ADVANCED IMAGE ANALYSIS FOR MODELING THE AGING BRAIN

W. HUIZINGA

Cover design by W. Huizinga.
Thesis layout was adopted from Tino Wagner.

The work in this thesis was conducted at the departments of Medical Informatics and Radiology & Nuclear Medicine of the Erasmus MC, Rotterdam, the Netherlands.

For financial support for the publication of this thesis, the following organizations are gratefully acknowledged: Alzheimer Nederland, Quantib BV, and the Erasmus MC.

ISBN: 978-94-6375-036-3
Printed by Ridderprint BV | `www.ridderprint.nl`

ADVANCED IMAGE ANALYSIS FOR MODELING THE AGING BRAIN

Geavanceerde beeldanalyse om het verouderende brein te modelleren

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit van Rotterdam
op gezag van de rector magnificus

Prof. dr. R.C.M.E. Engels

en volgens besluit van het College voor Promoties.

de openbare verdediging zal plaatsvinden op
07 November 2018 om 11:30 uur

door

Wyke Huizinga
geboren te Alkmaar

Erasmus University Rotterdam

PROMOTIECOMMISSIE:

Promotor: Prof. dr. W.J. Niessen

Overige leden: Prof. S. Durrleman
Prof. dr. A. van der Lugt
Prof. D. Rueckert

Copromotores: Dr. ir. S. Klein
Dr. ir. D.H.J. Poot

CONTENTS

1

GENERAL INTRODUCTION

As the brain ages naturally, it progressively loses structure due to the death of neurons and connections between them, a process called neurodegeneration. This causes the morphology of the brain to change. Very distinct changes are the increasing ventricular size and decreasing white and gray matter volumes. The effect of normal aging on brain morphology is illustrated in Figures 1.1a and 1.1b, which shows magnetic resonance (MR) images of a 46-year-old and a 92-year-old person, respectively.

When neurodegeneration occurs in an abnormal manner, we speak of neurodegenerative diseases, such as Parkinson's disease and Alzheimer's disease (AD). Figure 1.1c shows the image of an 85-year-old AD patient. Please note the morphological similarities to the 92-year-old brain in 1.1b. As neurodegeneration due to disease may be difficult to distinguish from that of normal aging, interpretation of brain images in the context of diagnosis of neurodegenerative diseases is challenging, especially in the early stages of the disease. This thesis presents comprehensive models of the aging brain and novel computer-aided diagnosis methods, based on advanced, quantitative analysis of brain MR images, facilitating the differentiation between normal and abnormal neurodegeneration.

The work described in this thesis makes extensive use of advanced image processing, machine learning, and pattern recognition techniques. In each chapter, pointers to the relevant literature are given and, where necessary, basic concepts of the used methodology are explained. In the section below, background information of some of the important techniques is discussed, in order to set the stage for a more precise definition of my research aims. The chapter ends with an outline of the thesis.

1.1 BACKGROUND

Neuro-image analysis is a broad field where many techniques and concepts play an important role. Specifically, in this thesis, image segmentation, image registration, diagnostic classification, and normative modeling are used to extract quantitative biomarkers, establish spatial correspondence between images, and develop models in order to support the development for clinical decision making. These key techniques and concepts are briefly discussed in the following sections.

1.1.1 *Image segmentation*

Segmenting the brain into its different tissue types and regions of interest is necessary when one wants to study their diagnostic relevance, heritability or structural connectivity. The man-

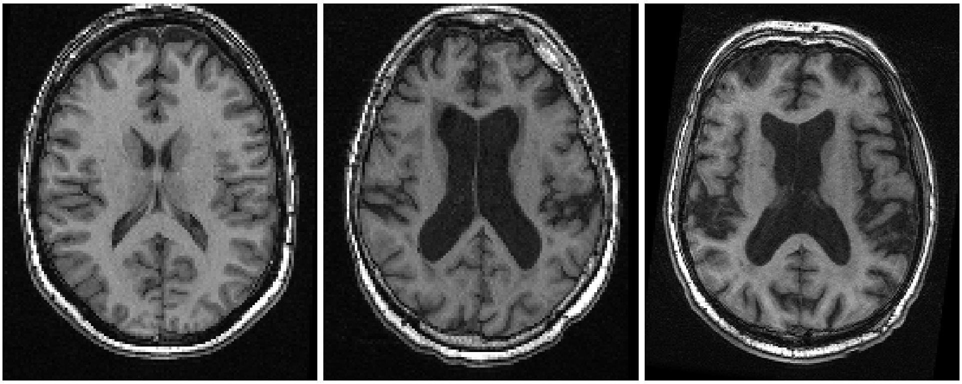

(a) A 46-year-old brain. (b) A 92-year-old brain. (c) An 85-year-old brain of an AD patient.

