J. E. McKenzie, D. I. Graham, G. W. Roberts, and R. E. Mraz. Glial-neuronal interactions in Alzheimer’s disease: the potential role of a ‘cytokine cycle’ in disease progression. *Brain pathology*, 8(1):65–72, 1998.
- R. Guerrero, R. Wolz, A. W. Rao, and D. Rueckert. Manifold population modeling as a neuroimaging biomarker: Application to ADNI and ADNI-GO. *NeuroImage*, 94:275–286, 2014.

- P. Hagmann, L. Jonasson, P. Maeder, J.-P. Thiran, V. J. Wedeen, and R. Meuli. Understanding Diffusion MR Imaging Techniques: From Scalar Diffusion-weighted Imaging to Diffusion Tensor Imaging and Beyond. *RadioGraphics*, 26(Supplement 1):S205–S223, 2006.
- J. Hamm, D. H. Ye, R. Verma, and C. Davatzikos. GRAM: A framework for geodesic registration on anatomical manifolds. *Medical Image Analysis*, 14(5):633–642, 2010. Special Issue on the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2009.
- A. Hammers, R. Allom, M. Koepp, 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.
- A. Hammers, C.-H. Chen, L. Lemieux, R. Allom, S. Vossos, S. L. Free, R. Myers, D. J. Brooks, J. S. Duncan, and M. J. Koepp. Statistical neuroanatomy of the human inferior frontal gyrus and probabilistic atlas in a standard stereotaxic space. *Human Brain Mapping*, 28(1):34–48, 2007.
- X. Han, M. S. Hoogeman, P. C. Levendag, L. S. Hibbard, D. N. Teguh, P. Voet, A. C. Cowen, and T. K. Wolf. Atlas-Based Auto-segmentation of Head and Neck CT Images. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 5242 of *Lecture Notes in Computer Science*, pages 434–441. 2008.
- S. Haneuse, J. Schildcrout, P. Crane, J. Sonnen, B. J., and E. Larson. Adjustment for Selection Bias in Observational Studies with Application to the Analysis of Autopsy Data. *Neuroepidemiology*, 32(3):229–239, 2009.
- 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.
- R. A. Heckemann, S. Keihaninejad, P. Aljabar, D. Rueckert, J. V. Hajnal, and A. Hammers. Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. *NeuroImage*, 51(1):221–227, 2010.
- R. A. Heckemann, S. Keihaninejad, P. Aljabar, K. R. Gray, C. Nielsen, D. Rueckert, J. V. Hajnal,

- A. Hammers, and The AD Neuroimaging Initiative. Automatic morphometry in Alzheimer's disease and mild cognitive impairment. *NeuroImage*, 56(4):2024–2037, 2011.
- R. A. Heckemann, C. Ledig, K. R. Gray, P. Aljabar, D. Rueckert, J. V. Hajnal, and A. Hammers. Brain Extraction Using Label Propagation and Group Agreement: Pincram. *PLoS ONE*, 10(7): e0129211, 2015.
- T. Heimann, B. Van Ginneken, M. A. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, G. Bekes, F. Bello, G. Binnig, H. Bischof, A. Bornik, P. Cashman, Y. Chi, A. Cordova, B. Dawant, M. Fidrich, J. Furst, D. Furukawa, L. Grenacher, J. Hornegger, D. Kainmuller, R. Kitney, H. Kobatake, H. Lamecker, T. Lange, J. Lee, B. Lennon, R. Li, S. Li, H.-P. Meinzer, G. Nemeth, D. Raicu, A.-M. Rau, E. van Rikxoort, M. Rousson, L. Rusko, K. A. Saddi, G. Schmidt, D. Seghers, A. Shimizu, P. Slagmolen, E. Sorantin, G. Soza, R. Susomboon, J. Waite, A. Wimmer, and I. Wolf. Comparison and Evaluation of Methods for Liver Segmentation From CT Datasets. *IEEE Transactions on Medical Imaging*, 28(8):1251–1265, 2009.
- M. Heneka and M. Obanion. Inflammatory processes in Alzheimer's disease. *Journal of Neuroimmunology*, 184(1-2):69–91, 2007.
- F. L. Heppner, R. M. Ransohoff, and B. Becher. Immune attack: the role of inflammation in Alzheimer disease. *Nature Reviews Neuroscience*, 16(6):358–372, 2015.
- C. J. Holmes, R. Hoge, L. Collins, R. Woods, A. W. Toga, and A. C. Evans. Enhancement of MR Images Using Registration for Signal Averaging. *Journal of Computer Assisted Tomography*, 22(2):324–333, 1998.
- X. Hua, A. D. Leow, N. Parikshak, S. Lee, M.-C. Chiang, A. W. Toga, C. R. Jack Jr, M. W. Weiner, and P. M. Thompson. Tensor-based morphometry as a neuroimaging biomarker for Alzheimer's disease: An MRI study of 676 AD, MCI, and normal subjects. *NeuroImage*, 43(3):458–469, 2008.
- D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, 1993.

- A. Irimia, M. C. Chambers, J. R. Alger, M. Filippou, M. W. Prastawa, B. Wang, D. A. Hovda, G. Gerig, A. W. Toga, R. Kikinis, P. M. Vespa, and J. D. Van Horn. Comparison of acute and chronic traumatic brain injury using semi-automatic multimodal segmentation of MR volumes. *Journal of Neurotrauma*, 28(11):2287–2306, 2011.
- A. Irimia, B. Wang, S. R. Aylward, M. W. Prastawa, D. F. Pace, G. Gerig, D. A. Hovda, R. Kikinis, P. M. Vespa, and J. D. Van Horn. Neuroimaging of structural pathology and connectomics in traumatic brain injury: Toward personalized outcome prediction. *NeuroImage: Clinical*, 1(1): 1–17, 2012.
- P. Jaccard. THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE. *New Phytologist*, 11(2):37–50, 1912.
- C. R. Jack Jr., R. C. Petersen, Y. C. Xu, S. C. Waring, P. C. O’Brien, E. G. Tangalos, G. E. Smith, R. J. Ivnik, and E. Kokmen. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer’s disease. *Neurology*, 49(3):786–794, 1997.
- C. R. Jack Jr, D. S. Knopman, W. J. Jagust, L. M. Shaw, P. S. Aisen, M. W. Weiner, R. C. Petersen, and J. Q. Trojanowski. Hypothetical model of dynamic biomarkers of the Alzheimer’s pathological cascade. *The Lancet Neurology*, 9(1):119–128, 2010a.
- C. R. Jack Jr, H. J. Wiste, P. Vemuri, S. D. Weigand, M. L. Senjem, G. Zeng, M. A. Bernstein, J. L. Gunter, V. S. Pankratz, P. S. Aisen, M. W. Weiner, R. C. Petersen, L. M. Shaw, J. Q. Trojanowski, and D. S. Knopman. Brain beta-amyloid measures and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer’s disease. *Brain*, 133(11):3336–3348, 2010b.
- C. R. Jack Jr., M. S. Albert, D. S. Knopman, G. M. McKhann, R. A. Sperling, M. C. Carrillo, B. Thies, and C. H. Phelps. Introduction to the recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease. *Alzheimer’s & Dementia*, 7(3):257–262, 2011.
- B. Jacobs, T. Beems, M. Stulemeijer, A. B. van Vugt, T. M. van der Vliet, G. F. Borm, and P. E. Vos. Outcome Prediction in Mild Traumatic Brain Injury: Age and Clinical Variables Are Stronger Predictors than CT Abnormalities. *Journal of Neurotrauma*, 27(4):655–668, 2010.

- B. Jennett and M. Bond. Assessment of Outcome after severe brain damage: A practical scale. *The Lancet*, 306(7905):480–484, 1975.
- B. Jennett, J. Snoek, M. R. Bond, and N. Brooks. Disability after severe head injury: observations on the use of the Glasgow Outcome Scale. *Journal of Neurology, Neurosurgery, and Psychiatry*, 44:285–293, 1981.
- S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23(Supplement 1):S151–S160, 2004.
- P. Juneja, P. M. Evans, and E. J. Harris. The Validation Index: A new metric for validation of segmentation algorithms using two or more expert outlines with application to radiotherapy planning. *IEEE Transactions on Medical Imaging*, 32(8):1481–1489, 2013.
- G. B. Karas, P. Scheltens, S. A. R. B. Rombouts, P. J. Visser, R. A. van Schijndel, N. C. Fox, and F. Barkhof. Global and local gray matter loss in mild cognitive impairment and Alzheimer’s disease. *NeuroImage*, 23(2):708–716, 2004.
- E. Karran and J. Hardy. A critique of the drug discovery and phase 3 clinical programs targeting the amyloid hypothesis for Alzheimer disease. *Annals of Neurology*, 76(2):185–205, 2014.
- E. Karran, M. Mercken, and B. D. de Strooper. The amyloid cascade hypothesis for Alzheimer’s disease: an appraisal for the development of therapeutics. *Nature Reviews Drug Discovery*, 10(9):698–712, 2011.
- M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.
- V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola. From Local Kernel to Nonlocal Multiple-Model Image Denoising. *International Journal of Computer Vision*, 86(1):1–32, 2010.
- M. J. Kempton, T. S. Underwood, S. Brunton, F. Stylios, A. Schmechtig, U. Ettinger, M. S. Smith, S. Lovestone, W. R. Crum, S. Frangou, S. C. R. Williams, and A. Simmons. A comprehensive testing protocol for MRI neuroanatomical segmentation techniques: Evaluation of a novel lateral ventricle segmentation method. *NeuroImage*, 58(4):1051–1059, 2011.
- C. Kervrann and J. Boulanger. Optimal Spatial Adaptation for Patch-Based Image Denoising. *IEEE Transactions on Image Processing*, 15(10):2866–2878, 2006.

- J. Kim, B. Avants, S. Patel, J. Whyte, B. H. Coslett, J. Pluta, J. A. Detre, and J. C. Gee. Structural consequences of diffuse traumatic brain injury: A large deformation tensor-based morphometry study. *NeuroImage*, 39(3):1014–1026, 2008.
- K. M. Kinnunen, R. Greenwood, J. H. Powell, R. Leech, P. C. Hawkins, V. Bonnelle, M. C. Patel, S. J. Counsell, and D. J. Sharp. White matter damage and cognitive impairment after traumatic brain injury. *Brain*, 134(2):449–463, 2011.
- 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.
- H.-W. Klafki, M. Staufenbiel, J. Kornhuber, and J. Wiltfang. Therapeutic approaches to Alzheimer’s disease. *Brain*, 129(11):2840–2855, 2006.
- A. Klein and J. Tourville. 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in Neuroscience*, 6(171):1–12, 2012.
- A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M.-C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier, J. H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R. P. Woods, J. J. Mann, and R. V. Parsey. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3):786–802, 2009.
- S. Klein, U. A. van der Heide, B. W. Raaymakers, A. N. T. J. Kotte, M. Staring, and J. P. W. Pluim. Segmentation of the prostate in MR images by atlas matching. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, pages 1300–1303, 2007.
- S. Klöppel, C. M. Stonnington, J. Barnes, F. Chen, C. Chu, C. D. Good, I. Mader, L. A. Mitchell, A. C. Patel, C. C. Roberts, N. C. Fox, C. R. Jack, J. Ashburner, and R. S. J. Frackowiak. Accuracy of dementia diagnosis—a direct comparison between radiologists and a computerized method. *Brain*, 131(11):2969–2974, 2008.
- S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack, J. Ashburner, and R. S. J. Frackowiak. Automatic classification of MR scans in Alzheimer’s disease. *Brain*, 131(3):681–689, 2008.
- S. Klöppel, A. Abdulkadir, C. R. Jack Jr., N. Koutsouleris, J. Mourão Miranda, and P. Vemuri. Diagnostic neuroimaging across diseases. *NeuroImage*, 61(2):457–463, 2012.

- H. Koikkalainen, J. and Pölönen, J. Mattila, M. van Gils, H. Soininen, J. Lötjönen, and for the Alzheimer's Disease Neuroimaging Initiative. Improved Classification of Alzheimer's Disease Data via Removal of Nuisance Variability. *PLoS ONE*, 7(2):e31112, 2012.
- J. Koikkalainen, J. Lötjönen, L. Thurfjell, D. Rueckert, G. Waldemar, and H. Soininen. Multi-template tensor-based morphometry: Application to analysis of Alzheimer's disease. *NeuroImage*, 56(3):1134–1144, 2011.
- L. I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. John Wiley & Sons, Inc., 2004.
- B. Landman and S. K. Warfield. MICCAI Grand Challenge and Workshop on Multi-Atlas Labeling. 2012. URL https://masi.vuse.vanderbilt.edu/workshop2012/index.php/Main_Page.
- B. A. Landman, A. J. Asman, A. G. Scoggins, J. A. Bogovic, F. Xing, and J. L. Prince. Robust Statistical Fusion of Image Labels. *IEEE Transactions on Medical Imaging*, 31(2):512–522, 2012.
- J. A. Langlois, W. Rutland-Brown, and K. E. Thomas. Traumatic Brain Injury in the United States: Emergency Department Visits, Hospitalizations, and Deaths. *Atlanta (GA): Centers for Disease Control and Prevention, National Center for Injury Prevention and Control*, 2004.
- C. Ledig, R. A. Heckemann, P. Aljabar, R. Wolz, J. V. Hajnal, A. Hammers, and D. Rueckert. Segmentation of MRI brain scans using MALP-EM. pages 79–82. MICCAI Grand Challenge and Workshop on Multi-Atlas Labeling, 2012a. URL https://masi.vuse.vanderbilt.edu/workshop2012/images/c/c8/MICCAI_2012_Workshop_v2.pdf.
- C. Ledig, R. Wolz, P. Aljabar, J. Lötjönen, R. A. Heckemann, A. Hammers, and D. Rueckert. Multi-class brain segmentation using atlas propagation and EM-based refinement. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, pages 896–899, 2012b.
- C. Ledig, R. A. Heckemann, A. Makropoulos, A. Hammers, J. Lötjönen, D. Menon, and D. Rueckert. Robust whole-brain segmentation: Application to traumatic brain injury. *Medical Image Analysis*, 21(1):40–58, 2015.
- M. Lehtovirta, H. Soininen, M. P. Laakso, K. Partanen, S. Helisalmi, A. Mannermaa, M. Ryynnen, P. H. J. Kuikka, and P. J. Sr Riekkinen. SPECT and MRI analysis in Alzheimer's disease: relation

- to apolipoprotein E epsilon 4 allele. *Journal of Neurology, Neurosurgery, and Psychiatry*, 60: 644–649, 1996.
- K. K. Leung, J. Barnes, G. R. Ridgway, J. W. Bartlett, M. J. Clarkson, K. Macdonald, N. Schuff, N. C. Fox, and S. Ourselin. Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer’s disease. *NeuroImage*, 51(4):1345–1359, 2010a.
- K. K. Leung, M. J. Clarkson, J. W. Bartlett, S. Clegg, C. R. Jack Jr., M. W. Weiner, N. C. Fox, and S. Ourselin. Robust atrophy rate measurement in Alzheimer’s disease using multi-site serial MRI: Tissue-specific intensity normalization and parameter selection. *NeuroImage*, 50(2):516–523, 2010b.
- K. K. Leung, J. Barnes, M. Modat, G. R. Ridgway, J. W. Bartlett, N. C. Fox, and S. Ourselin. Brain MAPS: An automated, accurate and robust brain extraction technique using a template library. *NeuroImage*, 55(3):1091–1108, 2011.
- K. K. Leung, G. R. Ridgway, S. Ourselin, and N. C. Fox. Consistent multi-time-point brain atrophy estimation from the boundary shift integral. *NeuroImage*, 59(4):3995–4005, 2012.
- E. B. Lewis and N. C. Fox. Correction of differential intensity inhomogeneity in longitudinal MR images. *NeuroImage*, 58(23):75–83, 2004.
- L. Liang, C. Liu, Y.-Q. Xu, B. Guo, and H.-Y. Shum. Real-time texture synthesis by patch-based sampling. *Graphics, ACM Transactions on*, 20(3):127–150, July 2001.
- X. Liu, D. Tosun, M. W. Weiner, and N. Schuff. Locally linear embedding (LLE) for MRI based Alzheimer’s disease classification. *NeuroImage*, 83(0):148–157, 2013.
- X. Liu, M. Niethammer, R. Kwitt, M. McCormick, and S. Aylward. Low-Rank to the Rescue - Atlas-Based Analyses in the Presence of Pathologies. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 8675 of *Lecture Notes in Computer Science*, pages 97–104. 2014.
- M. Lorenzo-Valdes, G. I. Sanchez-Ortiz, A. G. Elkington, R. H. Mohiaddin, and D. Rueckert. Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. *Medical Image Analysis*, 8(3):255–265, 2004.

- J. M. Lötjönen, R. Wolz, J. R. Koikkalainen, L. Thurfjell, G. Waldemar, H. Soininen, and D. Rueckert. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *NeuroImage*, 49(3):2352–2365, 2010.
- E. Luders, C. Gaser, K. L. Narr, and A. W. Toga. Why Sex Matters: Brain Size Independent Differences in Gray Matter Distributions between Men and Women. *The Journal of Neuroscience*, 29(45):14265–14270, 2009.
- A. I. R. Maas, E. W. Steyerberg, I. Butcher, R. Dammers, J. Lu, A. Marmarou, N. A. Mushkudiani, G. S. McHugh, and G. D. Murray. Prognostic Value of Computerized Tomography Scan Characteristics in Traumatic Brain Injury: Results from The IMPACT Study. *Journal of Neurotrauma*, 24(2):303–314, 2007.
- A. I. R. Maas, N. Stocchetti, and R. Bullock. Moderate and severe traumatic brain injury in adults. *The Lancet Neurology*, 7(8):728–741, 2008.
- B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Pélégriani-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.
- J. F. Malec, A. W. Brown, C. L. Leibson, J. T. Flaada, J. N. Mandrekar, N. N. Diehl, and P. K. Perkins. The mayo classification system for traumatic brain injury severity. *Journal of Neurotrauma*, 24(9):1417–1424, 2007.
- J. V. Manjón, S. F. Eskildsen, P. Coupé, J. E. Romero, and D. L. Collins. Nonlocal Intracranial Cavity Extraction. *International Journal of Biomedical Imaging*, 2014(Article ID: 820205):11 pages, 2014.
- D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner. Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. *Journal of Cognitive Neuroscience*, 19(9):1498–1507, 2007.
- L. F. Marshall, S. Bowers Marshall, M. R. Klauber, M. van Berkum Clark, H. M. Eisenberg, J. A. Jane, T. G. Luerksen, A. Marmarou, and M. A. Foulkes. A new classification of head injury based on computerized tomography. *Journal of Neurosurgery*, 75(1, Supplement):S14–S20, 1991.

- J. C. Mazziotta, A. W. Toga, A. Evans, P. Fox, and J. Lancaster. A Probabilistic Atlas of the Human Brain: Theory and Rationale for Its Development: The International Consortium for Brain Mapping (ICBM). *NeuroImage*, 2(2, Part A):89–101, 1995.
- T. W. McAllister, L. A. Flashman, B. C. McDonald, and A. J. Saykin. Mechanisms of Working Memory Dysfunction after Mild and Moderate TBI: Evidence from Functional MRI and Neurogenetics. *Journal of Neurotrauma*, 23(10):1450–1467, 2006.
- B. C. McDonald, A. J. Saykin, and T. W. McAllister. Functional MRI of mild traumatic brain injury (mTBI): progress and perspectives from the first decade of studies. *Brain Imaging and Behavior*, 6(2):193–207, 2012.
- I. G. McKeith, D. W. Dickson, J. Lowe, M. Emre, J. T. O’Brien, H. Feldman, and et al. Diagnosis and management of dementia with Lewy bodies: Third report of the DLB consortium. *Neurology*, 65(12):1863–1872, 2007.
- 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.
- G. M. McKhann, M. S. Albert, M. Grossman, B. Miller, D. Dickson, and J. Q. Trojanowski. Clinical and pathological diagnosis of frontotemporal dementia: Report of the work group on frontotemporal dementia and Pick’s disease. *Archives of Neurology*, 58(11):1803–1809, 2001.
- G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack 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 & Dementia*, 7(3):263–269, 2011.
- J. M. Meythaler, J. D. Peduzzi, E. Eleftheriou, and T. A. Novack. Current concepts: Diffuse axonal injury-associated traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, 82(10):1461–1471, 2001.

- C. Misra, Y. Fan, and C. Davatzikos. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: Results from ADNI. *NeuroImage*, 44(4):1415–1422, 2009.
- M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin. Fast free-form deformation using graphics processing units. *Computer Methods and Programs in Biomedicine*, 98(3):278–284, 2010.
- S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett. The Alzheimer’s Disease Neuroimaging Initiative. *Neuroimaging Clinics of North America*, 15(4):869–877, 2005.
- I. Nasrallah and J. Dubroff. An Overview of PET Neuroimaging. *Seminars in Nuclear Medicine*, 43(6):449–461, 2013.
- V. F. J. Newcombe, M. M. Correia, C. Ledig, M. G. Abate, J. G. Outtrim, D. Chatfield, T. Geeraerts, A. E. Manktelow, E. Garyfallidis, J. D. Pickard, B. J. Sahakian, P. J. A. Hutchinson, D. Rueckert, J. P. Coles, G. B. Williams, and D. K. Menon. Dynamic Changes in White Matter Abnormalities Correlate With Late Improvement and Deterioration Following TBI: A Diffusion Tensor Imaging Study. *Neurorehabilitation and Neural Repair*, 2015. doi: 10.1177/1545968315584004.
- M. Niethammer, G. L. Hart, D. F. Pace, P. M. Vespa, A. Irimia, J. D. Van Horn, and S. R. Aylward. Geometric Metamorphosis. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 6892 of *Lecture Notes in Computer Science*, pages 639–646. 2011.
- A. C. Nugent, D. A. Luckenbaugh, S. E. Wood, W. Bogers, C. A. Zarate, and W. C. Drevets. Automated subcortical segmentation using FIRST: Test-retest reliability, interscanner reliability, and comparison to manual segmentation. *Human Brain Mapping*, 34(9):2313–2329, 2013.
- L. G. Nyul and J. K. Udupa. On Standardizing the MR Image Intensity Scale. *Magnetic Resonance in Medicine*, 42(6):1072–1081, 1999.
- S. M. Paling, E. D. Williams, R. Barber, E. J. Burton, W. R. Crum, N. C. Fox, and J. T. O’Brien. The application of serial MRI analysis techniques to the study of cerebral atrophy in late-onset dementia. *Medical Image Analysis*, 8(1):69–79, 2004.

- S. Parisot, H. Duffau, S. Chemouny, and N. Paragios. Joint tumor segmentation and dense deformable registration of brain MR images. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 7511 of *Lecture Notes in Computer Science*, pages 651–658. 2012.
- S. Periaswamy and H. Farid. Medical image registration with partial data. *Medical Image Analysis*, 10(3):452–464, 2006.
- 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.
- C. Petitjean, S. Ruan, D. Grosgeorge, J. Caudron, and J.-N. Dacher. Right ventricle segmentation in cardiac MRI: a MICCAI 2012 challenge. In *MICCAI Right Ventricle Segmentation Challenge Workshop*. 2012. URL <http://www.litislab.eu/rvsc>.
- C. Petitjean, M. A. Zuluaga, W. Bai, J.-N. Dacher, D. Grosgeorge, J. Caudron, S. Ruan, I. B. Ayed, M. J. Cardoso, H.-C. Chen, D. Jimenez-Carretero, M. J. Ledesma-Carbayo, C. Davatzikos, J. Doshi, G. Erus, O. M. O. Maier, C. M. S. Nambakhsh, Y. Ou, S. Ourselin, C.-W. Peng, N. S. Peters, T. M. Peters, M. Rajchl, D. Rueckert, A. Santos, W. Shi, C.-W. Wang, H. Wang, and J. Yuan. Right ventricle segmentation from cardiac MRI: A collation study. *Medical Image Analysis*, 19(1):187–202, 2015.
- E. Pichon, A. Tannenbaum, and R. Kikinis. A statistically based flow for image segmentation. *Medical Image Analysis*, 8(3):267–274, 2004.
- B. L. Plassman, R. J. Havlik, D. C. Steffens, M. J. Helms, T. N. Newman, D. Drosdick, C. Phillips, B. A. Gau, K. A. Welsh-Bohmer, J. R. Burke, J. M. Guralnik, and J. C. S. Breitner. Documented head injury in early adulthood and risk of Alzheimer’s disease and other dementias. *Neurology*, 55(8):1158–1166, 2000.
- M. Prince, R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, and C. P. Ferri. The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimer’s & Dementia*, 9(1):63–75.e2, 2013.

- S. Pszczolkowski, S. Zafeirou, C. Ledig, and D. Rueckert. A Robust Similarity Measure for Non-rigid Image Registration with Outliers. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, pages 568–571, 2014.
- A. Ramani, J. H. Jensen, and J. A. Helpert. Quantitative MR Imaging in Alzheimer Disease. *Radiology*, 241(1):26–44, 2006.
- A. F. Ramlackhansingh, D. J. Brooks, R. J. Greenwood, S. K. Bose, F. E. Turkheimer, K. M. Kinnunen, S. Gentleman, R. A. Heckemann, K. Gunanayagam, G. Gelosa, and D. J. Sharp. Inflammation after trauma: Microglial activation and traumatic brain injury. *Annals of Neurology*, 70(3):374–383, 2011.
- A. Rao, C. Ledig, V. Newcombe, D. Menon, and D. Rueckert. Contusion Segmentation from subjects with traumatic brain injury: a random forest framework. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, pages 333–336, 2014.
- C. Restif. Revisiting the Evaluation of Segmentation Results: Introducing Confidence Maps. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 4792 of *Lecture Notes in Computer Science*, pages 588–595. 2007.
- M. Reuter, N. J. Schmansky, H. D. Rosas, and B. Fischl. Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage*, 61(4):1402–1418, 2012.
- A. Ribbens, J. Hermans, F. Maes, D. Vandermeulen, and P. Suetens. Unsupervised Segmentation, Clustering, and Groupwise Registration of Heterogeneous Populations of Brain MR Images. *IEEE Transactions on Medical Imaging*, 33(2):201–224, 2014.
- E. D. Roberson and L. Mucke. 100 Years and Counting: Prospects for Defeating Alzheimer’s Disease. *Science*, 314(5800):781–784, 2006.
- A. Roche, G. Malandain, and N. Ayache. Unifying Maximum Likelihood Approaches in Medical Image Registration. Technical Report RR-3741, 1999.
- T. Rohlfing, R. Brandt, R. Menzel, and C. R. 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.

- G. C. Roman, T. K. Tatemichi, T. Erkinjuntti, J. L. Cummings, M. J. C., J. H. Garcia, and et al. Vascular dementia Diagnostic criteria for research studies: Report of the NINDSAIREN International Workshop. *Neurology*, 43(2):250–260, 1993.
- F. Rousseau, P. A. Habas, and C. Studholme. A Supervised Patch-Based Approach for Human Brain Labeling. *IEEE Transactions on Medical Imaging*, 30(10):1852–1862, 2011.
- D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, and D. J. Leach, M. O. and Hawkes. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.
- L. Ryan and G. E. Alexander. Neuroimaging: Overview of Methods and Applications. In L. Luecken and L. Gallo, editors, *Handbook of Physiological Research Methods in Health Psychology*, pages 371–395. Elsevier, 2007.
- M. R. Sabuncu, B. T. T. Yeo, K. Van Leemput, B. Fischl, and P. Golland. A Generative Model for Image Segmentation Based on Label Fusion. *IEEE Transactions on Medical Imaging*, 29(10):1714–1729, 2010.
- C. H. Salmond, D. A. Chatfield, D. K. Menon, J. D. Pickard, and B. J. Sahakian. Cognitive sequelae of head injury: involvement of basal forebrain and associated structures. *Brain*, 128(1):189–200, 2004.
- S. Sandor and R. Leahy. Surface-based labeling of cortical anatomy using a deformable atlas. *IEEE Transactions on Medical Imaging*, 16(1):41–54, 1997.
- R. I. Scahill, J. M. Schott, J. M. Stevens, M. N. Rossor, and N. C. Fox. Mapping the evolution of regional atrophy in Alzheimer’s disease: Unbiased analysis of fluid-registered serial MRI. *Proceedings of the National Academy of Sciences*, 99(7):4703–4707, 2002.
- P. Scheltens, N. C. Fox, F. Barkhof, and C. De Carli. Structural magnetic resonance imaging in the practical assessment of dementia: beyond exclusion. *The Lancet Neurology*, 1(1):13–21, 2002.
- M. Schoenberger, J. Ponsford, D. Reutens, R. Beare, and R. O’Sullivan. The relationship between age, injury severity, and mri findings after traumatic brain injury. *Journal of Neurotrauma*, 26(12):2157–2167, 2009.

- F. Ségonne, A. M. Dale, E. Busa, M. Glessner, D. Salat, H. K. Hahn, and B. Fischl. A hybrid approach to the skull stripping problem in MRI. *NeuroImage*, 22(3):1060–1075, 2004.
- D. J. Sharp and T. E. Ham. Investigating white matter injury after mild traumatic brain injury. *Current Opinion in Neurology*, 24(6):558–563, 2011.
- D. J. Sharp, C. F. Beckmann, R. Greenwood, K. M. Kinnunen, V. Bonnelle, X. De Boissezon, J. H. Powell, S. J. Counsell, M. C. Patel, and R. Leech. Default mode network functional and structural connectivity after traumatic brain injury. *Brain*, 134(8):2233–2247, 2011.
- D. J. Sharp, G. Scott, and R. Leech. Network dysfunction after traumatic brain injury. *Nature Reviews Neurology*, 10(3):156–166, 2014.
- D. W. Shattuck, M. Mirza, V. Adisetiyo, C. Hojatkashani, G. Salamon, K. L. Narr, R. A. Poldrack, R. M. Bilder, and A. W. Toga. Construction of a 3D probabilistic atlas of human cortical structures. *NeuroImage*, 39(3):1064–1080, 2008.
- D. Shen, S. Moffat, S. M. Resnick, and C. Davatzikos. Measuring Size and Shape of the Hippocampus in MR Images Using a Deformable Shape Model. *NeuroImage*, 15(2):422–434, 2002.
- M. E. Shenton, H. M. Hamoda, J. S. Schneiderman, S. Bouix, O. Pasternak, Y. Rathi, M.-A. Vu, M. P. Purohit, K. Helmer, I. Koerte, A. P. Lin, C.-F. Westin, R. Kikinis, M. Kubicki, R. A. Stern, and R. Zafonte. A review of magnetic resonance imaging and diffusion tensor imaging findings in mild traumatic brain injury. *Brain Imaging and Behavior*, 6(2):137–192, 2012.
- M. Sherer, M. A. Struchen, S. A. Yablon, Y. Wang, and T. G. Nick. Comparison of indices of traumatic brain injury severity: Glasgow coma scale, length of coma and post-traumatic amnesia. *Journal of Neurology, Neurosurgery & Psychiatry*, 79(6):678–685, 2008.
- W. Shi, H. Lombaert, W. Bai, C. Ledig, X. Zhuang, A. Marvao, T. Dawes, D. O’Regan, and D. Rueckert. Multi-atlas Spectral PatchMatch: Application to cardiac image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 8673 of *Lecture Notes in Computer Science*, pages 348–355. 2014.
- S. Shively, A. I. Scher, D. P. Perl, and R. Diaz-Arrastia. Dementia resulting from traumatic brain injury: What is the pathology? *Archives of Neurology*, 69(10):1245–1251, 2012.

- P. E. Shrout and J. L. Fleiss. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86:420–428, 1979.
- A. Sidaros, A. Skimminge, M. G. Liptrot, K. Sidaros, A. W. Engberg, M. Herning, O. B. Paulson, T. L. Jernigan, and E. Rostrup. Long-term global and regional brain volume changes following severe traumatic brain injury: A longitudinal study with clinical correlates. *NeuroImage*, 44(1): 1–8, 2009.
- C. Smith. Neuropathology. In J. M. Silver, T. W. Mc Allister, and S. C. Yudofsky, editors, *Textbook of Traumatic Brain Injury - 2nd Edition*, pages 23–35. American Psychiatric Publishing, Inc., Washington DC, 2011.
- C. D. Smith, H. Chebrolu, D. R. Wekstein, F. A. Schmitt, and W. R. Markesbery. Age and gender effects on human brain anatomy: A voxel-based morphometric study in healthy elderly. *Neurobiology of Aging*, 28(7):1075–1087, 2007.
- D. H. Smith, V. E. Johnson, and W. Stewart. Chronic neuropathologies of single and repetitive TBI: substrates of dementia? *Nature Reviews Neurology*, 9(4):211–221, 2013.
- S. M. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–155, 2002.
- Z. Song, N. Tustison, B. Avants, and J. C. Gee. Integrated Graph Cuts for Brain MRI Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 4191 of *Lecture Notes in Computer Science*, pages 831–838. 2006.
- A. Sotiras, C. Davatzikos, and N. Paragios. Deformable Medical Image Registration: A Survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190, 2013.
- 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 & Dementia*, 7(3):280–292, 2011.
- R. Stefanescu, O. Commowick, G. Malandain, P.-Y. Bondiau, N. Ayache, and X. Pennec. Non-rigid Atlas to Subject Registration with Pathologies for Conformal Brain Radiotherapy. In *Medical*

- Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 3216 of *Lecture Notes in Computer Science*, pages 704–711. 2004.
- G. E. Strangman, T. M. O’Neil-Pirozzi, C. Supelana, R. Goldstein, D. I. Katz, and M. B. Glenn. Regional brain morphometry predicts memory rehabilitation outcome after traumatic brain injury. *Frontiers in Human Neuroscience*, 4:182, 2010.
- 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.
- G. M. Sullivan and R. Feinn. Using Effect Size or Why the P Value Is Not Enough. *Journal of Graduate Medical Education*, 4(3):279–282, 2012.
- C. A. Szekely, J. E. Thorne, P. P. Zandi, M. Ek, E. Messias, J. C. S. Breitner, and S. N. Goodman. Nonsteroidal Anti-Inflammatory Drugs for the Prevention of Alzheimer’s Disease: A Systematic Review. *Neuroepidemiology*, 23(4):159–169, 2004.
- V.-T. Ta, R. Giraud, D. L. Collins, and P. Coupé. Optimized PatchMatch for Near Real Time and Accurate Label Fusion. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 8675 of *Lecture Notes in Computer Science*, pages 398–406. 2014.
- K. H. Taber, D. L. Warden, and R. A. Hurley. Blast-Related Traumatic Brain Injury: What Is Known? *The Journal of Neuropsychiatry and Clinical Neurosciences*, 18(2):141–145, 2006.
- J. Talairach and P. Tournoux. *Co-planar Stereotaxic Atlas of the Human Brain: 3-dimensional Proportional System: an Approach to Cerebral Imaging*. Thieme Publishers Series. Thieme, 1988.
- G. Teasdale and B. Jennett. Assessment of coma and impaired consciousness: a practical scale. *The Lancet*, 2(7872):81–84, 1974.
- S. J. Teipel, C. Born, M. Ewers, A. L. W. Bokde, M. F. Reiser, H.-J. Möller, and H. Hampel. Multivariate deformation-based analysis of brain atrophy to predict Alzheimer’s disease in mild cognitive impairment. *NeuroImage*, 38(1):13–24, 2007.
- J.-P. Thirion. Image matching as a diffusion process: an analogy with Maxwell’s demons. *Medical Image Analysis*, 2(3):243–260, 1998.

- S. Thornhill, G. M. Teasdale, G. D. Murray, J. McEwen, C. W. Roy, and K. I. Penny. Disability in young people and adults one year after head injury: prospective cohort study. *BMJ*, 320(7250):1631–1635, 2000.
- T. Tong, R. Wolz, P. Coupé, J. V. Hajnal, and D. Rueckert. Segmentation of MR images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling. *NeuroImage*, 76(0):11–23, 2013.
- T. Tong, R. Wolz, Q. Gao, R. Guerrero, J. V. Hajnal, and D. Rueckert. Multiple instance learning for classification of dementia in brain MRI. *Medical Image Analysis*, 18(5):808–818, 2014.
- N. Tustison, B. Avants, P. Cook, Y. Zheng, A. Egan, P. Yushkevich, and J. Gee. N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320, 2010.
- A. Tversky. Features of Similarity. *Psychological Review*, 84(4):327–352, 1977.
- N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage*, 15(1):273–289, 2002.
- J. K. Udupa, V. R. LeBlanc, Y. Zhuge, C. Imielinska, H. Schmidt, L. M. Currie, B. E. Hirsch, and J. Woodburn. A framework for evaluating image segmentation algorithms. *Computerized Medical Imaging and Graphics*, 30(2):75–87, 2006.
- R. Unnikrishnan, C. Pantofaru, and M. Hebert. Toward Objective Evaluation of Image Segmentation Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):929–944, 2007.
- F. van der Lijn, T. den Heijer, M. M. B. Breteler, and W. J. Niessen. Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. *NeuroImage*, 43(4):708–720, 2008.
- K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18(10):897–908, 1999.

- 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.
- T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1, Supplement 1):S61–S72, 2009.
- P. Viola and W. M. Wells III. Alignment by Maximization of Mutual Information. *International Journal of Computer Vision*, 24(2):137–154, 1997.
- H. Vojodi, A. Fakhari, and A. M. Eftekhari Moghadam. A new evaluation measure for color image segmentation based on genetic programming approach. *Image and Vision Computing*, 31(11):877–886, 2013.
- U. Vovk, F. Pernuš, and B. Likar. A Review of Methods for Correction of Intensity Inhomogeneity in MRI. *IEEE Transactions on Medical Imaging*, 26(3):405–421, 2007.
- B. Wang, M. Prastawa, S. P. Awate, A. Irimia, M. C. Chambers, P. M. Vespa, J. D. van Horn, and G. Gerig. Segmentation of serial MRI of TBI patients using personalized atlas construction and topological change estimation. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, pages 1152–1155, 2012a.
- B. Wang, M. Prastawa, A. Irimia, M. C. Chambers, P. M. Vespa, J. D. Van Horn, and G. Gerig. A patient-specific segmentation framework for longitudinal MR images of traumatic brain injury. *Proceedings of SPIE 8314, Medical Imaging 2012*, pages 831402–831402–7, 2012b.
- B. Wang, M. Prastawa, A. Irimia, M. C. Chambers, N. Sadeghi, P. M. Vespa, and G. Gerig. Analyzing Imaging Biomarkers for Traumatic Brain Injury using 4D modeling of longitudinal MRI. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, pages 1392–1395, 2013a.
- H. Wang, S. R. Das, J. W. Suh, M. Altinay, J. Pluta, C. Craige, B. B. Avants, and P. A. Yushkevich. A Learning-Based Wrapper Method to Correct Systematic Errors in Automatic Image Segmentation: Consistently Improved Performance in Hippocampus, Cortex and Brain. *NeuroImage*, 55(3):968–985, 2011a.