Figure 1.1: MR brain images showing how neurodegeneration affects brain morphology. One clear hallmark is the increasing ventricular size and the decreasing volumes of white and gray matter.

ual segmentation of a brain image is a time-consuming task, which has to be performed by an expert and is therefore too expensive and impractical for a routine clinical setting [1]. To automatically obtain brain region volumes from MRI brain data, numerous fully automated brain segmentation methods have been proposed in literature. Each method relies on different techniques to segment either the full brain or a specific region as accurately as possible, where manual segmentation serves as the golden standard. We can distinguish methods that are based on prior probability maps [2], statistical shape and appearance models [3–5], multi-atlas registration and labeling [6–12], deep-learning approaches [13–15], but also several other approaches [16–19]. Figure 1.2 shows a T1-weighted MR brain image with a colored overlay of several automatically segmented brain regions.

1.1.2 *Image registration*

Image registration is a technique that finds transformations to obtain spatial correspondence between images, such that image coordinates correspond to the same anatomical location in

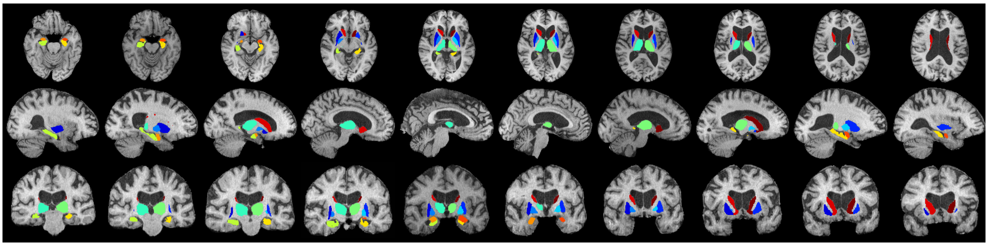


Figure 1.2: An MR brain image of a non-demented subject, with a colored overlay of the sub-cortical brain regions, as well as the hippocampus and amygdala. Slices in the axial direction are shown in the top row, slices in the sagittal direction are shown in the middle row, and slices in the coronal direction are shown in the bottom row. This illustration is based on a figure in Chapter 3.

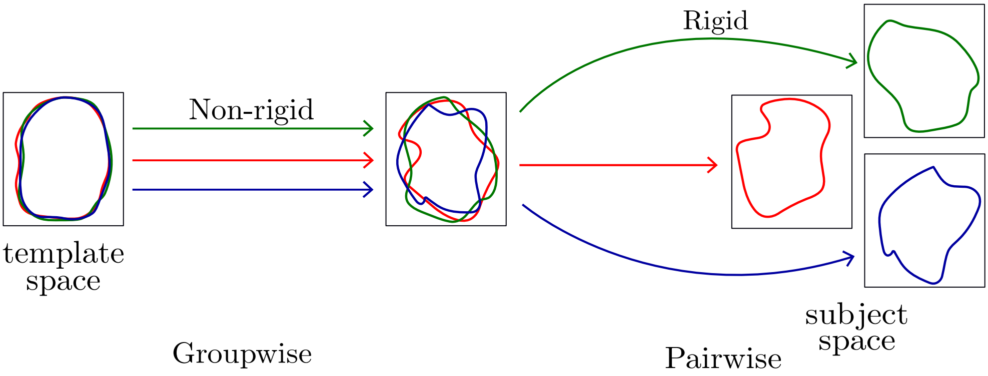


Figure 1.3: Illustration of the concepts of rigid, non-rigid, pairwise and groupwise registration. On the left three non-rigid transformations are simultaneously estimated, to transform the images into the template space. On the right, three rigid transformations are estimated via separate pairwise registrations, and are used as initialization of the non-rigid groupwise registration. This illustration is based on a figure in Chapter 4.

each of the images. In intra-subject registration, images of the same subject are registered. This is necessary to align images acquired with different imaging modalities, to compensate motion in dynamic imaging data, or to evaluate change in a longitudinal setting. In inter-subject image registration, images of different subjects are registered, which is for example used in atlas-based segmentation and template construction.

Two transformation types are distinguished: rigid transformations, which have limited degrees of freedom, e.g. translation, rotation, and possibly also scaling, and non-rigid or deformable transformations, with many degrees of freedom. Often, a rigid transformation is used as an initialization when estimating the non-rigid transformation. In the case of two images, a pairwise registration technique is often used. Here, one image is used as reference and the other image is spatially aligned with this reference. When more than two images are involved, a groupwise registration can be considered. Here, all images are simultaneously registered to an intrinsic average space, often called the template space. These transformation types and registration techniques are illustrated in Figure 1.3. For comprehensive surveys of the literature on image registration, the reader is referred to [20, 21].

1.1.3 *Diagnostic classification*

In neuro-image analysis, diagnostic classification of subjects using machine-learning approaches is an area of active research [23, 24]. Usually, the aim of diagnostic classification is to classify subjects into one or more classes: healthy or diseased, with sometimes multiple disease stages. This diagnostic classification is supported by models that are constructed using data from subjects for which the class is known. A diagnostic classification model uses features, which are biomarker values that distinguish between healthy and diseased subjects, and possibly the disease state of the subject. Such features are for example blood pressure,

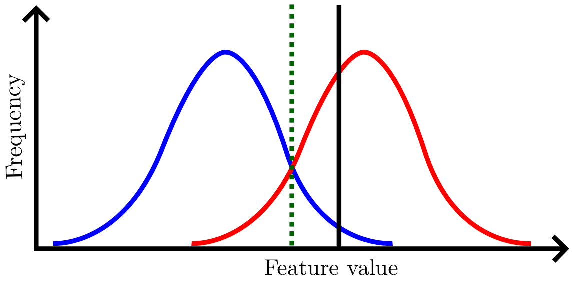


Figure 1.4: Illustration of univariate classification. The graph shows the feature value distributions of the subjects labeled as healthy (blue) and the subjects labeled as diseased (red). The green dotted line is the optimal decision boundary, i.e. the boundary where the model's classification performance is maximal. The black line is the feature value of a new subject, which would be classified as diseased according to this model. The larger the separation between the two distributions, the higher the feature's diagnostic relevance becomes. This principle is used in the Disease State Index (DSI) classifier [22], which is evaluated in Chapter 2.

cognitive test scores, but also imaging based features. Medical image analysis is the field where image processing techniques are used to extract features from imaging data, such as tissue or regional volumes, but also more advanced features such as white matter integrity, or brain deformation. When a diagnostic classification model uses a single feature it is called univariate. When multiple features are used, it is called multivariate. Figure 1.4 shows the concept of univariate classification.

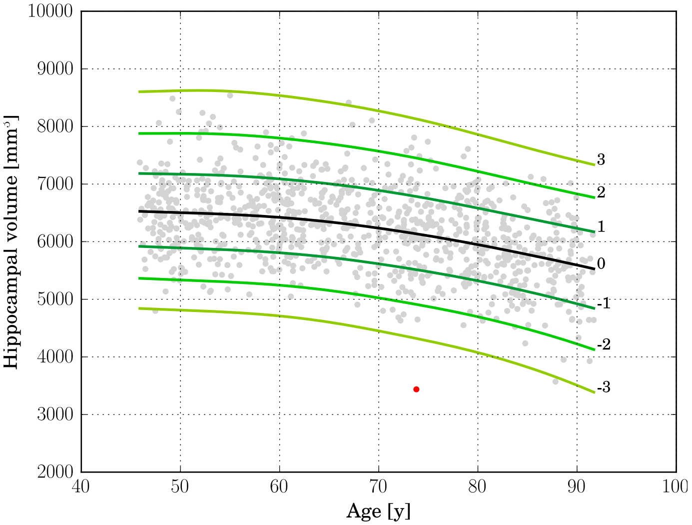


Figure 1.5: Example of a normative hippocampal volume distribution, visualized in iso-z-score lines from -3 to 3 SD. The light gray dots show the normative volumes and the red dot shows the volume of one patient with Alzheimer's disease. This illustration is based on a figure in Chapter 3.

1.1.4 *Normative modeling*

In order for a clinician to interpret qualitative information extracted from medical images, the clinician must know the range of these values for representative healthy reference persons. This range is determined using normative data, which can e.g. be acquired from population imaging studies and to which a patient's measurement can be compared [1]. The distribution of the normative data, the patient's measurement and its distance to the normative distribution can then be used to aid clinical decision making. With normative modeling, subjects with abnormal biomarker values are identified given a feature value distribution of a reference population. This approach could therefore be considered as an example of "one-class" diagnostic classification [25].