- H. Wang, J. W. Suh, S. R. Das, J. Pluta, C. Craige, and P. A. Yushkevich. Multi-Atlas Segmentation with Joint Label Fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3):611–623, 2013b.
- L. Wang, J. S. Swank, I. E. Glick, M. H. Gado, M. I. Miller, J. C. Morris, and J. G. Csernansky. Changes in hippocampal volume and shape across time distinguish dementia of the Alzheimer type from healthy aging. *NeuroImage*, 20(2):667–682, 2003.
- L. Wang, S. Feng, P. Yap, J. H. Gilmore, W. Lin, and D. Shen. Accurate and Consistent 4D Segmentation of Serial Infant Brain MR Images. In *Multimodal Brain Image Analysis (MBIA)*, volume 7012, pages 93–101. Lecture Notes in Computer Science, 2011b.
- S. K. Warfield, K. H. Zou, and W. M. Wells. 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.
- M. A. Warner, C. M. de la Plata, J. Spence, J. Y. Wang, C. Harper, C. Moore, M. Devous, and R. Diaz-Arrastia. Assessing spatial relationships between axonal integrity, regional brain volumes, and neuropsychological outcomes after traumatic axonal injury. *Journal of Neurotrauma*, 27(12):2121–2130, 2010a.
- M. A. Warner, T. S. Youn, T. Davis, A. Chandra, M. C. de la Plata, C. Moore, C. Harper, C. J. Madden, J. Spence, R. McColl, M. Devous, R. D. King, and R. Diaz-Arrastia. Regionally selective atrophy after traumatic axonal injury. *Archives of Neurology*, 67(10):1336–1344, 2010b.
- C.-Y. Wee, P.-T. Yap, D. Shen, and for the Alzheimer’s Disease Neuroimaging Initiative. Prediction of Alzheimer’s disease and mild cognitive impairment using cortical morphological patterns. *Human Brain Mapping*, 34(12):3411–3425, 2013.
- W. M. Wells III, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz. Adaptive Segmentation of MRI Data. *IEEE Transactions on Medical Imaging*, 15(4):429–442, 1996.
- G. Wenk. Neuropathologic changes in Alzheimer’s disease. *Journal of Clinical Psychiatry*, 64 (Supplement 9):7–10, 2003.
- M. J. West, P. D. Coleman, D. G. Flood, and J. C. Troncoso. Differences in the pattern of hippocam-

- pal neuronal loss in normal ageing and Alzheimer's disease. *The Lancet*, 344(8925):769–772, 1994.
- E. Westman, C. Aguilar, J.-S. Muehlboeck, and A. Simmons. Regional Magnetic Resonance Imaging Measures for Multivariate Analysis in Alzheimer's Disease and Mild Cognitive Impairment. *Brain Topography*, 26(1):9–23, 2013.
- R. Wolz, P. Aljabar, D. Rueckert, R. A. Heckemann, and A. Hammers. Segmentation of subcortical structures and the hippocampus in brain MRI using graph-cuts and subject-specific a-priori information. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, pages 470–473, 2009.
- R. Wolz, P. Aljabar, J. V. Hajnal, A. Hammers, and D. Rueckert. LEAP: Learning embeddings for atlas propagation. *NeuroImage*, 49(2):1316–1325, 2010a.
- R. Wolz, R. A. Heckemann, P. Aljabar, J. V. Hajnal, A. Hammers, J. Lötjönen, and D. Rueckert. Measurement of hippocampal atrophy using 4D graph-cut segmentation: Application to ADNI. *NeuroImage*, 52(1):109–118, 2010b.
- R. Wolz, V. Julkunen, J. Koikkalainen, E. Niskanen, D. P. Zhang, D. Rueckert, H. Soininen, J. Lötjönen, and the Alzheimer's Disease Neuroimaging Initiative. Multi-Method Analysis of MRI Images in Early Diagnostics of Alzheimer's Disease. *PLoS ONE*, 6(10):e25446, 10 2011.
- A. Worth and J. Tourville. An ever-improving model of the structure of the living human. *Proceedings of 10th Annual World Congress of Society for Brain Mapping and Therapeutics (SBMT)*, page 20, 2013. URL <http://www.worldbrainmapping.org/docs/2013-Oral%20Poster%20Abstracts.pdf>.
- M. Wu, C. Rosano, P. Lopez-Garcia, C. S. Carter, and H. J. Aizenstein. Optimum template selection for atlas-based segmentation. *NeuroImage*, 34(4):1612–1618, 2007.
- L. Xu, A. Krzyzak, and C. Y. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 22(3):418–435, 1992.
- Z. Xue, D. Shen, and C. Davatzikos. CLASSIC: consistent longitudinal alignment and segmentation for serial image computing. *NeuroImage*, 30(2):388–399, 2006.

- W. A. Yasnoff, J. K. Mui, and J. W. Bacus. Error measures for scene segmentation. *Pattern Recognition*, 9(4):217–231, 1977.
- P. A. Yushkevich, B. B. Avants, S. R. Das, J. Pluta, M. Altinay, and C. Craige. Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: An illustration in ADNI 3T MRI data. *NeuroImage*, 50(2):434–445, 2010.
- E. I. Zacharaki, C. S. Hoge, D. Shen, G. Biros, and C. Davatzikos. Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth. *NeuroImage*, 46(3):762–774, 2009.
- H. Zhang, J. E. Fritts, and S. A. Goldman. Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding*, 110(2):260–280, 2008.
- J. Zhang. The Mean Field Theory in EM Procedures for Markov Random Fields. *IEEE Transactions on Signal Processing*, 40(10):2570–2583, 1992.
- Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57, 2001.
- Y. J. Zhang. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335–1346, 1996.
- B. Zitová and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, 2003.

Appendix

A Publications

A.1 Journal publications

- R. Heckemann, **C. Ledig**, K. R. Gray, P. Aljabar, D. Rueckert, J. V. Hajnal, and A. Hammers, “Brain extraction using label propagation and group agreement: pincram”, PLoS ONE, vol. 10, no. 7, pp. e0129211, 2015.
- V. F. J. Newcombe, M. M. Correia, **C. Ledig**, M. G. Abate, J. G. Outtrim, D. Chatfield, T. Geeraerts, A. E. Manktelow, E. Garyfallidis, J. D. Pickard, B. J. Sahakian, P. J. A. Hutchinson, D. Rueckert and J. P. Coles, G. B. Williams and D. K. Menon, “Dynamic Changes in White Matter Abnormalities Correlate With Late Improvement and Deterioration Following TBI: A Diffusion Tensor Imaging Study”, Neurorehabilitation and Neural Repair, doi: 10.1177/1545968315584004, 2015.
- E. E. Bron, M. Smits, W. M. van der Flier, H. Vrenken, F. Barkhof, P. Scheltens, J. M. Papma, R. M. E. Steketee, C. Méndez Orellana, R. Meijboom, M. Pinto, J. R. Meireles, C. Garrett, A. J. Bastos-Leite, A. Abdulkadir, O. Ronneberger, N. Amoroso, R. Bellotti, D. Cárdenas-Peña, A. M. Álvarez-Meza, C. V. Dolph, K. M. Iftekharuddin, S. F. Eskildsen, P. Coupé, V. S. Fonov, K. Franke, C. Gaser, **C. Ledig**, R. Guerrero, T. Tong, K. R. Gray, E. Moradi, J. Tohka, A. Routier, S. Durrleman, A. Sarica, G. Di Fatta, F. Sensi, A. Chincarini, G. M. Smith, Z. V. Stoyanov, L. Sørensen, M. Nielsen, S. Tangaro, P. Inglese, C. Wachinger, M. Reuter, J. C. van Swieten, W. J. Niessen and S. Klein, “Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge”, Neuroimage, vol. 111, pp. 562-579, 2015.
- **C. Ledig**, R. A. Heckemann, A. Hammers, J. C. Lopez, V. F. J. Newcombe, A. Makropoulos,

- J. Loejoenen, D. Menon and D. Rueckert, “Robust whole-brain segmentation: application to traumatic brain injury”, *Medical Image Analysis*, vol. 21, no. 1, pp. 40-58, 2015.
- W. Bai, W. Shi, **C. Ledig**, and D. Rueckert, “Multi-Atlas Segmentation with Augmented Features for Cardiac MR Images”, *Medical Image Analysis*, vol. 19, no. 1, pp. 98-109, 2015.
 - R. Wright, V. Kyriakopoulou, **C. Ledig**, M. Rutherford, J. V. Hajnal, D. Rueckert, and P. Aljabar, “Automatic quantification of normal cortical folding patterns from foetal brain MRI”, *NeuroImage*, vol. 91, pp. 21-32, 2014.
 - A. Makropoulos, I. S. Gousias, **C. Ledig**, P. Aljabar, A. Serag, J. V. Hajnal, D. A. Edwards, S. J. Counsell, and D. Rueckert, “Automatic Whole Brain MRI Segmentation of the Developing Neonatal Brain”, *IEEE Transactions on Medical Imaging*, vol. 33, no. 9, pp. 1818-1831, 2014.
 - F. Deligianni, G. Varoquaux, B. Thirion, D. J. Sharp, **C. Ledig**, R. Leech, and D. Rueckert, “A framework for inter-subject prediction of functional connectivity from structural networks”, *IEEE Transactions on Medical Imaging*, vol. 32, no. 12, pp. 2200-2214, 2013.

A.2 Conference publications

- A. Schmidt-Richberg, R. Guerrero, **C. Ledig**, H. Molina-Abril, A. F. Frangi and D. Rueckert, “Multi-stage biomarker models for progression estimation in Alzheimer’s disease”, *Information Processing in Medical Imaging (IPMI)*, vol. 9123, pp. 387-398, 2015.
- W. Shi, H. Lombaert, W. Bai, **C. Ledig**, X. Zhuang, A. Marvao, T. Dawes, D. O’Regan, and D. Rueckert, “Multi-atlas Spectral PatchMatch: Application to cardiac image segmentation”, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 8673 of *Lecture Notes in Computer Science*, pp. 348-355, 2014.
- **C. Ledig**, W. Shi, W. Bai, and D. Rueckert, “Patch-based evaluation of image segmentation”, *Computer Vision and Pattern Recognition (CVPR)*, pp. 3065-3072, 2014.
- J. Lötjönen, **C. Ledig**, J. Koikkalainen, R. Wolz, L. Thurfjell, H. Soininen, S. Ourselin, and D. Rueckert, “Extended Boundary Shift Integral”, *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, pp. 854-857, 2014.

- **C. Ledig**, W. Shi, A. Makropoulos, J. Koikkalainen, R. A. Heckemann, A. Hammers, J. Loetjoenen, and D. Rueckert, “Consistent and robust 4D whole-brain segmentation: application to traumatic brain injury”, IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), pp. 673-676, 2014.
- A. Rao, **C. Ledig**, V. Newcombe, D. Menon, and D. Rueckert, “Contusion Segmentation from subjects with traumatic brain injury: a random forest framework”, IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), pp. 333-336, 2014.
- S. Pszczolkowski, S. Zafeirou, **C. Ledig**, and D. Rueckert, “A Robust Similarity Measure for Nonrigid Image Registration with Outliers”, IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), pp. 868-871, 2014.
- J. Koikkalainen, J. Loetjoenen, **C. Ledig**, D. Rueckert, O. Tenovuo, and D. Menon, “Automatic Analysis of CT images for traumatic brain injury”, IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), pp. 125-128, 2014.
- **C. Ledig**, R. A. Heckemann, A. Hammers, and D. Rueckert, “Improving whole-brain segmentations through incorporating regional image intensity statistics”, Proceedings of SPIE 8669, Medical Imaging 2013, pp. 86691M-86691M-7, 2013.
- W. Shi, J. Caballero, **C. Ledig**, X. Zhuang, W. Bai, K. Bhatia, A. Marvao, T Dawes, D. ORegan, and D. Rueckert, “Cardiac image super-resolution with global correspondence using multi-atlas patchmatch”, Medical Image Computing and Computer-Assisted Intervention (MICCAI), vol. 8151 of Lecture Notes in Computer Science, pp. 9-16, 2013.
- **C. Ledig**, R. Wolz, P. Aljabar, J. Loetjoenen, R. A. Heckemann, A. Hammers, and D. Rueckert, “Multi-class brain segmentation using atlas propagation and EM-based refinement”, IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), pp. 896-899, 2012.
- J. Lötjönen, R. Wolz, J. Koikkalainen, V. Manna, **C. Ledig**, L. Thurfjell, R. Lundqvist, G. Waldemar, H. Soininen, and D. Rueckert, “Hippocampal atrophy rate using an expectation maximization classifier with a disease-specific prior”, IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), pp. 1164-1167, 2012.

A.3 Grand challenges

- **C. Ledig**, R. Guerrero, T. Tong, K. Gray, A. Schmidt-Richberg, A. Makropoulos, R. A. Heckemann, and D. Rueckert, “Alzheimer’s disease state classification using structural volumetry, cortical thickness and intensity features”, MICCAI workshop Challenge on Computer-Aided Diagnosis of Dementia based on structural MRI data, pp. 55-64, 2014, **(3rd Place)**.
- **C. Ledig**, R. A. Heckemann, P. Aljabar, R. Wolz, J. V. Hajnal, A. Hammers, and D. Rueckert, “Segmentation of MRI brain scans using MALP-EM”, MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling, pp. 79-82, 2012, **(3rd Place)**.
- R. A. Heckemann, **C. Ledig**, P. Aljabar, K. R. Gray, D. Rueckert, J. V. Hajnal, and A. Hammers, “Label propagation using group agreement DISPATCH”, MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling, pp. 75-78, 2012.
- R. A. Heckemann, S. Keihaninejad, **C. Ledig**, P. Aljabar, D. Rueckert, J. V. Hajnal, and A. Hammers, “Multi-atlas propagation with enhanced registration - MAPER”, MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling, pp. 83- 86, 2012.
- A. Makropoulos, **C. Ledig**, P. Aljabar, A. Serag, J. V. Hajnal, D. A. Edwards, J. C. Serena, and D. Rueckert, “Automatic tissue and structural segmentation of neonatal brain MRI using Expectation-Maximization”, MICCAI Grand Challenge: Neonatal Brain Segmentation 2012 (NeoBrainS12), 2012. **(1st Place)**

A.4 Workshop publications

- A. Schuh, M. Murgasova, A. Makropoulos, **C. Ledig**, S. J. Counsell, J. V. Hajnal, P. Aljabar, and D. Rueckert, “Construction of a 4D Brain Atlas and Growth Model using Diffeomorphic Registration”, MICCAI Workshop on Spatio-temporal Image Analysis, 2014.
- R. Guerrero, **C. Ledig**, and D. Rueckert, “Manifold alignment and transfer learning for classification of Alzheimer’s disease”, Machine Learning in Medical Imaging, vol. 8679, pp. 77-84, 2014, **(Best Paper Award)**.
- R. A. Heckemann, R. Husson, **C. Ledig**, D. Rueckert, and A. Hammers, “Positional normalization as a first step in processing magnetic resonance brain images: work in progress”, Swedish Society for Automated Image Analysis (SSBA), 2013.

- **C. Ledig**, R. Wolz, P. Aljabar, J. Lötjönen and D. Rueckert, “PBSI: A symmetric probabilistic extension of the Boundary Shift Integral”, MICCAI 2012 Workshop on Novel Imaging Biomarkers for Alzheimer’s Disease and Related Disorders (NIBAD), pp. 125-132, 2012.
- **C. Ledig**, R. Wright, A. Serag, P. Aljabar, and D. Rueckert, “Neonatal brain segmentation using second order neighbourhood information”, MICCAI Workshop on Perinatal and Paediatric Imaging (PAPI), pp. 33-40, 2012.
- R. Wright, D. Vatansever, V. Kyriakopoulou, **C. Ledig**, R. Wolz, A. Serag, D. Rueckert, M. A. Rutherford, J. V. Hajnal, and P. Aljabar, “Age dependent fetal MR segmentation using manual and automated approaches”, MICCAI Workshop on Perinatal and Paediatric Imaging (PAPI), pp. 97-104, 2012.

A.5 Conference abstracts

- **C. Ledig**, R. Wolz, M. Austin, K. R. Gray, D. Rueckert, D. L. G. Hill, “Improved AD disease classification using structural volumes of the whole brain,” 12th International Conference on Alzheimer’s and Parkinson’s Diseases and Related Neurological Disorders (AD/PD), 2015.
- O. Tenovuo, D. Menon, M. van Gils, D. Rueckert, A. Katila, J. P. Coles, J. Mattila, **C. Ledig**, J. Frantzen, J. G. Outtrim, and J. Loetjoenen, “Improving the Individual Diagnostics of TBI the International TBicare project”, Proceedings of Tenth World Congress on Brain Injury, 2014.
- V. Newcombe, **C. Ledig**, G. Abate, J. Outtrim, D. Chatfield, T. Geeraerts, A. Manktelow, P. J. Hutchinson, J. Coles, G. Williams, D. Rueckert, and D. K. Menon, “Dynamic Evolution of Atrophy after moderate to severe Traumatic Brain Injury”, Journal of Neurotrauma, vol. 31, no. 5, pp. A36-A37, 2014.
- **C. Ledig**, V. Newcombe, M. G. Abate, J. G. Outtrim, D. Chatfield, T. Geeraerts, A. Manktelow, P. J. Hutchinson, J. P. Coles, G. Williams, D. Rueckert, and D. Menon, “Dynamic evolution of atrophy after traumatic brain injury”, Abstract accepted at ISMRM, 2014.
- A. Makropoulos, I. S. Gousias, **C. Ledig**, P. Aljabar, A. Serage, J. V. Hajnal, D. A. Edwards, S. J. Counsell, and D. Rueckert, “Automatic multi-label segmentation of the preterm brain with the use of adaptive atlases”, Abstract accepted at ISMRM, 2012.

B ROIs defined in the NMM atlas

Table B.1: Structure names of non-cortical structures in NMM brain atlas. Original ROI IDs were remapped to be consecutive.

non-cortical structure	abbreviation	ID (right/left)
3rd ventricle	3rdVent	1
4th Ventricle	4thVent	2
Accumbens Area	AccA	3/4
Amygdala	Am	5/6
Brain Stem	BS	7
Caudate	Cau	8/9
Cerebellum Exterior	CblmExt	10/11
Cerebellum White Matter	CblmWM	12/13
<i>Cerebral Exterior</i> ^{excluded}	<i>CrblExt</i>	<i>14/15</i>
Cerebral White Matter	CrblWM	16/17
Cerebrospinal Fluid	CSF	18
Hippocampus	Hc	19/20
Inf. Lateral Ventricle	infLV	21/22
Lateral Ventricle	LV	23/24
Pallidum	Pa	25/26
Putamen	Pu	27/28
Thalamus Proper	Th	29/30
Ventral DC	vDC	31/32
<i>Vessel</i> ^{excluded}	<i>Vsl</i>	<i>33/34</i>
Optic Chiasm	OptC	35
Cerebellar Vermal Lobules I-V	CVL1t5	36
Cerebellar Vermal Lobules VI-VII	CVL6t7	37
Cerebellar Vermal Lobules VIII-X	CVL8t10	38
Basal Forebrain	BF	40/39

Table B.2: Structure names of cortical structures in NMM brain atlas.

cortical structure	abbreviation	ID (right/left)
anterior cingulate gyrus	ACgG	41/42
anterior insula	AIns	43/44
anterior orbital gyrus	AOrG	45/46
angular gyrus	AnG	47/48
calcarine cortex	Calc	49/50
central operculum	CO	51/52
cuneus	Cun	53/54
entorhinal area	Ent	55/56
frontal operculum	FO	57/58
frontal pole	FRP	59/60
fusiform gyrus	FuG	61/62
gyrus rectus	GRe	63/64
inferior occipital gyrus	IOG	65/66
inferior temporal gyrus	ITG	67/68
lingual gyrus	LiG	69/70
lateral orbital gyrus	LOrG	71/72
middle cingulate gyrus	MCgG	73/74
medial frontal cortex	MFC	75/76
middle frontal gyrus	MFG	77/78
middle occipital gyrus	MOG	79/80
medial orbital gyrus	MOrG	81/82
postcentral gyrus medial segment	MPoG	83/84
precentral gyrus medial segment	MPrG	85/86
superior frontal gyrus medial segment	MSFG	87/88
middle temporal gyrus	MTG	89/90
occipital pole	OCP	91/92
occipital fusiform gyrus	OFuG	93/94
opercular part of the inferior frontal gyrus	OpIFG	95/96
orbital part of the inferior frontal gyrus	OrIFG	97/98
posterior cingulate gyrus	PCgG	99/100
precuneus	PCu	101/102
parahippocampal gyrus	PHG	103/104
posterior insula	PIns	105/106
parietal operculum	PO	107/108
postcentral gyrus	PoG	109/110
posterior orbital gyrus	POrG	111/112
planum polare	PP	113/114
precentral gyrus	PrG	115/116
planum temporale	PT	117/118
subcallosal area	SCA	119/120
superior frontal gyrus	SFG	121/122
supplementary motor cortex	SMC	123/124
supramarginal gyrus	SMG	125/126
superior occipital gyrus	SOG	127/128
superior parietal lobule	SPL	129/130
superior temporal gyrus	STG	131/132
temporal pole	TMP	133/134
triangular part of the inferior frontal gyrus	TrIFG	135/136
transverse temporal gyrus	TTG	137/138

C The Marshall Classification System

Table C.1: Marshall Classification system based on and modified from Marshall et al. (1991).

Marshall class	Description
1 Diffuse injury (DI) I	no visible intracranial pathological changes seen on CT
2 Diffuse injury II	cisterns are present with midline shift of 0–5 mm and/or: lesions densities present; no high or mixed density lesion $>25 \text{ cm}^3$ may include bone fragments and foreign bodies*
3 Diffuse injury III (swelling)	cisterns compressed or absent with midline shift of 0–5 mm; no high or mixed density lesion $>25 \text{ cm}^3$
4 Diffuse injury IV (shift)	midline shift $>5 \text{ mm}$; no high or mixed density lesion $> 25 \text{ cm}^3$
5 Evacuated mass lesion (EML)	any lesion surgically evacuated
6 Non-evacuated mass lesion (NEML) [†]	high or mixed density lesion $>25 \text{ cm}^3$; not surgically evacuated

*As may be the case in depressed skull fractures.

[†]Merged into group 5 in data provided by Turku University Hospital.

D The Glasgow Outcome Scale

Table D.1: Glasgow Outcome Scale based on and modified from Jennett and Bond (1975); Jennett et al. (1981).