Normative data may incorporate covariates such as age or gender, when the distribution is expected to vary significantly as a function of these variables. To illustrate how normative volumetric MR data can be used in clinical practice, Figure 1.5 shows the normative distribution of hippocampus volumes as iso-z-score lines. The red dot shows the hippocampus volume of an AD patient that clearly lies outside the normative distribution.

1.2 RESEARCH AIMS

This thesis aims to develop and evaluate novel methods based on advanced, quantitative analysis of brain MR images, facilitating the differentiation between normal and abnormal neurodegeneration to support clinical decision making. Specifically, the following research objectives have been pursued:

1. To evaluate the accuracy of predicting global cognitive decline in the general population using a multivariate classification framework based on a wide variety of input features, including MRI, age, gender, cardiovascular risk factors, gait, cognitive, and genetic features.
2. To evaluate the impact of differences in automated brain region segmentation methods on single-subject analysis in a normative modeling framework.
3. To develop and evaluate a novel approach for extracting and modeling the brain morphology changes due to normal aging, leading to a spatio-temporal reference model of the aging brain.

Besides these three main research objectives, novel image registration methods have been developed that were crucial for completing the third objective, but also have many other applications in the field of medical image analysis:

4. A novel method for intra-subject non-rigid groupwise registration of multiple images with contrast differences.
5. A highly efficient algorithm for B-spline interpolation and transformation, which leads to substantial acceleration of non-rigid image registration methods.

1.3 OUTLINE

This thesis is divided into two parts. In the first part, comprising Chapters 2, 3 and 4, methods for clinical decision support using features derived from MR brain images are developed and evaluated. The second part presents methods that were developed to enable the work presented in Chapter 4, but have many applications in the field of medical image analysis.

In **Chapter 2** we validate the possibility to predict global cognitive decline in the general population using a previously proposed, multivariate classification framework, the Disease State Index (DSI) [22]. This prediction is relevant, because identifying persons at risk for global cognitive decline may aid in early detection of risk at dementia to support preventive strategies. We assess the prediction performance of the DSI with various sets of features. These features include MRI features, and non-imaging features such as age, sex, cognitive test results, cardiovascular risk factors, genetics, gait, and education.

In **Chapter 3** we assess differences between automated brain region segmentation methods in a normative modeling framework. Many automated methods have been proposed to extract region-based MRI features, several comparison studies have been done to evaluate their performance and to determine the difference between the methods. However, the impact of using different segmentation methods on the analyses of individual patients within a normative modeling framework was unknown. We therefore compare five automated brain segmentation methods, by measuring correlation and absolute agreement on non-demented subjects of six regional volumes. We also compare the absolute agreement on the position of AD patients relative to the normative distributions.

In **Chapter 4** we propose a method to build a reference model of the entire brain as a function of age, i.e. a spatio-temporal reference model, to which an individual brain morphology can be compared. This is achieved by computing voxel-wise features which are used to derive a description of the brain morphology. Brain deformation as a function of age is computed using groupwise image registration. Because this model was built on many images, a computationally efficient groupwise image registration method is applied. This was enabled by the novel techniques developed in Chapters 5 and 6.

In **Chapter 5** a groupwise image registration method is developed for the purpose of spatially aligning images acquired in a quantitative MRI acquisition. The anatomical correspondence between those images is crucial, because quantitative tissue parameters are subsequently determined by voxel-wise fitting the acquisition model on the images. Misalignment may lead to wrong estimation of these tissue parameters. Due to the large contrast differences between the acquired images, the registration is a challenging task. The method presented in this chapter aligns these images simultaneously regardless of the contrast differences. This method is also used to register the MRI brain images of the model presented in Chapter 4.

In **Chapter 6**, algorithms that are widely used in image registration, B-spline interpolation and transformation, are reformulated and efficiently implemented using an advanced C++ programming language feature called template metaprogramming. This feature simultaneously allows generic program code and runtime efficiency. The methods presented in Chapter 5 and Chapter 6 have been made publicly available in the image registration software package elastix.

Finally, **Chapter 7** discusses the presented work and provides recommendations for future research.