GOS	extended GOS	Description
1 (D)	1 (D)	<i>Dead</i>
2 (VS)	2 (VS)	<i>Vegetative state</i> : no evidence of meaningful responsiveness.
3 (SD)	3 (SD-)	<i>Severe disability</i> : conscious, but unable to live independently
	4 (SD+)	due to mental or physical disability.
4 (MD)	5 (MD-)	<i>Moderate disability</i> : able to live independently,
	6 (MD+)	limited ability to return to work or school.
5 (GR)	7 (GR-)	<i>Good recovery</i> : capacity to resume normal occupational and social activities,
	8 (GR+)	minor deficits possible.

E ADNI classification results

E.1 Cross-sectional analysis at baseline (AD vs. HC, uncorrected volumes)

Table E.1: Classification results in % (6-fold CV, LDA 100 runs, RF/SVM 20 runs) for separating AD from CN. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses. Sorted by effect size.

AD patients (N = 322, Positives ^a) vs. Healthy Controls (N = 404, Negatives ^b) (baseline analysis, uncorrected volumes)									
structure	ACC (bACC)	SENS	SPEC	mean (SD) [mm ³] ^P	mean (SD) [mm ³] ^N	effect size (d)	p-value	sig.	(corr.)
RandomForest (all features)	85 (85)	81	88						
SVM (all features)	90 (90)	86	93						
Gender (female = 0, male = 1)	52 (52)	55	50	0.5 (0.5)	0.5 (0.5)	0.099	0.18349	o (o)	
Age	51 (51)	51	51	75.1 (7.7)	74.9 (5.7)	0.027	0.71340	o (o)	
Ventricles	66 (65)	55	75	63295.6 (28189.0)	46227.4 (20971.9)	0.698	<0.00001	++ (++)	
CorticalGreyMatter	61 (61)	62	60	491782.0 (56040.1)	519904.6 (50579.3)	0.530	<0.00001	++ (++)	
BrainTissue	56 (56)	57	55	1133344.3 (118326.9)	1163549.9 (110731.1)	0.265	0.00042	++ (+)	
DeepGreyMatter	55 (55)	57	54	172879.2 (18868.2)	175625.6 (16551.2)	0.156	0.03725	+	(o)
Brain	54 (55)	56	53	1199158.5 (132290.9)	1212115.6 (118623.9)	0.104	0.16526	o (o)	
WhiteMatter	48 (47)	47	48	468683.1 (62373.7)	468019.7 (60538.5)	0.011	0.88497	o (o)	
LeftHippocampus	76 (76)	76	75	2608.0 (479.7)	3221.8 (434.3)	1.349	<0.00001	++ (++)	
Amygdala	75 (75)	73	78	1825.9 (388.6)	2296.6 (314.5)	1.348	<0.00001	++ (++)	
Hippocampus	75 (75)	75	75	5429.2 (943.8)	6601.4 (825.8)	1.332	<0.00001	++ (++)	
RightAmygdala	75 (75)	73	76	925.6 (209.5)	1168.0 (171.6)	1.280	<0.00001	++ (++)	
EntA	73 (73)	71	75	3427.5 (675.5)	4245.4 (609.2)	1.279	<0.00001	++ (++)	
LeftAmygdala	75 (75)	73	77	900.3 (205.1)	1128.6 (162.7)	1.250	<0.00001	++ (++)	
RightHippocampus	73 (73)	71	75	2821.2 (546.5)	3379.6 (427.4)	1.154	<0.00001	++ (++)	
InfLatVent	73 (72)	58	85	3432.1 (1556.6)	2107.1 (825.3)	1.099	<0.00001	++ (++)	
LeftInfLatVent	75 (74)	63	85	1652.9 (765.4)	1006.8 (420.7)	1.079	<0.00001	++ (++)	
RightInfLatVent	70 (68)	55	82	1779.3 (902.2)	1100.3 (447.5)	0.988	<0.00001	++ (++)	
ITG	66 (66)	66	65	21776.6 (3325.8)	24529.3 (2916.2)	0.887	<0.00001	++ (++)	
PHG	66 (66)	68	65	6447.4 (977.5)	7217.3 (899.2)	0.824	<0.00001	++ (++)	
AnG	67 (67)	67	67	15507.5 (2580.5)	17559.5 (2480.9)	0.812	<0.00001	++ (++)	
MTG	64 (64)	66	62	25677.9 (4030.6)	28511.8 (3713.2)	0.735	<0.00001	++ (++)	
TMP	64 (63)	61	65	13730.4 (2468.4)	15339.5 (2004.0)	0.724	<0.00001	++ (++)	
LateralVentricle	65 (64)	55	73	55184.7 (26190.1)	39883.0 (19481.1)	0.674	<0.00001	++ (++)	
LeftLateralVentricle	66 (65)	55	75	29062.5 (14264.5)	20827.4 (10483.3)	0.669	<0.00001	++ (++)	
SMG	62 (63)	64	61	13731.4 (2077.6)	15086.0 (2006.8)	0.664	<0.00001	++ (++)	
RightLateralVentricle	63 (62)	54	71	26122.3 (12557.2)	19055.6 (9403.3)	0.647	<0.00001	++ (++)	
STG	63 (63)	62	63	12671.1 (2031.8)	13889.0 (1933.7)	0.616	<0.00001	++ (++)	
OFuG	60 (60)	61	60	7382.1 (1324.3)	8129.7 (1267.7)	0.578	<0.00001	++ (++)	
ThalamusProper	61 (61)	63	60	12486.4 (1477.5)	13304.8 (1426.0)	0.565	<0.00001	++ (++)	
3rdVentricle	60 (60)	54	65	2319.7 (828.7)	1891.2 (721.5)	0.556	<0.00001	++ (++)	
BasalForebrain	62 (62)	64	60	795.6 (215.5)	914.0 (213.2)	0.552	<0.00001	++ (++)	
FuG	60 (61)	63	59	15951.4 (2428.3)	17131.2 (2281.3)	0.503	<0.00001	++ (++)	
AIIns	62 (62)	63	60	8241.7 (1225.1)	8820.0 (1142.4)	0.490	<0.00001	++ (++)	
AccumbensArea	60 (60)	57	62	617.0 (184.6)	695.7 (148.3)	0.475	<0.00001	++ (++)	
MOG	59 (59)	61	57	10087.5 (1773.4)	10837.5 (1524.1)	0.457	<0.00001	++ (++)	
IOG	58 (58)	57	59	11413.2 (2016.2)	12233.2 (1815.4)	0.430	<0.00001	++ (++)	
SPL	58 (58)	59	57	16943.2 (2469.3)	17977.7 (2441.6)	0.422	<0.00001	++ (++)	
PIIns	58 (59)	60	57	4482.8 (701.8)	4771.4 (703.1)	0.411	<0.00001	++ (++)	
PP	58 (58)	60	55	3682.5 (589.7)	3912.8 (549.4)	0.406	<0.00001	++ (++)	
FO	57 (57)	58	55	3504.1 (549.2)	3718.8 (539.9)	0.395	<0.00001	++ (++)	
MFG	59 (59)	60	59	33109.0 (5125.3)	34968.2 (4791.0)	0.376	<0.00001	++ (++)	
SFG	56 (57)	58	55	26326.6 (3837.4)	27676.8 (3441.8)	0.373	<0.00001	++ (++)	
PT	57 (57)	61	54	3658.7 (712.6)	3926.4 (769.4)	0.359	<0.00001	++ (++)	
TrIFG	54 (54)	55	54	6113.5 (1082.3)	6494.0 (1071.5)	0.354	<0.00001	++ (++)	
Caudate	58 (56)	43	69	7535.3 (2705.8)	6812.7 (1705.2)	0.328	0.00001	++ (+)	
Putamen	56 (56)	57	55	7455.9 (1342.8)	7878.8 (1331.8)	0.316	0.00003	++ (+)	
SCA	56 (57)	58	55	2367.0 (496.3)	2510.7 (434.7)	0.310	0.00004	++ (+)	
PCu	55 (56)	58	53	21742.5 (3583.1)	22743.6 (3039.0)	0.304	0.00005	++ (+)	
PCgG	56 (56)	57	55	8494.1 (1311.2)	8862.3 (1128.5)	0.304	0.00005	++ (+)	
CSF	57 (56)	51	61	2419.4 (579.6)	2250.7 (598.6)	0.286	0.00014	++ (+)	
POrG	55 (55)	57	54	6068.4 (927.5)	6312.4 (859.2)	0.274	0.00026	++ (+)	
MCgG	55 (54)	48	60	10326.7 (1674.5)	9905.0 (1511.3)	0.266	0.00040	++ (+)	
MSFG	56 (57)	61	53	12685.2 (2000.5)	13151.5 (1735.4)	0.251	0.00082	++ (+)	
SOG	55 (55)	57	54	7063.8 (1242.6)	7354.8 (1136.5)	0.246	0.00106	++ (+)	
MFC	54 (54)	54	54	3510.2 (716.9)	3663.4 (603.6)	0.233	0.00185	++ (+)	
OCP	53 (53)	57	50	4494.6 (931.2)	4728.0 (1076.0)	0.230	0.00215	++ (+)	
CerebellarVermalLobulesI-V	54 (54)	56	52	4464.0 (672.4)	4616.9 (672.9)	0.227	0.00244	++ (+)	
OrIFG	55 (55)	60	50	2749.4 (587.8)	2879.5 (568.2)	0.225	0.00263	++ (+)	
LORg	54 (55)	57	52	3713.5 (688.9)	3853.3 (600.1)	0.218	0.00363	++ (+)	
OpIFG	52 (53)	57	49	5649.2 (1036.1)	5864.9 (1065.5)	0.205	0.00624	++ (+)	
GRe	53 (53)	55	52	3228.0 (549.0)	3335.1 (512.8)	0.202	0.00691	++ (+)	
PO	54 (55)	58	52	4251.3 (792.1)	4409.9 (796.3)	0.200	0.00770	++ (+)	
CerebellumWhiteMatter	54 (52)	40	65	32795.7 (8328.0)	31344.3 (7506.8)	0.184	0.01393	++ (+)	
SMC	53 (53)	57	49	10285.9 (1713.6)	10567.8 (1418.9)	0.181	0.01558	++ (+)	
AORg	51 (51)	52	50	3804.7 (617.0)	3906.9 (598.1)	0.168	0.02448	++ (+)	
TTG	56 (57)	62	52	2387.0 (557.6)	2477.1 (532.0)	0.166	0.02684	++ (+)	
CerebellarVermalLobulesVI-VII	53 (53)	50	55	2417.3 (432.6)	2355.7 (390.0)	0.150	0.04435	++ (+)	
POg	53 (53)	55	51	18093.5 (2551.3)	18449.6 (2321.8)	0.147	0.04984	++ (+)	
MPoG	54 (54)	56	52	1644.7 (394.0)	1699.2 (370.6)	0.143	0.05604	++ (+)	
PrG	52 (52)	53	51	23676.0 (3462.6)	24096.7 (3052.7)	0.130	0.08268	++ (+)	
CO	53 (53)	55	51	7785.8 (1114.5)	7923.9 (1090.1)	0.125	0.09374	++ (+)	
Cale	53 (53)	50	55	7085.9 (1722.2)	6884.2 (1531.2)	0.125	0.09586	++ (+)	
FRP	51 (52)	53	50	5444.2 (1113.1)	5565.8 (996.7)	0.116	0.12123	++ (+)	
MPrg	53 (53)	55	51	5428.2 (957.0)	5531.1 (847.0)	0.115	0.12515	++ (+)	
MORg	52 (52)	54	51	8243.5 (1150.7)	8361.5 (1010.3)	0.110	0.14236	++ (+)	
ACg	51 (51)	54	49	8487.7 (1576.8)	8613.8 (1458.5)	0.083	0.26466	++ (+)	
VentralDC	51 (51)	52	50	8448.8 (965.4)	8523.1 (931.8)	0.079	0.29369	++ (+)	
BrainStem	50 (50)	52	49	18742.3 (2233.9)	18885.5 (2141.8)	0.066	0.38005	++ (+)	
Pallidum	51 (51)	49	52	2668.2 (454.5)	2645.4 (394.7)	0.054	0.47061	++ (+)	
Cun	49 (49)	49	48	10691.9 (2079.7)	10753.0 (1794.8)	0.032	0.67110	++ (+)	
CerebellarVermalLobulesVIII-X	48 (48)	47	49	2823.6 (474.2)	2814.5 (428.5)	0.020	0.78685	++ (+)	
4thVentricle	48 (48)	45	51	2359.0 (691.3)	2346.0 (690.8)	0.019	0.80132	++ (+)	
CerebralWhiteMatter	48 (48)	49	48	435887.4 (58972.0)	436675.4 (56880.4)	0.014	0.85528	++ (+)	
CerebellumExterior	48 (48)	48	49	97168.8 (12890.5)	97279.6 (11151.4)	0.009	0.90128	++ (+)	
LiG	48 (48)	47	48	18512.9 (2794.7)	18524.4 (2476.3)	0.004	0.95334	++ (+)	

E.2 Cross-sectional analysis at baseline (AD vs. HC, corrected volumes)

Table E.2: Classification results in % (6-fold CV, LDA 100 runs, RF/SVM 20 runs) for separating AD from CN. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses. Sorted by effect size.

AD patients (N = 322, Positives ^P) vs. Healthy Controls (N = 404, Negatives ^N) (baseline analysis, [†] volumes corrected for age/gender/brain size)									
structure	ACC (bACC)	SENS	SPEC	mean (SD) [mm ³] ^{P,†}	mean (SD) [mm ³] ^{N,†}	effect size (d)	p-value	sig. (corr.)	
RandomForest (all features)	87 (86)	83	90						
SVM (all features)	90 (89)	86	92						
Gender (female = 0, male = 1)	52 (52)	55	50	0.5 (0.5)	0.5 (0.5)	0.099	0.18349	o (o)	
Age	50 (50)	50	50	75.1 (7.7)	74.9 (5.7)	0.027	0.71340	o (o)	
BrainTissue	72 (71)	63	78	-17942.8 (22718.7)	0 (17403.6)	0.900	<0.00001	++ (++)	
CorticalGreyMatter	68 (67)	63	72	-23635.9 (28907.3)	0 (24078.1)	0.898	<0.00001	++ (++)	
Ventricles	72 (71)	63	79	17757.3 (22639.6)	0 (17293.9)	0.895	<0.00001	++ (++)	
WhiteMatter	57 (56)	51	61	7242.2 (29309.8)	0 (28686.1)	0.250	0.00086	++ (o)	
DeepGreyMatter	52 (52)	50	54	-1549.1 (13310.2)	0 (11142.2)	0.127	0.08834	o (o)	
Brain	54 (55)	56	53	1199158.5 (132290.9)	1212115.6 (118623.9)	0.104	0.16526	o (o)	
Amygdala	80 (80)	76	84	-452.0 (332.0)	0 (250.5)	1.561	<0.00001	++ (++)	
Hippocampus	78 (78)	75	80	-1115.4 (817.7)	0 (660.7)	1.519	<0.00001	++ (++)	
EntA	78 (78)	76	80	-801.3 (583.3)	0 (485.3)	1.509	<0.00001	++ (++)	
LeftHippocampus	79 (78)	76	81	-588.1 (423.1)	0 (364.6)	1.502	<0.00001	++ (++)	
RightAmygdala	80 (80)	77	83	-232.2 (182.6)	0 (139.4)	1.452	<0.00001	++ (++)	
LeftAmygdala	78 (78)	75	81	-219.8 (179.7)	0 (135.5)	1.403	<0.00001	++ (++)	
RightHippocampus	76 (75)	71	79	-527.3 (488.0)	0 (339.1)	1.280	<0.00001	++ (++)	
InfLatVent	78 (77)	65	89	1330.8 (1367.4)	0 (702.2)	1.267	<0.00001	++ (++)	
LeftInfLatVent	77 (76)	66	86	649.5 (677.1)	0 (360.6)	1.237	<0.00001	++ (++)	
ITG	71 (71)	69	73	-2588.0 (2506.9)	0 (1954.5)	1.168	<0.00001	++ (++)	
RightInfLatVent	75 (73)	59	87	-681.3 (814.9)	0 (391.3)	1.106	<0.00001	++ (++)	
PHG	68 (68)	64	72	-739.8 (832.6)	0 (702.4)	0.970	<0.00001	++ (++)	
MTG	69 (69)	69	70	-2534.0 (2853.0)	0 (2411.3)	0.969	<0.00001	++ (++)	
AnG	68 (67)	65	70	-1880.5 (2345.6)	0 (1973.2)	0.876	<0.00001	++ (++)	
TMP	66 (65)	64	67	-1521.7 (1967.2)	0 (1539.3)	0.873	<0.00001	++ (++)	
LateralVentricle	71 (70)	62	78	15972.2 (21196.2)	0 (16214.4)	0.859	<0.00001	++ (++)	
LeftLateralVentricle	70 (69)	61	76	8584.7 (11586.9)	0 (8816.1)	0.847	<0.00001	++ (++)	
RightLateralVentricle	70 (69)	59	78	7387.5 (10387.1)	0 (7883.2)	0.814	<0.00001	++ (++)	
STG	65 (65)	64	66	-1100.3 (1610.7)	0 (1478.6)	0.715	<0.00001	++ (++)	
3rdVentricle	67 (66)	63	70	434.3 (699.9)	0 (591.1)	0.677	<0.00001	++ (++)	
SMG	64 (64)	63	65	-1196.5 (2003.1)	0 (1697.8)	0.650	<0.00001	++ (++)	
ThalamusProper	63 (63)	61	65	-680.6 (1214.1)	0 (1005.2)	0.617	<0.00001	++ (++)	
FuG	62 (62)	62	61	-1072.1 (1750.6)	0 (1769.3)	0.609	<0.00001	++ (++)	
OFuG	60 (60)	58	62	-690.7 (1225.8)	0 (1061.2)	0.607	<0.00001	++ (++)	
BasalForebrain	63 (63)	66	61	-109.6 (195.4)	0 (189.3)	0.571	<0.00001	++ (++)	
AIIns	59 (60)	62	58	-509.3 (995.9)	0 (915.4)	0.535	<0.00001	++ (++)	
MOG	60 (60)	61	59	-643.3 (1375.1)	0 (1195.7)	0.503	<0.00001	++ (++)	
IOG	59 (59)	60	59	-697.2 (1571.8)	0 (1386.2)	0.474	<0.00001	++ (++)	
AccumbensArea	60 (60)	59	61	-71.7 (177.2)	0 (138.5)	0.457	<0.00001	++ (++)	
PP	58 (58)	61	56	-200.0 (464.5)	0 (433.4)	0.447	<0.00001	++ (++)	
MCgG	58 (57)	54	60	544.3 (1270.6)	0 (1190.1)	0.444	<0.00001	++ (++)	
PIns	58 (58)	58	58	-231.1 (524.6)	0 (547.9)	0.430	<0.00001	++ (++)	
MFG	58 (58)	56	60	-1467.9 (3689.2)	0 (3194.1)	0.429	<0.00001	++ (++)	
SPL	56 (56)	57	56	-867.5 (2175.8)	0 (1972.5)	0.420	<0.00001	++ (++)	
SFG	59 (59)	60	58	-1124.3 (3032.2)	0 (2632.7)	0.399	<0.00001	++ (++)	
Caudate	58 (57)	46	68	799.5 (2447.8)	0 (1597.5)	0.396	<0.00001	++ (++)	
FO	60 (60)	59	60	-187.3 (506.5)	0 (458.7)	0.390	<0.00001	++ (++)	
PT	56 (56)	58	55	-232.4 (590.0)	0 (636.4)	0.377	<0.00001	++ (++)	
TriFG	57 (57)	56	58	-331.8 (985.1)	0 (933.4)	0.347	<0.00001	++ (++)	
PCu	56 (56)	57	54	-788.1 (2502.1)	0 (2124.3)	0.343	<0.00001	++ (++)	
PCgG	57 (57)	59	56	-306.1 (960.5)	0 (898.4)	0.330	0.00001	++ (++)	
SCA	58 (58)	58	59	-124.5 (441.7)	0 (388.6)	0.301	0.00006	++ (+)	
POrG	57 (57)	54	59	-211.0 (738.1)	0 (685.9)	0.297	0.00008	++ (+)	
CSF	58 (58)	53	62	172.5 (621.2)	0 (545.8)	0.297	0.00008	++ (+)	
Putamen	57 (57)	58	56	-367.5 (1300.2)	0 (1243.1)	0.290	0.00012	++ (+)	
MSFG	54 (54)	55	53	-349.9 (1447.4)	0 (1334.5)	0.253	0.00076	++ (o)	
MFC	53 (53)	54	53	-123.9 (577.8)	0 (493.6)	0.233	0.00192	++ (o)	
CerebellumWhiteMatter	55 (54)	44	64	1677.9 (7982.2)	0 (6947.8)	0.226	0.00257	++ (o)	
OCF	53 (54)	57	50	-207.2 (815.9)	0 (999.5)	0.225	0.00274	++ (o)	
SOG	55 (56)	58	53	-220.0 (1044.7)	0 (974.8)	0.219	0.00354	++ (o)	
CerebralWhiteMatter	56 (55)	52	58	5564.3 (26392.1)	0 (25736.3)	0.214	0.00434	++ (o)	
OrIFG	55 (56)	59	52	-109.5 (518.0)	0 (510.3)	0.213	0.00446	++ (o)	
LOrG	52 (52)	53	51	-111.8 (565.6)	0 (492.2)	0.213	0.00456	++ (o)	
CerebellarVermalLobulesI-V	56 (56)	57	55	-128.8 (620.1)	0 (632.2)	0.205	0.00610	++ (o)	
Calc	53 (53)	51	55	275.0 (1536.7)	0 (1432.9)	0.186	0.01309	++ (o)	
CerebellarVermalLobulesVI-VII	54 (54)	52	55	71.0 (401.3)	0 (370.4)	0.185	0.01368	++ (o)	
OpIFG	51 (51)	53	50	-153.5 (890.2)	0 (909.9)	0.170	0.02290	++ (o)	
GRe	52 (52)	54	51	-77.1 (463.7)	0 (463.0)	0.166	0.02614	++ (o)	
SMC	52 (52)	54	51	-192.1 (1309.9)	0 (1108.0)	0.160	0.03273	++ (o)	
PO	53 (53)	54	52	-105.0 (682.3)	0 (637.7)	0.160	0.03303	++ (o)	
AOrG	52 (52)	52	52	-71.6 (483.0)	0 (464.7)	0.151	0.04301	++ (o)	
TTG	54 (54)	56	53	-70.4 (481.9)	0 (455.3)	0.151	0.04398	++ (o)	
MPoG	53 (53)	53	53	-44.8 (377.8)	0 (348.2)	0.124	0.09769	o (o)	
PrG	52 (53)	55	50	-262.1 (2723.4)	0 (2472.1)	0.101	0.17534	o (o)	
Pallidum	54 (54)	55	53	38.0 (418.7)	0 (362.0)	0.098	0.19090	o (o)	
PoG	52 (52)	52	52	-189.2 (2124.4)	0 (1913.6)	0.094	0.20803	o (o)	
FRP	50 (50)	53	47	-76.0 (962.3)	0 (901.5)	0.082	0.27406	o (o)	
MPrG	51 (51)	53	50	-58.2 (819.9)	0 (747.4)	0.075	0.31847	o (o)	
CO	50 (50)	51	49	-58.3 (783.8)	0 (796.8)	0.074	0.32384	o (o)	
MOrG	51 (51)	53	49	-61.3 (943.0)	0 (819.7)	0.070	0.34945	o (o)	
CerebellarVermalLobulesVIII-X	51 (51)	51	51	22.1 (423.1)	0 (375.1)	0.056	0.45705	o (o)	
LiG	51 (51)	51	51	106.4 (2022.0)	0 (1846.7)	0.055	0.46010	o (o)	
CerebellumExterior	52 (52)	52	51	442.1 (10408.6)	0 (8931.0)	0.046	0.53842	o (o)	
4thVentricle	50 (50)	46	54	19.9 (646.4)	0 (643.3)	0.031	0.67923	o (o)	
Cun	48 (48)	46	49	33.2 (1619.0)	0 (1403.3)	0.022	0.76765	o (o)	
VentralDC	47 (47)	46	47	-13.7 (709.3)	0 (597.0)	0.021	0.77728	o (o)	
BrainStem	48 (48)	47	48	17.8 (1574.8)	0 (1484.5)	0.012	0.87624	o (o)	
ACg	47 (47)	47	48	-5.9 (1178.5)	0 (1101.8)	0.005	0.94417	o (o)	

E.3 Cross-sectional analysis at baseline (pMCI vs. sMCI, corrected volumes)

Table E.3: Classification results in % (6-fold CV, LDA 100 runs, RF/SVM 20 runs) for separating progressive from stable MCI. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses. Sorted by effect size.

progressive MCI (N = 177, Positives ^P) vs. stable MCI (N = 166, Negatives ^N) (baseline analysis, ¹ volumes corrected for age/gender/brain size)									
structure	ACC (bACC)	SENS	SPEC	mean (SD) [mm ³] ^{P,†}	mean (SD) [mm ³] ^{N,†}	effect size (d)	p-value	sig. (corr.)	
RandomForest (all features)	68 (68)	72	64						
SVM (all features)	67 (67)	70	64						
Age	47 (47)	45	48	73.5 (7.6)	73.8 (7.6)	0.029	0.78516	o (o)	
Gender (female = 0, male = 1)	45 (45)	45	45	0.6 (0.5)	0.6 (0.5)	0.006	0.95829	o (o)	
BrainTissue	60 (60)	51	69	-12652.2 (20144.6)	-4372.1 (20741.3)	0.405	0.00021	++ (+)	
Ventricles	60 (60)	51	70	12483.5 (20045.3)	4427.7 (20676.8)	0.396	0.00029	++ (+)	
CorticalGreyMatter	56 (56)	55	56	-16743.9 (27348.8)	-6929.4 (26948.6)	0.361	0.00091	++ (o)	
Brain	51 (51)	52	49	1222424.9 (131084.2)	1238635.5 (118830.2)	0.129	0.23200	o (o)	
WhiteMatter	48 (48)	47	49	5966.3 (27919.1)	4607.3 (26735.5)	0.050	0.64592	o (o)	
DeepGreyMatter	47 (47)	46	48	-1874.6 (11838.0)	-2050.0 (11988.8)	0.015	0.89167	o (o)	
Amygdala	65 (65)	63	68	-387.4 (308.7)	-149.4 (352.2)	0.720	<0.00001	++ (++)	
LeftAmygdala	62 (62)	61	64	-191.6 (161.2)	-70.4 (189.7)	0.690	<0.00001	++ (++)	
InfLatVent	64 (64)	54	74	1034.0 (1097.9)	358.6 (903.1)	0.670	<0.00001	++ (++)	
LeftInfLatVent	65 (65)	56	75	480.0 (538.2)	156.3 (451.1)	0.650	<0.00001	++ (++)	
RightAmygdala	65 (65)	63	67	-195.8 (176.7)	-79.0 (184.1)	0.648	<0.00001	++ (++)	
EntA	61 (61)	61	60	-678.3 (565.7)	-294.0 (622.8)	0.647	<0.00001	++ (++)	
RightInfLatVent	64 (64)	54	74	554.0 (689.1)	202.3 (527.5)	0.571	<0.00001	++ (++)	
Hippocampus	61 (61)	63	60	-1042.8 (817.1)	-604.7 (786.1)	0.546	<0.00001	++ (++)	
RightHippocampus	63 (63)	61	64	-514.7 (442.9)	-287.7 (408.3)	0.532	<0.00001	++ (++)	
MTG	60 (60)	59	60	-2305.3 (2698.1)	-925.8 (2708.9)	0.510	<0.00001	++ (++)	
AnG	60 (60)	61	59	-1579.0 (2223.0)	-495.7 (2084.0)	0.502	<0.00001	++ (++)	
LeftHippocampus	61 (61)	62	60	-528.1 (429.6)	-317.0 (425.8)	0.493	<0.00001	++ (++)	
CSF	59 (59)	54	64	156.4 (533.9)	-61.6 (491.3)	0.424	0.00010	++ (+)	
3rdVentricle	59 (59)	55	64	376.2 (661.3)	113.9 (639.9)	0.403	0.00023	++ (+)	
LateralVentricle	60 (61)	51	70	11080.2 (18700.0)	3986.3 (19536.4)	0.371	0.00066	++ (o)	
RightLateralVentricle	60 (60)	51	69	5078.1 (9151.1)	1656.3 (9510.1)	0.367	0.00077	++ (o)	
LeftLateralVentricle	59 (59)	51	68	6002.1 (10336.4)	2330.0 (10722.7)	0.349	0.00136	++ (o)	
STG	57 (57)	57	56	-1048.8 (1509.9)	-513.8 (1583.2)	0.346	0.00149	++ (o)	
ITG	57 (57)	57	56	-1510.5 (2300.7)	-691.2 (2482.3)	0.343	0.00165	++ (o)	
TMP	58 (58)	58	58	-875.3 (1703.2)	-310.1 (1766.7)	0.326	0.00275	++ (o)	
MCgG	55 (55)	50	60	472.7 (1159.1)	78.4 (1292.7)	0.322	0.00311	++ (o)	
BasalForebrain	57 (57)	57	57	-108.9 (190.9)	-45.6 (207.0)	0.318	0.00346	++ (o)	
IOG	56 (56)	59	52	-466.0 (1583.2)	17.9 (1546.6)	0.309	0.00449	++ (o)	
FuG	56 (56)	57	55	-776.1 (1755.3)	-257.5 (1632.4)	0.306	0.00495	++ (o)	
MOG	58 (58)	59	56	-445.8 (1531.4)	-25.6 (1245.4)	0.300	0.00579	++ (o)	
OFuG	56 (56)	55	57	-415.3 (1153.0)	-57.9 (1242.7)	0.299	0.00604	++ (o)	
AccumbensArea	59 (59)	55	63	-66.5 (162.8)	-24.3 (137.4)	0.279	0.01012	++ (o)	
ThalamusProper	56 (55)	57	54	-650.7 (1342.2)	-310.3 (1187.7)	0.268	0.01356	++ (o)	
Putamen	59 (59)	59	60	-356.0 (1372.3)	-28.0 (1071.7)	0.265	0.01454	++ (o)	
SMG	53 (53)	53	53	-880.2 (1933.5)	-376.0 (1924.4)	0.261	0.01608	++ (o)	
OrIFG	58 (58)	60	55	-111.1 (518.9)	16.4 (489.0)	0.253	0.01998	++ (o)	
PIus	58 (58)	59	57	-216.3 (569.6)	-82.3 (486.6)	0.252	0.02010	++ (o)	
AIus	56 (55)	60	51	-403.2 (892.4)	-175.9 (937.7)	0.248	0.02206	++ (o)	
FO	58 (58)	61	54	-189.9 (473.2)	-73.5 (468.1)	0.247	0.02266	++ (o)	
FRP	55 (55)	55	55	-178.4 (995.6)	58.1 (978.8)	0.239	0.02731	++ (o)	
PCgG	53 (53)	57	50	-176.8 (851.6)	35.2 (947.5)	0.236	0.02984	++ (o)	
SFG	51 (51)	50	53	-891.9 (2741.4)	-274.6 (2842.4)	0.221	0.04139	++ (o)	
MSFG	55 (55)	54	56	-383.6 (1538.4)	-69.4 (1404.0)	0.213	0.04948	++ (o)	
PO	58 (58)	58	57	-150.2 (668.2)	-12.6 (636.3)	0.211	0.05185	++ (o)	
PrG	53 (53)	51	54	327.2 (3003.8)	-269.9 (2644.4)	0.211	0.05213	++ (o)	
SPL	55 (55)	57	53	-513.7 (2105.3)	-91.4 (1985.5)	0.206	0.05715	++ (o)	
PCu	55 (55)	54	57	-543.4 (2447.9)	-80.3 (2204.2)	0.198	0.06707	++ (o)	
PHG	56 (56)	56	55	-553.9 (799.3)	-403.9 (766.7)	0.191	0.07742	++ (o)	
Calc	53 (54)	48	59	313.9 (1471.6)	72.0 (1551.3)	0.160	0.13935	++ (o)	
MFG	49 (49)	51	47	-1109.3 (3500.4)	-578.0 (3119.2)	0.160	0.13961	++ (o)	
SMC	53 (53)	52	53	-71.7 (1254.1)	-267.8 (1219.8)	0.158	0.14334	++ (o)	
Caudate	51 (51)	39	63	490.3 (2016.4)	172.7 (2103.6)	0.154	0.15432	++ (o)	
CerebellarVermalLobulesVI-VII	53 (54)	49	58	62.9 (350.8)	6.4 (396.3)	0.151	0.16198	++ (o)	
BrainStem	50 (51)	49	52	35.3 (1656.3)	-201.8 (1493.6)	0.150	0.16584	++ (o)	
TrIFG	53 (53)	54	51	-328.1 (972.6)	-183.2 (958.8)	0.150	0.16586	++ (o)	
GRG	54 (54)	57	51	-40.2 (484.7)	31.6 (507.6)	0.145	0.18101	++ (o)	
PP	54 (54)	56	53	-209.2 (444.7)	-153.5 (405.8)	0.131	0.22693	++ (o)	
MPrG	49 (49)	47	50	141.9 (981.1)	32.8 (760.2)	0.124	0.25265	++ (o)	
VentralDCP	53 (53)	52	53	-55.7 (672.9)	-132.9 (624.4)	0.119	0.27208	++ (o)	
OCF	52 (52)	53	51	25.5 (1055.1)	136.3 (908.3)	0.112	0.29959	++ (o)	
SCA	50 (50)	51	49	-90.4 (407.2)	-46.3 (381.1)	0.112	0.30165	++ (o)	
PT	55 (55)	55	54	-274.9 (635.1)	-208.9 (575.4)	0.109	0.31490	++ (o)	
ACg	51 (51)	52	50	-161.8 (1157.0)	-40.5 (1133.5)	0.106	0.32780	++ (o)	
CerebellumExterior	53 (53)	50	56	271.9 (9093.7)	-695.3 (9828.3)	0.102	0.34449	++ (o)	
OpIFG	52 (52)	50	54	-164.8 (946.5)	-74.3 (994.5)	0.093	0.38867	++ (o)	
LORg	49 (49)	48	49	-94.9 (526.9)	-46.4 (610.2)	0.085	0.43046	++ (o)	
SOG	52 (51)	54	49	-30.6 (1105.4)	51.4 (890.3)	0.081	0.45156	++ (o)	
MPoG	50 (50)	49	51	-50.0 (374.9)	-74.2 (365.0)	0.066	0.54466	++ (o)	
POrG	51 (51)	54	49	-127.2 (678.0)	-85.1 (691.7)	0.061	0.56966	++ (o)	
PoG	52 (52)	52	52	225.5 (2211.5)	100.1 (1992.6)	0.059	0.58233	++ (o)	
CerebellarVermalLobulesI-V	49 (49)	50	48	-88.1 (588.0)	-56.4 (617.9)	0.053	0.62643	++ (o)	
CerebellumWhiteMatter	49 (49)	42	56	1633.9 (8349.6)	1274.0 (6934.3)	0.047	0.66544	++ (o)	
LiG	49 (49)	49	49	26.2 (2037.7)	-61.0 (2098.3)	0.042	0.69632	++ (o)	
CerebralWhiteMatter	46 (46)	47	46	4332.4 (25782.2)	3333.3 (25445.0)	0.039	0.71838	++ (o)	
4thVentricle	48 (48)	45	52	-6.9 (571.0)	-31.1 (681.2)	0.039	0.72152	++ (o)	
TTG	46 (46)	47	45	-48.6 (506.4)	-36.1 (443.2)	0.026	0.80827	++ (o)	
MFC	48 (48)	49	46	-145.6 (598.9)	-131.3 (575.9)	0.024	0.82244	++ (o)	
Pallidum	47 (47)	48	47	9.3 (383.8)	2.4 (364.6)	0.018	0.86516	++ (o)	
Cun	47 (47)	46	48	129.0 (1484.4)	102.5 (1449.4)	0.018	0.86737	++ (o)	
MORg	46 (46)	47	45	4.1 (835.1)	-8.5 (900.9)	0.015	0.89268	++ (o)	
CO	46 (46)	46	46	-96.2 (783.1)	-107.0 (830.1)	0.013	0.90184	++ (o)	
CerebellarVermalLobulesVIII-X	46 (46)	47	45	11.7 (406.8)	17.1 (421.7)	0.013	0.90479	++ (o)	
AORg	46 (46)	46	46	-73.3 (451.4)	-78.7 (502.9)	0.011	0.91793	++ (o)	

E.4 Longitudinal analysis (baseline→m12, AD vs. HC)

Table E.4: Classification results in % (6-fold CV, LDA 100 runs, RF/SVM 20 runs) for separating AD from CN. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses. Sorted by effect size.

AD patients (N = 195, Positives ^P) vs. Healthy Controls (N = 290, Negatives ^N) (longitudinal analysis, bl→m12)								
structure	ACC (bACC)	SENS	SPEC	mean (SD) [%] ^P	mean (SD) [%] ^N	effect size (d)	p-value	sig. (corr.)
RandomForest (all features)	85 (84)	78	90					
SVM (all features)	84 (82)	71	93					
Age	54 (54)	53	54	75.3 (7.7)	74.8 (5.6)	0.072	0.43678	o (o)
Gender (female = 0, male = 1)	48 (48)	50	46	0.5 (0.5)	0.5 (0.5)	0.042	0.64808	o (o)
Ventricles	75 (74)	64	83	7.2 (4.6)	2.8 (3.0)	1.202	<0.00001	++ (++)
BrainTissue	70 (70)	67	72	-1.3 (1.1)	-0.5 (0.8)	0.862	<0.00001	++ (++)
WhiteMatter	73 (73)	68	77	-0.8 (1.0)	-0.2 (0.7)	0.736	<0.00001	++ (++)
CorticalGreyMatter	65 (64)	62	67	-1.9 (2.3)	-0.8 (1.4)	0.585	<0.00001	++ (++)
Brain	64 (64)	63	65	-0.9 (1.0)	-0.4 (0.7)	0.574	<0.00001	++ (++)
DeepGreyMatter	66 (64)	57	71	-1.4 (1.7)	-0.8 (1.3)	0.455	<0.00001	++ (++)
Hippocampus	80 (78)	67	88	-4.8 (3.7)	-1.1 (1.7)	1.400	<0.00001	++ (++)
InfLatVent	79 (77)	69	86	7.5 (5.5)	1.8 (3.3)	1.334	<0.00001	++ (++)
LeftHippocampus	80 (78)	70	87	-4.9 (4.2)	-1.1 (1.7)	1.287	<0.00001	++ (++)
MTG	76 (74)	67	82	-3.8 (2.8)	-1.1 (1.5)	1.274	<0.00001	++ (++)
RightInfLatVent	77 (75)	65	84	7.4 (6.4)	1.8 (3.3)	1.185	<0.00001	++ (++)
RightLateralVentricle	75 (73)	63	83	7.5 (5.1)	2.8 (3.1)	1.170	<0.00001	++ (++)
LeftInfLatVent	78 (76)	70	83	7.5 (5.8)	1.7 (4.4)	1.166	<0.00001	++ (++)
LateralVentricle	74 (72)	62	83	7.6 (4.9)	3.0 (3.2)	1.160	<0.00001	++ (++)
RightHippocampus	78 (75)	62	89	-4.8 (4.5)	-1.1 (2.1)	1.138	<0.00001	++ (++)
LeftLateralVentricle	74 (72)	63	81	7.6 (4.9)	3.1 (3.4)	1.085	<0.00001	++ (++)
PP	77 (76)	70	81	-6.7 (5.5)	-2.0 (3.5)	1.075	<0.00001	++ (++)
CerebralWhiteMatter	76 (74)	67	82	-0.9 (0.9)	-0.2 (0.6)	0.914	<0.00001	++ (++)
PT	70 (69)	62	76	-5.8 (5.3)	-2.1 (3.1)	0.898	<0.00001	++ (++)
ITG	72 (71)	67	75	-2.7 (2.7)	-1.0 (1.3)	0.866	<0.00001	++ (++)
FuG	69 (68)	63	73	-1.7 (1.7)	-0.5 (1.1)	0.852	<0.00001	++ (++)
Amygdala	76 (74)	66	83	-3.3 (5.0)	-0.5 (2.0)	0.787	<0.00001	++ (++)
EntA	69 (69)	65	72	-3.7 (4.1)	-1.2 (2.8)	0.747	<0.00001	++ (++)
TMP	72 (71)	67	74	-4.3 (4.6)	-1.8 (2.8)	0.700	<0.00001	++ (++)
LeftAmygdala	73 (70)	58	82	-3.6 (6.2)	-0.5 (2.5)	0.699	<0.00001	++ (++)
SMG	64 (64)	61	67	-2.9 (3.5)	-1.1 (2.1)	0.674	<0.00001	++ (++)
RightAmygdala	74 (73)	67	79	-3.3 (6.1)	-0.5 (2.5)	0.631	<0.00001	++ (++)
TTG	67 (66)	62	70	-4.6 (5.3)	-1.9 (3.6)	0.604	<0.00001	++ (++)
AnG	63 (62)	58	66	-2.6 (3.7)	-0.9 (2.2)	0.598	<0.00001	++ (++)
Alns	65 (64)	59	68	-2.0 (2.7)	-0.7 (1.9)	0.562	<0.00001	++ (++)
MSFG	60 (59)	57	61	-2.4 (3.4)	-0.7 (2.6)	0.551	<0.00001	++ (++)
PHG	61 (60)	57	63	-1.7 (2.5)	-0.6 (1.8)	0.521	<0.00001	++ (++)
Caudate	64 (63)	60	67	-4.7 (5.6)	-2.3 (3.7)	0.512	<0.00001	++ (++)
IOG	64 (63)	58	69	-2.2 (3.4)	-0.9 (2.2)	0.442	<0.00001	++ (++)
SMC	60 (60)	58	62	-1.9 (3.2)	-0.7 (2.5)	0.434	<0.00001	++ (++)
3rdVentricle	63 (62)	57	67	4.1 (4.4)	2.4 (3.6)	0.427	<0.00001	++ (++)
OrIFG	60 (60)	59	62	-2.8 (3.7)	-1.3 (3.6)	0.425	<0.00001	++ (++)
FO	58 (58)	54	61	-1.9 (3.8)	-0.6 (2.3)	0.406	0.00001	++ (+)
PCu	60 (60)	59	60	-1.2 (3.1)	-0.2 (2.0)	0.395	0.00002	++ (+)
PCgG	57 (56)	53	59	-0.8 (1.9)	-0.1 (1.7)	0.385	0.00004	++ (+)
MFG	60 (60)	61	59	-2.0 (3.8)	-0.9 (2.1)	0.384	0.00004	++ (+)
ACg	60 (60)	58	62	-1.6 (3.4)	-0.5 (2.3)	0.384	0.00004	++ (+)
SFG	59 (59)	57	61	-2.1 (4.1)	-0.9 (2.4)	0.354	0.00015	++ (+)
MORg	59 (58)	54	62	-1.9 (4.5)	-0.6 (3.0)	0.352	0.00016	++ (+)
TrIFG	60 (59)	56	62	-2.4 (3.6)	-1.4 (2.5)	0.348	0.00019	++ (+)
ThalamusProper	72 (71)	63	79	-2.2 (2.3)	-1.1 (3.9)	0.331	0.00038	++ (+)
BrainStem	55 (53)	45	61	-0.9 (1.3)	-0.6 (0.9)	0.310	0.00089	++ (o)
MFC	60 (60)	57	63	-1.8 (4.5)	-0.6 (3.4)	0.303	0.00116	++ (o)
MOG	58 (58)	56	60	-1.7 (4.1)	-0.7 (2.6)	0.299	0.00133	++ (o)
4thVentricle	58 (57)	51	62	2.4 (6.2)	0.9 (4.5)	0.295	0.00152	++ (o)
CSF	56 (55)	54	57	1.1 (3.7)	0.1 (3.2)	0.289	0.00192	++ (o)
OFuG	55 (55)	53	56	-1.0 (3.3)	-0.2 (2.5)	0.277	0.00294	++ (o)
SPL	60 (59)	58	60	-2.0 (4.0)	-1.0 (3.0)	0.274	0.00320	++ (o)
OpIFG	56 (56)	54	58	-2.1 (4.4)	-1.2 (2.8)	0.270	0.00369	++ (o)
PORg	56 (56)	56	56	-1.3 (3.3)	-0.6 (2.4)	0.236	0.01130	++ (o)
CerebellumExterior	60 (58)	49	68	-1.3 (2.6)	-0.8 (2.1)	0.229	0.01382	++ (o)
MPrG	55 (54)	53	56	-1.0 (3.9)	-0.3 (3.2)	0.194	0.03708	++ (o)
LORg	56 (56)	55	57	-2.7 (4.8)	-1.7 (5.1)	0.192	0.03847	++ (o)
SOG	58 (57)	56	59	-1.6 (4.6)	-0.8 (3.5)	0.187	0.04441	++ (o)
SCA	58 (58)	59	57	-2.6 (10.0)	-1.3 (4.5)	0.176	0.05788	o (o)
LiG	55 (56)	60	52	-0.3 (1.9)	-0.0 (1.6)	0.175	0.05999	o (o)
PrG	55 (55)	55	55	-1.4 (4.1)	-0.8 (2.9)	0.167	0.07119	o (o)
MCgG	54 (54)	55	54	-0.5 (2.6)	-0.1 (2.5)	0.158	0.08770	o (o)
Plns	55 (54)	50	58	-1.4 (3.4)	-1.0 (2.5)	0.148	0.11160	o (o)
Pallidum	58 (52)	21	83	0.3 (2.9)	-0.1 (2.2)	0.141	0.12864	o (o)
FRP	58 (57)	51	62	-3.9 (15.3)	-2.3 (14.0)	0.111	0.23076	o (o)
AccumbensArea	56 (49)	13	85	3.0 (50.3)	-0.5 (7.0)	0.109	0.23774	o (o)
VentralDC	53 (53)	54	53	-0.7 (2.0)	-0.5 (1.6)	0.107	0.24813	o (o)
GRc	56 (56)	55	57	-2.1 (6.7)	-1.5 (7.0)	0.097	0.29359	o (o)
CerebellarVermalLobulesVIII-X	54 (53)	52	55	-0.6 (2.8)	-0.4 (2.0)	0.096	0.30009	o (o)
PO	52 (52)	53	51	-0.1 (3.9)	-0.4 (2.7)	0.089	0.33509	o (o)
Cun	50 (51)	53	48	-0.3 (3.5)	-0.6 (3.9)	0.087	0.34872	o (o)
AORg	53 (52)	49	55	-1.5 (5.1)	-1.1 (3.8)	0.086	0.35288	o (o)
CerebellarVermalLobulesI-V	49 (49)	47	51	-0.5 (2.5)	-0.3 (1.8)	0.083	0.37161	o (o)
OCP	53 (51)	46	57	-3.1 (10.4)	-2.2 (11.5)	0.081	0.38148	o (o)
Putamen	54 (49)	28	71	-0.1 (1.4)	-0.0 (0.8)	0.075	0.41924	o (o)
MPoG	50 (50)	50	50	-0.9 (8.4)	-0.5 (5.0)	0.057	0.54043	o (o)
BasalForebrain	51 (50)	47	53	-2.1 (12.9)	-1.6 (9.1)	0.046	0.62158	o (o)
PoG	50 (50)	52	49	-1.1 (4.8)	-0.9 (3.3)	0.045	0.62950	o (o)
CerebellumWhiteMatter	53 (50)	31	68	0.6 (7.2)	0.4 (5.3)	0.036	0.69917	o (o)
CerebellarVermalLobulesVI-VII	49 (48)	47	50	0.2 (4.3)	0.1 (3.7)	0.029	0.75272	o (o)
CO	48 (48)	47	48	-0.5 (3.7)	-0.6 (2.6)	0.025	0.78620	o (o)
Calc	48 (47)	45	50	-0.1 (3.5)	-0.2 (3.1)	0.019	0.83368	o (o)
STG	48 (48)	46	50	-0.6 (2.6)	-0.6 (1.8)	0.010	0.91377	o (o)

E.5 Longitudinal analysis (baseline→m12, pMCI vs. sMCI)

Table E.5: Classification results in % (6-fold CV, LDA 100 runs, RF/SVM 20 runs) for separating progressive from stable MCI. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses. Sorted by effect size.

progressive MCI (N = 168, Positives ^P) vs. stable MCI (N = 149, Negatives ^N) (longitudinal analysis, bl→m12)								
structure	ACC (hACC)	SENS	SPEC	mean (SD) [%] ^P	mean (SD) [%] ^N	effect size (d)	p-value	sig. (corr.)
RandomForest (all features)	74 (73)	77	70					
SVM (all features)	74 (74)	72	75					
Age	49 (49)	46	51	73.6 (7.6)	74.0 (7.4)	0.050	0.65912	o (o)
Gender (female = 0, male = 1)	45 (45)	45	45	0.6 (0.5)	0.6 (0.5)	0.001	0.99199	o (o)
Ventricles	68 (69)	63	74	6.8 (4.3)	3.4 (3.1)	0.890	<0.00001	++ (++)
BrainTissue	63 (63)	64	63	-1.2 (1.0)	-0.6 (0.8)	0.652	<0.00001	++ (++)
CorticalGreyMatter	64 (64)	67	61	-1.8 (1.7)	-0.8 (1.3)	0.608	<0.00001	++ (++)
Brain	58 (58)	62	54	-0.8 (0.8)	-0.4 (0.6)	0.495	0.00002	++ (+)
DeepGreyMatter	64 (64)	62	66	-1.3 (1.2)	-0.7 (1.3)	0.428	0.00017	++ (+)
WhiteMatter	63 (63)	61	64	-0.6 (0.9)	-0.4 (0.7)	0.310	0.00615	+ (o)
Hippocampus	67 (67)	59	76	-4.1 (3.2)	-1.7 (2.2)	0.867	<0.00001	++ (++)
LateralVentricle	67 (68)	61	75	7.1 (4.6)	3.6 (3.4)	0.866	<0.00001	++ (++)
InfLatVent	69 (69)	63	75	6.7 (5.6)	2.6 (3.8)	0.845	<0.00001	++ (++)
LeftHippocampus	67 (68)	60	75	-4.2 (3.3)	-1.7 (2.5)	0.845	<0.00001	++ (++)
LeftLateralVentricle	66 (67)	60	73	7.2 (4.8)	3.6 (3.5)	0.836	<0.00001	++ (++)
RightLateralVentricle	67 (67)	62	72	7.0 (4.6)	3.6 (3.6)	0.828	<0.00001	++ (++)
MTG	70 (70)	65	75	-3.1 (2.4)	-1.4 (1.8)	0.807	<0.00001	++ (++)
ITG	67 (67)	64	70	-2.5 (2.0)	-1.1 (1.6)	0.793	<0.00001	++ (++)
LeftInfLatVent	68 (69)	62	76	6.7 (6.2)	2.4 (4.3)	0.783	<0.00001	++ (++)
EntA	67 (67)	67	68	-4.0 (3.7)	-1.4 (3.3)	0.749	<0.00001	++ (++)
PT	68 (68)	63	73	-5.5 (4.5)	-2.6 (3.3)	0.727	<0.00001	++ (++)
RightInfLatVent	64 (64)	58	71	6.7 (6.2)	2.7 (4.4)	0.726	<0.00001	++ (++)
FuG	66 (66)	63	69	-1.5 (1.6)	-0.5 (1.1)	0.694	<0.00001	++ (++)
PP	66 (67)	64	70	-5.8 (4.6)	-2.9 (3.9)	0.692	<0.00001	++ (++)
RightHippocampus	65 (65)	57	73	-4.0 (3.9)	-1.7 (2.8)	0.677	<0.00001	++ (++)
3rdVentricle	64 (64)	61	67	4.7 (3.8)	2.5 (3.7)	0.590	<0.00001	++ (++)
Amygdala	65 (66)	58	73	-3.2 (4.1)	-1.2 (2.8)	0.577	<0.00001	++ (++)
RightAmygdala	64 (64)	56	73	-3.7 (5.5)	-1.1 (3.3)	0.567	<0.00001	++ (++)
TTG	61 (62)	61	62	-4.4 (4.2)	-2.2 (3.5)	0.558	<0.00001	++ (++)
TMP	64 (64)	64	64	-3.9 (3.7)	-2.2 (3.0)	0.524	<0.00001	++ (++)
Caudate	62 (62)	62	62	-4.8 (5.2)	-2.4 (3.9)	0.510	<0.00001	++ (++)
SPL	64 (64)	67	62	-2.1 (2.8)	-0.7 (2.9)	0.502	0.00001	++ (++)
SMG	62 (62)	61	62	-2.6 (2.9)	-1.3 (2.5)	0.489	0.00002	++ (+)
PHG	58 (58)	57	60	-1.7 (2.2)	-0.7 (1.9)	0.478	0.00003	++ (+)
MOG	58 (58)	51	65	-2.0 (3.2)	-0.8 (2.1)	0.442	0.00011	++ (+)
AnG	61 (61)	59	63	-2.1 (2.9)	-1.0 (2.2)	0.432	0.00015	++ (+)
MSFG	59 (59)	58	59	-2.2 (3.0)	-1.0 (2.4)	0.428	0.00017	++ (+)
AIns	56 (56)	55	57	-1.5 (2.2)	-0.7 (1.8)	0.387	0.00065	++ (o)
LeftAmygdala	62 (63)	53	72	-2.8 (4.5)	-1.3 (3.5)	0.376	0.00095	++ (o)
TrIFG	60 (60)	57	63	-2.6 (3.0)	-1.6 (2.5)	0.366	0.00127	+ (o)
BrainStem	62 (61)	68	55	-0.8 (0.9)	-0.3 (1.7)	0.362	0.00144	+ (o)
IOG	56 (57)	52	62	-1.8 (2.5)	-1.0 (2.1)	0.347	0.00220	+ (o)
SMC	54 (54)	53	55	-1.7 (2.9)	-0.8 (2.4)	0.333	0.00328	+ (o)
OCF	57 (57)	53	61	-4.4 (13.1)	-0.5 (10.0)	0.332	0.00337	+ (o)
OrIFG	55 (55)	51	59	-3.0 (3.4)	-1.9 (2.9)	0.332	0.00339	+ (o)
SFG	58 (58)	55	61	-2.1 (3.0)	-1.1 (2.9)	0.318	0.00498	+ (o)
MFG	58 (58)	58	58	-1.9 (2.7)	-1.1 (2.4)	0.317	0.00516	+ (o)
PCgG	58 (58)	52	63	-0.7 (1.8)	-0.1 (1.6)	0.314	0.00562	+ (o)
PCu	55 (56)	53	58	-1.0 (2.5)	-0.4 (1.9)	0.295	0.00923	+ (o)
CerebralWhiteMatter	62 (63)	60	65	-0.6 (0.8)	-0.4 (0.7)	0.284	0.01201	+ (o)
FO	55 (55)	51	60	-1.9 (3.3)	-1.0 (2.8)	0.284	0.01224	+ (o)
ThalamusProper	65 (65)	55	76	-2.6 (7.3)	-1.1 (2.8)	0.273	0.01582	+ (o)
LiG	55 (55)	54	56	-0.4 (1.5)	-0.0 (1.4)	0.268	0.01794	+ (o)
OFuG	54 (54)	51	57	-1.2 (2.9)	-0.5 (2.8)	0.249	0.02748	+ (o)
MOrG	55 (55)	53	56	-1.6 (2.6)	-0.9 (3.1)	0.245	0.02999	+ (o)
LOrG	54 (54)	53	55	-2.4 (4.1)	-1.3 (4.7)	0.245	0.03029	+ (o)
Cun	53 (53)	53	54	-0.5 (3.6)	0.3 (3.0)	0.231	0.04104	+ (o)
SOG	55 (55)	55	55	-1.6 (4.7)	-0.6 (3.9)	0.229	0.04284	+ (o)
SCA	57 (57)	58	56	-2.9 (6.8)	-1.4 (7.1)	0.225	0.04654	+ (o)
MFC	53 (53)	51	56	-1.4 (3.9)	-0.5 (4.1)	0.222	0.04980	+ (o)
OpIFG	58 (58)	60	56	-1.7 (3.2)	-1.1 (2.3)	0.217	0.05489	o (o)
GRc	53 (53)	56	51	-1.9 (5.4)	-0.5 (7.2)	0.215	0.05712	o (o)
AOrG	52 (53)	47	58	-2.1 (4.1)	-1.2 (4.0)	0.210	0.06272	o (o)
POrG	55 (55)	55	54	-0.8 (2.9)	-0.2 (2.7)	0.210	0.06305	o (o)
CSF	53 (53)	52	55	1.0 (3.4)	0.4 (3.6)	0.186	0.10005	o (o)
PrG	55 (55)	59	51	-1.1 (2.9)	-0.6 (3.1)	0.173	0.12492	o (o)
Putamen	48 (50)	18	82	-0.4 (3.5)	0.1 (1.2)	0.170	0.13126	o (o)
Plns	54 (54)	53	55	-0.9 (2.8)	-0.5 (1.9)	0.168	0.13708	o (o)
AccumbensArea	55 (56)	38	75	-1.3 (9.7)	0.1 (7.8)	0.167	0.13805	o (o)
ACg	52 (53)	47	58	-1.3 (2.7)	-0.9 (2.7)	0.160	0.15511	o (o)
CerebellarVermalLobulesI-V	55 (55)	51	59	-0.6 (1.3)	-0.3 (2.1)	0.153	0.17461	o (o)
CO	52 (52)	48	56	-0.7 (2.9)	-0.3 (2.3)	0.150	0.18303	o (o)
BasalForebrain	54 (54)	55	53	-1.5 (10.3)	-0.1 (9.8)	0.138	0.22106	o (o)
CerebellumExterior	55 (56)	52	59	-1.0 (1.6)	-0.8 (2.1)	0.137	0.22484	o (o)
CerebellumWhiteMatter	56 (55)	66	44	-0.3 (3.3)	0.2 (4.1)	0.128	0.25500	o (o)
STG	52 (52)	57	48	-0.6 (2.4)	-0.3 (1.8)	0.127	0.26171	o (o)
Calc	52 (52)	54	51	-0.3 (3.4)	0.0 (2.4)	0.126	0.26340	o (o)
MCgG	49 (48)	51	46	-0.3 (2.2)	-0.0 (2.2)	0.118	0.29487	o (o)
PoG	52 (52)	55	49	-1.1 (3.9)	-0.8 (2.9)	0.095	0.39793	o (o)
MPoG	50 (50)	54	45	-0.1 (7.2)	0.4 (5.0)	0.091	0.42142	o (o)
FRP	48 (48)	54	41	-2.4 (9.7)	-1.6 (16.5)	0.056	0.62168	o (o)
CerebellarVermalLobulesVIII-X	48 (48)	47	49	-0.4 (2.4)	-0.5 (2.8)	0.056	0.62187	o (o)
PO	49 (49)	51	48	-0.1 (3.7)	0.0 (2.8)	0.052	0.64691	o (o)
Pallidum	49 (50)	38	62	-0.1 (3.0)	0.0 (3.0)	0.051	0.65324	o (o)
CerebellarVermalLobulesVI-VII	47 (47)	46	47	-0.1 (4.1)	0.1 (3.7)	0.050	0.65863	o (o)
MPgG	47 (47)	47	46	-0.5 (2.8)	-0.5 (2.8)	0.028	0.80310	o (o)
VentralDC	49 (49)	51	47	-0.6 (1.9)	-0.5 (3.1)	0.021	0.85372	o (o)
4thVentricle	46 (47)	45	48	2.1 (7.0)	2.0 (5.5)	0.018	0.87648	o (o)

E.6 Longitudinal analysis (baseline→m24, AD vs. HC)

Table E.6: Classification results in % (6-fold CV, LDA 100 runs, RF/SVM 20 runs) for separating AD from CN. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses. Sorted by effect size.

AD patients (N=117, Positives ^P) vs. Healthy Controls (N=168, Negatives ^N) (longitudinal analysis, bl→m24)									
structure	ACC (bACC)	SENS	SPEC	mean (SD) [%] ^P	mean (SD) [%] ^N	effect size (d)	p-value	sig. (corr.)	
RandomForest (all features)	89 (88)	81	94						
SVM (all features)	89 (88)	80	95						
Age	47 (47)	46	47	75.5 (7.7)	75.9 (5.2)	0.061	0.61542		o (o)
Gender (female=0, male=1)	49 (48)	44	52	0.5 (0.5)	0.6 (0.5)	0.059	0.62275		o (o)
Ventricles	83 (81)	70	92	14.7 (7.9)	5.8 (3.7)	1.531	<0.00001	++	(++)
BrainTissue	80 (78)	71	86	-2.6 (1.3)	-1.1 (0.8)	1.353	<0.00001	++	(++)
WhiteMatter	75 (74)	65	82	-1.8 (1.4)	-0.6 (0.8)	1.121	<0.00001	++	(++)
CorticalGreyMatter	73 (73)	69	77	-3.5 (2.4)	-1.6 (1.7)	0.928	<0.00001	++	(++)
Brain	68 (67)	65	70	-1.5 (1.0)	-0.8 (0.7)	0.847	<0.00001	++	(++)
DeepGreyMatter	67 (66)	63	69	-2.3 (1.8)	-1.3 (1.3)	0.604	<0.00001	++	(++)
Hippocampus	87 (85)	77	94	-10.2 (6.2)	-2.0 (2.4)	1.880	<0.00001	++	(++)
InfLatVent	84 (83)	76	90	15.9 (8.8)	3.5 (4.5)	1.871	<0.00001	++	(++)
RightInfLatVent	83 (82)	73	90	16.1 (9.7)	3.5 (4.9)	1.728	<0.00001	++	(++)
LeftHippocampus	86 (84)	77	92	-9.9 (6.6)	-1.9 (2.7)	1.701	<0.00001	++	(++)
LeftInfLatVent	80 (78)	69	88	15.5 (9.5)	3.5 (5.2)	1.652	<0.00001	++	(++)
MTG	83 (81)	70	92	-6.8 (4.1)	-2.0 (1.6)	1.651	<0.00001	++	(++)
RightHippocampus	84 (81)	69	94	-10.5 (7.6)	-2.0 (2.6)	1.617	<0.00001	++	(++)
ITG	83 (81)	71	91	-5.3 (3.3)	-1.6 (1.4)	1.603	<0.00001	++	(++)
FuG	85 (83)	72	93	-3.7 (2.7)	-0.7 (0.9)	1.588	<0.00001	++	(++)
LeftLateralVentricle	82 (80)	71	89	15.6 (8.2)	6.5 (4.4)	1.470	<0.00001	++	(++)
PP	80 (79)	73	85	-10.9 (6.1)	-3.9 (3.5)	1.466	<0.00001	++	(++)
LateralVentricle	80 (78)	65	90	15.4 (8.5)	6.3 (4.0)	1.454	<0.00001	++	(++)
TMP	84 (82)	75	90	-8.2 (5.1)	-2.8 (2.5)	1.430	<0.00001	++	(++)
RightLateralVentricle	80 (77)	63	92	15.1 (9.3)	6.2 (4.0)	1.331	<0.00001	++	(++)
PT	78 (77)	70	83	-10.0 (6.2)	-3.8 (3.5)	1.276	<0.00001	++	(++)
EntA	80 (79)	72	86	-7.0 (5.5)	-2.2 (3.1)	1.151	<0.00001	++	(++)
CerebralWhiteMatter	75 (74)	65	83	-1.9 (1.3)	-0.7 (0.9)	1.138	<0.00001	++	(++)
Amygdala	81 (79)	70	89	-7.3 (6.7)	-1.7 (3.7)	1.076	<0.00001	++	(++)
RightAmygdala	79 (77)	69	85	-7.6 (7.0)	-1.8 (4.2)	1.050	<0.00001	++	(++)
SMG	71 (70)	61	78	-5.7 (4.6)	-2.2 (2.6)	0.968	<0.00001	++	(++)
AIms	73 (73)	67	78	-3.6 (3.4)	-1.1 (2.0)	0.965	<0.00001	++	(++)
3rdVentricle	70 (69)	65	74	8.4 (5.7)	4.1 (4.0)	0.904	<0.00001	++	(++)
PHG	73 (71)	65	78	-3.4 (3.7)	-0.9 (2.0)	0.901	<0.00001	++	(++)
LeftAmygdala	79 (77)	65	89	-7.1 (8.7)	-1.7 (3.9)	0.863	<0.00001	++	(++)
ThalamusProper	78 (77)	67	86	-4.5 (3.5)	-2.0 (2.8)	0.838	<0.00001	++	(++)
MSFG	71 (70)	68	73	-4.7 (3.9)	-2.1 (2.8)	0.789	<0.00001	++	(++)
PCgG	67 (65)	57	73	-2.1 (2.7)	-0.6 (1.7)	0.729	<0.00001	++	(++)
AnG	69 (67)	60	75	-4.5 (4.4)	-2.0 (2.7)	0.718	<0.00001	++	(++)
FO	69 (68)	58	77	-4.8 (5.5)	-1.8 (3.0)	0.706	<0.00001	++	(++)
TTG	70 (70)	65	74	-7.0 (6.8)	-3.4 (3.3)	0.705	<0.00001	++	(++)
IOG	64 (63)	57	70	-3.8 (3.5)	-1.8 (2.5)	0.671	<0.00001	++	(++)
POrG	61 (61)	63	59	-2.7 (3.0)	-0.9 (2.7)	0.628	<0.00001	++	(++)
PCu	69 (67)	58	76	-2.7 (3.7)	-1.0 (2.2)	0.601	<0.00001	++	(++)
SMC	62 (62)	57	66	-3.9 (3.9)	-2.0 (2.5)	0.595	<0.00001	++	(++)
OrIFG	63 (62)	58	67	-4.8 (4.7)	-2.5 (3.6)	0.573	<0.00001	++	(++)
MFG	61 (60)	58	63	-3.5 (3.8)	-1.9 (2.6)	0.521	0.00002	++	(+)
MORg	64 (63)	61	66	-3.3 (4.5)	-1.4 (3.1)	0.509	0.00003	++	(+)
ACg	62 (60)	50	70	-3.3 (4.0)	-1.7 (2.4)	0.505	0.00004	++	(+)
BrainStem	60 (59)	53	65	-1.5 (1.4)	-1.0 (1.0)	0.456	0.00018	++	(+)
MOG	61 (60)	58	63	-3.1 (4.0)	-1.3 (4.2)	0.452	0.00021	++	(+)
SPL	64 (64)	63	65	-3.6 (5.0)	-1.7 (3.9)	0.446	0.00025	++	(+)
TriFG	62 (60)	54	67	-3.8 (3.5)	-2.4 (2.9)	0.443	0.00028	++	(+)
LORg	62 (61)	53	69	-4.9 (6.1)	-2.5 (5.5)	0.426	0.00047	++	(+)
OpIFG	60 (59)	57	62	-3.5 (3.7)	-2.1 (2.9)	0.423	0.00051	++	(+)
CSF	59 (58)	58	59	2.0 (5.1)	0.1 (4.1)	0.412	0.00072	++	(o)
SCA	63 (63)	61	64	-5.9 (13.2)	-2.1 (6.2)	0.388	0.00141	+	(o)
MFC	61 (60)	55	65	-3.4 (5.6)	-1.7 (4.0)	0.354	0.00356	+	(o)
GrE	64 (63)	61	66	-4.3 (8.0)	-1.7 (6.9)	0.347	0.00427	+	(o)
OFuG	59 (59)	60	58	-1.6 (3.1)	-0.6 (3.0)	0.331	0.00636	+	(o)
Caudate	61 (61)	58	63	-5.9 (9.5)	-3.6 (4.4)	0.328	0.00683	+	(o)
STG	55 (55)	52	58	-1.7 (3.2)	-0.9 (2.0)	0.306	0.01145	+	(o)
PrG	58 (58)	60	57	-2.7 (3.8)	-1.8 (3.4)	0.252	0.03747	+	(o)
AORg	54 (54)	53	55	-2.6 (4.9)	-1.6 (3.4)	0.243	0.04413	+	(o)
4thVentricle	60 (59)	55	63	3.8 (7.0)	2.3 (5.8)	0.240	0.04731	+	(o)
SFG	57 (57)	59	55	-2.7 (4.0)	-2.0 (3.2)	0.209	0.08426		o (o)
CerebellumExterior	60 (60)	56	63	-1.9 (2.9)	-1.4 (2.0)	0.200	0.09756		o (o)
LiG	53 (53)	56	51	-0.8 (3.1)	-0.3 (2.3)	0.193	0.10963		o (o)
Putamen	59 (53)	22	84	-0.4 (3.0)	0.1 (2.0)	0.192	0.11187		o (o)
MCgG	57 (57)	56	57	-1.3 (3.4)	-0.7 (2.3)	0.190	0.11579		o (o)
VentralDC	57 (57)	53	61	-1.3 (1.8)	-1.0 (2.4)	0.180	0.13700		o (o)
SOG	58 (58)	59	56	-2.6 (5.8)	-1.5 (6.6)	0.175	0.14836		o (o)
Pallidum	52 (53)	62	45	0.2 (3.5)	-0.2 (2.0)	0.164	0.17299		o (o)
CerebellarVermalLobulesI-V	54 (52)	43	62	-1.3 (3.5)	-0.7 (4.4)	0.164	0.17424		o (o)
Plns	54 (53)	52	55	-1.7 (4.1)	-1.2 (2.7)	0.141	0.24371		o (o)
PO	52 (51)	48	54	-1.6 (4.6)	-1.2 (3.2)	0.109	0.36791		o (o)
FRP	58 (57)	50	63	-5.3 (18.9)	-3.6 (14.8)	0.107	0.37505		o (o)
CO	53 (52)	48	56	-1.9 (4.2)	-1.6 (3.1)	0.105	0.38495		o (o)
Cun	54 (53)	49	58	-0.1 (4.8)	-0.6 (7.4)	0.077	0.52528		o (o)
MPoG	48 (48)	44	51	-1.1 (8.5)	-1.6 (5.5)	0.074	0.54109		o (o)
BasalForebrain	48 (48)	48	48	-2.6 (16.4)	-3.4 (9.7)	0.062	0.60717		o (o)
AccumbensArea	58 (52)	18	86	2.6 (60.0)	0.1 (16.1)	0.061	0.61555		o (o)
CerebellumWhiteMatter	49 (52)	69	35	0.1 (7.1)	0.4 (4.8)	0.057	0.63523		o (o)
MPrG	48 (48)	46	50	-1.8 (3.9)	-1.7 (3.7)	0.043	0.71987		o (o)
CerebellarVermalLobulesVIII-X	46 (47)	53	40	-1.0 (3.0)	-1.2 (6.2)	0.039	0.74661		o (o)
PoG	48 (47)	47	48	-1.9 (5.0)	-1.7 (4.1)	0.027	0.82435		o (o)
OCF	49 (49)	48	49	-2.4 (17.0)	-2.1 (17.1)	0.015	0.90207		o (o)
CerebellarVermalLobulesVI-VII	47 (47)	46	48	-0.3 (4.7)	-0.4 (4.5)	0.014	0.90530		o (o)
Calc	47 (46)	43	49	-0.8 (4.8)	-0.8 (3.3)	0.013	0.91102		o (o)

E.7 Longitudinal analysis (baseline→m24, pMCI vs. sMCI)

Table E.7: Classification results in % (6-fold CV, LDA 100 runs, RF/SVM 20 runs) for separating progressive from stable MCI. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses. Sorted by effect size.

progressive MCI (N = 140, Positives ^P) vs. stable MCI (N = 107, Negatives ^N) (longitudinal analysis, bl→m24)								
structure	ACC (hACC)	SENS	SPEC	mean (SD) [%] ^P	mean (SD) [%] ^N	effect size (d)	p-value	sig. (corr.)
RandomForest (all features)	79 (78)	82	74					
SVM (all features)	76 (76)	76	76					
Age	54 (54)	51	58	73.3 (7.8)	74.7 (7.5)	0.176	0.17108	o (o)
Gender (female = 0, male = 1)	46 (47)	40	54	0.6 (0.5)	0.6 (0.5)	0.053	0.67739	o (o)
Ventricles	71 (72)	65	80	13.7 (7.4)	7.1 (5.6)	0.989	<0.00001	++ (++)
BrainTissue	68 (69)	62	76	-2.3 (1.5)	-1.2 (1.0)	0.845	<0.00001	++ (++)
DeepGreyMatter	66 (67)	64	69	-2.1 (1.4)	-1.2 (1.0)	0.714	<0.00001	++ (++)
CorticalGreyMatter	65 (66)	61	71	-3.2 (2.5)	-1.7 (1.6)	0.707	<0.00001	++ (++)
Brain	63 (63)	60	67	-1.5 (1.1)	-0.8 (0.7)	0.707	<0.00001	++ (++)
WhiteMatter	64 (65)	55	74	-1.6 (1.3)	-0.8 (1.0)	0.661	<0.00001	++ (++)
FuG	75 (76)	69	83	-3.1 (2.1)	-1.0 (1.4)	1.141	<0.00001	++ (++)
Hippocampus	73 (74)	68	80	-8.9 (5.1)	-3.7 (3.8)	1.118	<0.00001	++ (++)
LeftHippocampus	74 (74)	72	77	-9.3 (5.6)	-3.8 (4.2)	1.083	<0.00001	++ (++)
EntA	71 (72)	66	77	-7.4 (4.7)	-2.8 (3.6)	1.070	<0.00001	++ (++)
InflLatVent	70 (72)	62	81	14.6 (8.9)	6.2 (6.6)	1.053	<0.00001	++ (++)
MTG	72 (74)	64	83	-6.1 (3.9)	-2.6 (2.6)	1.016	<0.00001	++ (++)
LeftInflLatVent	68 (69)	60	79	14.7 (9.4)	6.4 (6.8)	0.987	<0.00001	++ (++)
PT	68 (68)	62	75	-9.7 (5.6)	-4.7 (4.2)	0.971	<0.00001	++ (++)
LateralVentricle	70 (71)	64	78	14.4 (7.9)	7.5 (6.0)	0.960	<0.00001	++ (++)
ITG	72 (73)	66	80	-4.7 (2.9)	-2.1 (2.2)	0.958	<0.00001	++ (++)
LeftLateralVentricle	69 (70)	62	77	14.6 (8.2)	7.6 (6.1)	0.956	<0.00001	++ (++)
RightInflLatVent	66 (68)	60	76	14.3 (10.1)	5.9 (7.4)	0.923	<0.00001	++ (++)
RightHippocampus	71 (72)	65	79	-8.6 (5.7)	-3.7 (4.6)	0.920	<0.00001	++ (++)
RightLateralVentricle	69 (70)	61	79	14.1 (8.1)	7.5 (6.3)	0.907	<0.00001	++ (++)
Amygdala	67 (68)	61	76	-6.4 (5.2)	-2.9 (3.5)	0.770	<0.00001	++ (++)
TTG	65 (66)	62	70	-8.0 (5.3)	-4.4 (4.1)	0.752	<0.00001	++ (++)
PHG	66 (68)	58	77	-3.5 (2.8)	-1.5 (2.4)	0.746	<0.00001	++ (++)
3rdVentricle	64 (65)	61	68	9.1 (6.2)	4.7 (5.5)	0.740	<0.00001	++ (++)
PP	68 (69)	65	72	-10.0 (5.7)	-5.9 (5.5)	0.728	<0.00001	++ (++)
TMP	68 (69)	64	74	-6.9 (4.5)	-3.9 (3.6)	0.721	<0.00001	++ (++)
RightAmygdala	67 (68)	62	73	-6.5 (6.3)	-2.7 (4.3)	0.688	<0.00001	++ (++)
CerebralWhiteMatter	65 (66)	58	73	-1.6 (1.3)	-0.9 (1.0)	0.635	<0.00001	++ (++)
SMG	62 (64)	54	74	-4.8 (4.0)	-2.5 (2.8)	0.633	<0.00001	++ (++)
LeftAmygdala	65 (66)	60	72	-6.4 (6.0)	-3.2 (4.2)	0.605	<0.00001	++ (++)
PCg	62 (62)	57	67	-1.7 (2.3)	-0.5 (1.6)	0.577	0.00001	++ (++)
MOG	63 (64)	55	74	-3.3 (4.1)	-1.3 (2.8)	0.570	0.00001	++ (++)
MSFG	63 (63)	59	67	-4.2 (4.0)	-2.2 (3.0)	0.551	0.00003	++ (++)
AIms	58 (59)	53	66	-2.8 (3.0)	-1.4 (2.0)	0.529	0.00005	++ (++)
SCA	65 (65)	69	61	-5.8 (8.0)	-2.1 (7.1)	0.480	0.00023	++ (++)
AnG	61 (62)	56	69	-3.9 (4.9)	-2.0 (2.8)	0.478	0.00024	++ (++)
BrainStem	58 (58)	59	57	-1.3 (1.4)	-0.6 (1.4)	0.462	0.00038	++ (++)
FO	61 (63)	52	73	-4.5 (5.6)	-2.3 (3.7)	0.462	0.00039	++ (++)
CerebellumExterior	65 (65)	64	66	-1.9 (2.0)	-1.1 (1.3)	0.434	0.00085	++ (o)
MFG	58 (59)	54	64	-3.3 (3.7)	-2.0 (2.4)	0.432	0.00090	++ (o)
PCu	61 (62)	58	65	-2.3 (3.9)	-1.0 (2.0)	0.430	0.00093	++ (o)
IOG	63 (64)	59	69	-3.1 (3.8)	-1.7 (2.4)	0.417	0.00133	++ (o)
SPL	62 (61)	63	59	-3.4 (4.7)	-1.7 (3.3)	0.408	0.00167	++ (o)
LiG	58 (57)	59	55	-0.9 (2.2)	-0.1 (2.0)	0.405	0.00181	++ (o)
MORg	59 (59)	58	60	-3.0 (2.9)	-1.8 (3.5)	0.398	0.00218	++ (o)
OCP	59 (58)	61	56	-3.9 (11.9)	0.0 (9.7)	0.352	0.00660	++ (o)
AccumbensArea	63 (65)	52	77	-2.3 (5.5)	-0.5 (5.1)	0.342	0.00827	++ (o)
ThalamusProper	66 (67)	56	79	-4.7 (8.7)	-2.4 (2.5)	0.341	0.00844	++ (o)
SMC	54 (55)	51	59	-3.3 (3.8)	-2.1 (2.8)	0.335	0.00972	++ (o)
TrIFG	58 (59)	52	65	-3.7 (3.7)	-2.7 (2.6)	0.320	0.01334	++ (o)
CSF	58 (58)	60	56	1.3 (5.1)	-0.1 (4.0)	0.318	0.01404	++ (o)
LORg	57 (58)	53	62	-4.2 (4.4)	-2.8 (5.1)	0.309	0.01673	++ (o)
OrIFG	59 (59)	53	65	-4.5 (5.0)	-3.1 (3.4)	0.309	0.01695	++ (o)
Caudate	59 (59)	57	61	-5.6 (8.1)	-3.5 (5.8)	0.302	0.01965	++ (o)
CerebellumWhiteMatter	61 (59)	73	45	-0.6 (2.7)	0.3 (3.3)	0.300	0.02030	++ (o)
STG	58 (58)	52	65	-1.5 (3.3)	-0.7 (2.1)	0.294	0.02284	++ (o)
OpIFG	58 (58)	53	63	-3.1 (4.5)	-2.0 (2.4)	0.287	0.02644	++ (o)
AORg	55 (56)	51	60	-3.2 (5.1)	-1.9 (4.0)	0.285	0.02752	++ (o)
SOG	57 (57)	56	57	-2.9 (5.7)	-1.6 (3.7)	0.269	0.03737	++ (o)
ACg	55 (55)	52	59	-2.7 (3.8)	-1.8 (3.1)	0.261	0.04334	++ (o)
SFG	59 (60)	56	64	-2.9 (3.8)	-2.0 (2.8)	0.253	0.05003	++ (o)
Calc	58 (57)	64	51	-0.7 (4.9)	0.4 (4.4)	0.241	0.06227	++ (o)
MFC	58 (60)	49	70	-3.9 (5.9)	-2.6 (4.6)	0.237	0.06573	++ (o)
PORg	55 (55)	55	55	-2.0 (3.2)	-1.3 (2.8)	0.236	0.06716	++ (o)
GRe	59 (60)	52	67	-4.4 (6.6)	-3.0 (5.7)	0.229	0.07521	++ (o)
PrG	55 (55)	56	53	-2.2 (3.7)	-1.4 (3.0)	0.229	0.07632	++ (o)
CO	54 (54)	50	59	-1.8 (4.5)	-1.0 (3.0)	0.193	0.13491	++ (o)
Plms	53 (54)	51	57	-1.8 (3.2)	-1.3 (2.4)	0.192	0.13540	++ (o)
OFuG	57 (57)	57	58	-1.5 (3.7)	-1.0 (2.3)	0.166	0.19645	++ (o)
PO	56 (56)	53	59	-1.2 (5.0)	-0.5 (3.5)	0.165	0.19971	++ (o)
FRP	56 (56)	55	57	-4.9 (13.3)	-3.0 (16.1)	0.129	0.31612	++ (o)
CerebellarVermalLobulesI-V	56 (56)	54	58	-1.1 (2.1)	-0.8 (2.1)	0.128	0.31896	++ (o)
Pallidum	48 (52)	23	81	0.4 (4.3)	-0.0 (2.1)	0.118	0.35737	++ (o)
CerebellarVermalLobulesVIII-X	51 (52)	49	55	-1.1 (2.7)	-0.8 (2.2)	0.118	0.36012	++ (o)
MCgG	51 (51)	51	50	-1.1 (2.9)	-0.8 (2.4)	0.117	0.36161	++ (o)
PoG	52 (52)	52	52	-1.7 (4.7)	-1.4 (3.2)	0.092	0.47252	++ (o)
CerebellarVermalLobulesVI-VII	51 (51)	49	52	-0.8 (4.7)	-0.4 (4.2)	0.085	0.50816	++ (o)
4thVentricle	54 (54)	50	58	2.7 (7.4)	2.2 (6.0)	0.076	0.55268	++ (o)
MPG	51 (51)	54	48	-1.5 (3.6)	-1.3 (3.0)	0.072	0.57329	++ (o)
BasalForebrain	47 (46)	48	45	-2.7 (13.1)	-1.9 (10.0)	0.066	0.60690	++ (o)
VentralDC	51 (51)	52	50	-1.1 (2.5)	-0.9 (2.3)	0.060	0.64346	++ (o)
Putamen	47 (49)	40	57	-0.1 (2.0)	-0.0 (0.7)	0.034	0.79321	++ (o)
Cun	46 (46)	45	47	-0.1 (4.3)	-0.1 (2.6)	0.003	0.98036	++ (o)
MPoG	47 (47)	46	47	-0.8 (6.7)	-0.8 (4.1)	0.001	0.99373	++ (o)

F TBI classification results

F.1 Cross-sectional analysis at acute stage

Table F.1: Classification results in % (6-fold cross-validation, LDA/RF/SVM 100 runs) obtained separating TBI patients with a severe disability from patients with moderate disability outcome based on structural volumes at the acute stage of the injury (top). Classification of patients with moderate and low disability outcome (bottom). Individual structures are sorted by effect size. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses.

Severe disability (N = 13, Positives ^P) vs. Moderate disability (N = 22, Negatives ^N) (cross-sectional analysis, acute stage)									
structure	ACC (bACC)	SENS	SPEC	mean (SD) [mm ³] ^P	mean (SD) [mm ³] ^N	effect size (d)	p-value	sig. (corr.)	
RandomForest (all features incl. asymmetry)	63 (57)	34	81						
SVM (all features incl. asymmetry)	64 (60)	46	74						
Gender (female = 0, male = 1)	66 (66)	69	64	0.3 (0.5)	0.6 (0.5)	0.673	0.06287	o (o)	
Age	64 (63)	55	70	67.2 (15.9)	58.8 (10.0)	0.672	0.06331	o (o)	
MarshallScore	57 (58)	62	55	3.9 (1.4)	3.0 (1.9)	0.510	0.15465	o (o)	
GlasgowComaScale	45 (40)	21	58	11.9 (4.6)	12.6 (3.9)	0.170	0.62934	o (o)	
InjurySeverity	38 (38)	35	40	2.8 (1.0)	2.8 (0.8)	0.083	0.81296	o (o)	
DeepGreyMatter	67 (68)	73	63	164564.8 (20276.5)	180076.7 (20887.9)	0.745	0.04082	+ (o)	
CorticalGreyMatter	67 (67)	66	68	459934.2 (63395.6)	494829.1 (39983.4)	0.701	0.05339	o (o)	
BrainTissue	55 (55)	54	56	1130166.1 (102464.8)	1181045.4 (89378.0)	0.539	0.13272	o (o)	
Ventricles	69 (67)	62	73	58766.8 (18808.4)	49592.9 (22303.5)	0.435	0.22267	o (o)	
Brain	59 (58)	58	59	1191022.6 (110493.8)	1232890.5 (98415.8)	0.407	0.25346	o (o)	
WhiteMatter	35 (34)	30	39	505667.1 (82222.8)	506139.5 (76357.7)	0.006	0.98638	o (o)	
CerebellarVermalLobulesVIII-X	77 (78)	84	73	2659.5 (304.5)	3110.6 (394.1)	1.239	0.00121	+ (o)	
AccumbensArea	76 (78)	84	71	460.0 (150.5)	609.9 (125.5)	1.110	0.00327	+ (o)	
Amygdala	65 (67)	75	59	1795.5 (321.1)	2161.0 (378.9)	1.018	0.00642	+ (o)	
ThalamusProper	60 (61)	63	59	11584.6 (1556.0)	13072.5 (1826.3)	0.859	0.01955	+ (o)	
BrainStem	61 (62)	67	57	17417.2 (2275.5)	19416.8 (2417.5)	0.845	0.02144	+ (o)	
BasalForebrain	63 (66)	77	55	548.7 (232.3)	773.2 (290.7)	0.829	0.02383	+ (o)	
Hippocampus	64 (64)	62	66	5136.2 (1030.6)	6033.6 (1261.9)	0.759	0.03743	+ (o)	
CerebellarVermalLobulesI-V	63 (64)	68	59	3471.9 (535.9)	3949.6 (704.4)	0.737	0.04283	+ (o)	
VentralDC	61 (59)	54	64	8247.1 (1062.4)	8844.7 (880.6)	0.629	0.08149	o (o)	
CerebellumExterior	61 (62)	66	57	94910.8 (12670.3)	103457.0 (14443.9)	0.618	0.08647	o (o)	
Putamen	73 (74)	76	72	6066.7 (2492.2)	7087.0 (1586.9)	0.519	0.14721	o (o)	
Pallidum	54 (54)	54	54	2801.2 (796.4)	3063.0 (442.4)	0.439	0.21786	o (o)	
LateralVentricle	69 (68)	62	74	51162.3 (17123.4)	42726.9 (20449.5)	0.437	0.22046	o (o)	
CerebellumWhiteMatter	65 (66)	69	63	36782.2 (7917.1)	33401.6 (8837.8)	0.397	0.26456	o (o)	
InfLatVent	61 (61)	58	63	3131.4 (1360.4)	2649.1 (1182.9)	0.386	0.27817	o (o)	
Caudate	63 (59)	43	74	7388.5 (3739.3)	6400.6 (1557.5)	0.384	0.28063	o (o)	
CSF	52 (52)	51	52	2025.5 (719.9)	2194.0 (558.1)	0.271	0.44433	o (o)	
3rdVentricle	60 (58)	54	63	2288.6 (814.2)	2076.9 (941.1)	0.236	0.50451	o (o)	
4thVentricle	38 (36)	30	42	2184.5 (673.2)	2140.1 (576.0)	0.072	0.83738	o (o)	
CerebralWhiteMatter	37 (36)	35	38	468884.8 (76690.5)	472738.0 (70658.1)	0.053	0.88084	o (o)	
CerebellarVermalLobulesVI-VII	39 (39)	38	40	2076.6 (384.4)	2096.5 (498.0)	0.043	0.90263	o (o)	

Moderate disability (N = 22, Positives ^P) vs. Low disability (N = 32, Negatives ^N) (cross-sectional analysis, acute stage)									
structure	ACC (bACC)	SENS	SPEC	mean (SD) [mm ³] ^P	mean (SD) [mm ³] ^N	effect size (d)	p-value	sig. (corr.)	
RandomForest (all features incl. asymmetry)	61 (56)	32	80						
SVM (all features incl. asymmetry)	68 (64)	43	85						
MarshallScore	72 (68)	45	91	3.0 (1.9)	1.6 (1.1)	0.975	0.00091	++ (+)	
InjurySeverity	69 (66)	55	78	2.8 (0.8)	2.3 (0.7)	0.666	0.01981	+ (o)	
GlasgowComaScale	65 (60)	32	87	12.6 (3.9)	14.1 (2.2)	0.483	0.08720	o (o)	
Age	58 (59)	62	56	58.8 (10.0)	61.1 (10.1)	0.223	0.42386	o (o)	
Gender (female = 0, male = 1)	40 (38)	28	48	0.6 (0.5)	0.7 (0.5)	0.041	0.88325	o (o)	
Ventricles	61 (61)	58	64	49592.9 (22303.5)	38131.9 (13556.0)	0.650	0.02268	+ (o)	
CorticalGreyMatter	66 (67)	71	63	494829.1 (39983.4)	524318.2 (64687.5)	0.526	0.06297	o (o)	
WhiteMatter	46 (44)	35	54	506139.5 (76357.7)	492768.2 (71084.9)	0.183	0.51279	o (o)	
BrainTissue	49 (49)	49	48	1181045.4 (89378.0)	1198466.6 (135087.2)	0.147	0.59861	o (o)	
DeepGreyMatter	42 (41)	38	44	180076.7 (20887.9)	181380.2 (22209.8)	0.060	0.82903	o (o)	
Brain	40 (40)	39	40	1232890.5 (98415.8)	1238799.0 (138060.0)	0.048	0.86362	o (o)	
InfLatVent	71 (69)	60	78	2649.1 (1182.9)	1913.3 (676.5)	0.804	0.00543	+ (o)	
3rdVentricle	61 (61)	56	65	2076.9 (941.1)	1552.5 (617.3)	0.686	0.01660	+ (o)	
AccumbensArea	58 (59)	64	53	609.9 (125.5)	696.3 (138.7)	0.647	0.02332	+ (o)	
LateralVentricle	62 (61)	57	65	42726.9 (20449.5)	32704.8 (12459.3)	0.620	0.02953	+ (o)	
Putamen	57 (56)	51	61	7087.0 (1586.9)	7849.7 (1338.7)	0.528	0.06207	o (o)	
Hippocampus	54 (53)	49	58	6033.6 (1261.9)	6572.2 (873.5)	0.514	0.06913	o (o)	
CerebellarVermalLobulesVIII-X	65 (65)	66	65	3110.6 (394.1)	2902.0 (420.4)	0.509	0.07185	o (o)	
CerebellarVermalLobulesI-V	60 (61)	64	58	3949.6 (704.4)	4242.3 (699.1)	0.417	0.13786	o (o)	
ThalamusProper	60 (61)	68	54	13072.5 (1826.3)	13832.8 (1845.3)	0.414	0.14127	o (o)	
CerebellumWhiteMatter	56 (54)	45	64	33401.6 (8837.8)	30605.9 (8306.2)	0.328	0.24176	o (o)	
4thVentricle	62 (62)	59	65	2140.1 (576.0)	1961.2 (526.9)	0.327	0.24336	o (o)	
Amygdala	60 (60)	58	62	2161.0 (378.9)	2280.9 (380.7)	0.315	0.26007	o (o)	
BasalForebrain	52 (51)	46	55	773.2 (290.7)	840.4 (233.9)	0.260	0.35230	o (o)	
Caudate	51 (51)	50	51	6400.6 (1557.5)	6613.9 (1066.8)	0.166	0.55236	o (o)	
CerebralWhiteMatter	43 (41)	33	50	472738.0 (70658.1)	462162.3 (67838.9)	0.153	0.58231	o (o)	
Pallidum	49 (48)	41	54	3063.0 (442.4)	2990.8 (504.4)	0.151	0.58912	o (o)	
CerebellarVermalLobulesVI-VII	48 (47)	43	52	2096.5 (498.0)	2150.0 (299.1)	0.137	0.62397	o (o)	
CerebellumExterior	45 (44)	40	49	103457.0 (14443.9)	101795.8 (13850.8)	0.118	0.67217	o (o)	
VentralDC	44 (44)	47	41	8844.7 (880.6)	8961.7 (1114.9)	0.114	0.68245	o (o)	
CSF	45 (45)	42	47	2194.0 (558.1)	2140.6 (538.2)	0.098	0.72593	o (o)	
BrainStem	44 (44)	47	42	19416.8 (2417.5)	19651.1 (2739.4)	0.090	0.74754	o (o)	

F.2 Longitudinal analysis (baseline→follow-up)

Table F.2: Classification results in % (6-fold cross-validation, LDA/RF/SVM 100 runs) obtained separating TBI patients with a severe disability from patients with moderate disability outcome based on structural volume changes between the acute and chronic disease stage (top). Classification of patients with moderate disability and low disability outcome (bottom). The individual structures are sorted by effect size. Significant group differences indicated by + ($p < 0.05$) and ++ ($p < 0.001$). Bonferroni corrected significance in parentheses.

Severe disability (N = 13, Positives ^P) vs. Moderate disability (N = 22, Negatives ^N) (longitudinal analysis, bl→follow-up)									
structure	ACC (bACC)	SENS	SPEC	mean (SD) [%] ^P	mean (SD) [%] ^N	effect size (d)	p-value	sig. (corr.)	
RandomForest (all features)	59 (53)	29	76						
SVM (all features)	56 (50)	25	75						
Gender (female = 0, male = 1)	66 (66)	69	64	0.3 (0.5)	0.6 (0.5)	0.673	0.06287	o (o)	
Age	65 (63)	55	70	67.2 (15.9)	58.8 (10.0)	0.672	0.06331	o (o)	
DeepGreyMatter	62 (59)	48	70	-2.7 (2.2)	-1.2 (2.3)	0.639	0.07664	o (o)	
Brain	61 (62)	65	59	-2.0 (1.7)	-1.3 (2.3)	0.352	0.32098	o (o)	
BrainTissue	56 (54)	48	61	-3.0 (2.4)	-2.1 (3.2)	0.310	0.38205	o (o)	
CorticalGreyMatter	51 (51)	51	52	-3.0 (3.2)	-1.9 (4.3)	0.279	0.43074	o (o)	
WhiteMatter	49 (46)	37	56	-3.0 (2.8)	-2.5 (3.3)	0.154	0.66288	o (o)	
Ventricles	37 (36)	35	37	11.6 (12.0)	11.6 (17.8)	0.001	0.99774	o (o)	
CerebellumExterior	69 (66)	57	75	-3.1 (3.1)	-0.6 (2.4)	0.933	0.01173	+ (o)	
CerebellarVermalLobulesVI-VII	60 (60)	61	59	-2.1 (3.9)	0.4 (3.8)	0.668	0.06477	o (o)	
CerebellumWhiteMatter	67 (64)	52	75	-1.3 (1.3)	-0.6 (1.4)	0.535	0.13539	o (o)	
Caudate	50 (51)	53	48	0.1 (12.6)	-4.7 (12.3)	0.385	0.27928	o (o)	
BrainStem	65 (62)	52	73	-3.1 (3.3)	-2.0 (3.2)	0.332	0.34973	o (o)	
AccumbensArea	47 (55)	89	22	-0.2 (3.0)	-3.6 (13.1)	0.316	0.37305	o (o)	
Hippocampus	58 (53)	36	70	-4.8 (10.6)	-2.2 (6.7)	0.313	0.37795	o (o)	
VentralDC	68 (67)	62	72	-2.5 (1.8)	-1.4 (4.3)	0.294	0.40675	o (o)	
Pallidum	37 (46)	78	13	0.2 (0.8)	0.8 (3.2)	0.233	0.50989	o (o)	
BasalForebrain	50 (50)	47	52	1.2 (16.6)	5.2 (17.1)	0.233	0.51037	o (o)	
CerebellarVermalLobulesI-V	42 (42)	44	40	-0.9 (1.6)	-1.3 (2.1)	0.201	0.56983	o (o)	
CerebellarVermalLobulesVIII-X	43 (43)	46	41	-0.3 (1.9)	-0.9 (3.9)	0.187	0.59611	o (o)	
Amygdala	44 (40)	21	58	-2.6 (6.7)	-1.6 (5.1)	0.168	0.63403	o (o)	
CSF	54 (51)	40	62	6.2 (6.5)	4.7 (10.8)	0.160	0.65057	o (o)	
4thVentricle	45 (47)	54	41	0.3 (7.5)	2.3 (15.1)	0.153	0.66533	o (o)	
CerebralWhiteMatter	47 (45)	35	54	-3.1 (3.0)	-2.7 (3.6)	0.138	0.69533	o (o)	
3rdVentricle	45 (46)	52	41	5.2 (8.4)	6.8 (17.0)	0.112	0.75025	o (o)	
Putamen	37 (43)	64	22	-0.1 (0.8)	-0.3 (2.2)	0.071	0.84134	o (o)	
ThalamusProper	41 (39)	33	45	-4.2 (3.5)	-3.9 (6.2)	0.053	0.88026	o (o)	
LateralVentricle	37 (37)	34	40	12.8 (12.0)	12.5 (18.8)	0.018	0.95937	o (o)	
InfLatVent	38 (39)	41	37	7.7 (19.6)	7.8 (24.3)	0.006	0.98701	o (o)	

Moderate disability (N = 22, Positives ^P) vs. Low disability (N = 32, Negatives ^N) (longitudinal analysis, bl→follow-up)									
structure	ACC (bACC)	SENS	SPEC	mean (SD) [%] ^P	mean (SD) [%] ^N	effect size (d)	p-value	sig. (corr.)	
RandomForest (all features)	62 (59)	47	72						
SVM (all features)	65 (58)	21	94						
Age	59 (59)	62	56	58.8 (10.0)	61.1 (10.1)	0.223	0.42386	o (o)	
Gender (female = 0, male = 1)	39 (38)	30	46	0.6 (0.5)	0.7 (0.5)	0.041	0.88325	o (o)	
WhiteMatter	69 (66)	47	85	-2.5 (3.3)	-0.6 (1.2)	0.820	0.00460	+ (o)	
BrainTissue	70 (67)	51	84	-2.1 (3.2)	-0.6 (2.0)	0.570	0.04449	+ (o)	
Ventricles	62 (60)	44	76	11.6 (17.8)	4.2 (9.7)	0.545	0.05460	o (o)	
Brain	72 (69)	55	83	-1.3 (2.3)	-0.4 (1.6)	0.461	0.10190	o (o)	
CorticalGreyMatter	70 (68)	58	78	-1.9 (4.3)	-0.7 (2.9)	0.346	0.21716	o (o)	
DeepGreyMatter	59 (57)	48	66	-1.2 (2.3)	-0.5 (2.5)	0.300	0.28432	o (o)	
CerebralWhiteMatter	69 (66)	47	85	-2.7 (3.6)	-0.7 (1.2)	0.817	0.00474	+ (o)	
LateralVentricle	59 (56)	36	75	12.5 (18.8)	4.7 (10.7)	0.531	0.06072	o (o)	
CerebellarVermalLobulesI-V	61 (59)	48	69	-1.3 (2.1)	-0.5 (1.0)	0.527	0.06274	o (o)	
ThalamusProper	65 (62)	45	78	-3.9 (6.2)	-1.6 (3.1)	0.515	0.06844	o (o)	
BrainStem	66 (64)	54	74	-2.0 (3.2)	-0.8 (2.1)	0.470	0.09576	o (o)	
4thVentricle	60 (58)	50	67	2.3 (15.1)	-2.2 (4.5)	0.444	0.11507	o (o)	
CerebellarVermalLobulesVIII-X	64 (61)	48	74	-0.9 (3.9)	0.4 (2.3)	0.441	0.11719	o (o)	
Amygdala	58 (57)	53	61	-1.6 (5.1)	-0.0 (2.5)	0.424	0.13221	o (o)	
Caudate	59 (57)	43	70	-4.7 (12.3)	-1.1 (6.6)	0.389	0.16652	o (o)	
AccumbensArea	59 (53)	20	87	-3.6 (13.1)	-0.5 (1.7)	0.367	0.19060	o (o)	
3rdVentricle	60 (58)	47	68	6.8 (17.0)	2.4 (9.2)	0.345	0.21816	o (o)	
CerebellumWhiteMatter	59 (58)	50	66	-0.6 (1.4)	-0.1 (1.3)	0.341	0.22407	o (o)	
InfLatVent	58 (53)	31	76	7.8 (24.3)	3.0 (8.0)	0.291	0.29794	o (o)	
Pallidum	55 (49)	14	83	0.8 (3.2)	0.2 (1.1)	0.287	0.30512	o (o)	
Putamen	60 (57)	38	75	-0.3 (2.2)	0.1 (0.4)	0.271	0.33165	o (o)	
Hippocampus	56 (53)	36	70	-2.2 (6.7)	-1.1 (2.9)	0.237	0.39550	o (o)	
CSF	51 (48)	32	64	4.7 (10.8)	3.2 (6.0)	0.188	0.49978	o (o)	
BasalForebrain	51 (50)	47	53	5.2 (17.1)	2.7 (14.6)	0.159	0.56781	o (o)	
CerebellarVermalLobulesVI-VII	44 (43)	40	47	0.4 (3.8)	-0.1 (3.1)	0.151	0.58850	o (o)	
CerebellumExterior	47 (46)	39	52	-0.6 (2.4)	-0.3 (2.9)	0.103	0.71168	o (o)	
VentralDC	49 (47)	36	57	-1.4 (4.3)	-1.2 (3.0)	0.066	0.81394	o (o)	