Part I

Clinical decision support using MR brain imaging

2

PREDICTING GLOBAL COGNITIVE DECLINE IN THE GENERAL POPULATION USING THE DISEASE STATE INDEX

Abstract. Identifying persons at risk for cognitive decline may aid in early detection of persons at risk of dementia and to select those that would benefit most from therapeutic or preventive measures for dementia. In this study we aimed to validate whether cognitive decline in the general population can be predicted with multi-variate data using a previously proposed supervised classification method: Disease State Index (DSI). We included 2,542 participants, non-demented and without mild cognitive impairment at baseline, from the population-based Rotterdam Study (mean age 60.9 ± 9.1 years). Participants with significant global cognitive decline were defined as the 5% of participants with the largest cognitive decline per year. We trained DSI to predict occurrence of significant global cognitive decline using a large variety of baseline features. Prediction performance was assessed as area under the receiver operating characteristic curve (AUC), using 500 repetitions of 2-fold cross-validation experiments. A mean AUC (95% confidence interval) for DSI prediction was 0.78 (0.77 - 0.79) using only age as input feature. When using all available features, a mean AUC of 0.77 (0.75 - 0.78) was obtained. Without age, and with age-corrected features and feature selection on MRI features, a mean AUC of 0.70 (0.63 - 0.76) was obtained, showing the potential of other features besides age. The best performance in the prediction of global cognitive decline in the general population by DSI was obtained using only age as input feature. Other features showed potential, but did not improve prediction. Future studies should evaluate whether the performance could be improved by new features, e.g., longitudinal features, and other prediction methods.

This chapter contains the content of *Predicting global cognitive decline in the general population using the Disease State Index*, L.G.M. Cremers & W. Huizinga et al., *submitted*.

2.1 INTRODUCTION

It is well established that neuropathological brain changes related to dementia accumulate over decades, and that the disease has a long preclinical phase. This may facilitate early disease detection and prediction [26]. A large amount of body of literature on potential features and risk factors for dementia exists. However, clinicians often struggle to integrate all the data obtained from a single patient for diagnostic or prognostic purposes. Therefore, there is a need for information technologies and computer-based methods that support clinical decision making [27].

Disease State Index (DSI) is a supervised machine learning method intended to aid clinical decision making [28]. This method compares a variety of patient variables with those variables from previously diagnosed cases, and computes an index that measures the similarity of the patient to the diagnostic group studied. The DSI method has previously been tested in specific patient populations and has shown to perform reasonably well in the early prediction of progression from mild cognitive impairment (MCI) to Alzheimer's disease and has been successful in the classification of different dementia subtypes [28–31]. In a recent study DSI has been validated in a population-based setting to predict late-life dementia [32]. Identification of persons at risk for global cognitive decline may aid in early detection of persons at risk of dementia and may help to develop therapeutic or preventive measures to postpone or even prevent further cognitive decline and dementia [33]. This is especially important since previous research has shown that preventive interventions for dementia were more effective in persons at risk than in unselected populations [34, 35]. We therefore used DSI to predict global cognitive decline in the general population to select the persons at risk.

The main aim of this study was to investigate whether multi-variate data can predict global cognitive decline in the general population. If a high risk group can be selected from the general population, a population screening program for this group might facilitate early detection of dementia. We evaluated the prediction performance using several sets of clinical features and features derived brain images acquired with magnetic resonance imaging (MRI), to assess whether the prediction is dependent on the combination of the input features. DSI was chosen as a classification method because this method is able to handle datasets with missing data, which is often the case in population study datasets. Also, this method has been successfully applied in previous studies and performed comparable to other state-of-the-art classifiers [29, 32, 36].

2.2 METHODS

2.2.1 *Study population*

We included participants from three independent cohorts within the Rotterdam Study (RS), a prospective population-based cohort study in a suburb of Rotterdam that investigates the determinants and occurrence of diseases in the middle-aged and elderly population [37]. Brain

MRI-scanning was implemented in the study protocol since 2005 [38]. The Rotterdam Study has been approved by the medical ethics committee according to the Population Study Act Rotterdam Study, executed by the Ministry of Health, Welfare and Sports of the Netherlands. Written informed consent was obtained from all participants [39].

We used data from RS cohorts I, II and III, of which each consists of multiple subcohorts. In this study a subcohort of RS cohort I, II and III were used, to which we refer as sI, sII and sIII, respectively. Baseline features of sI were collected during 2009-2011 and sII were collected during 2004-2006. The participants of the both cohorts were 55 years or older. For RS cohort III participants were 45 years or older at time of inclusion. Baseline features of sIII were collected during 2006-2008.

Participants with prevalent dementia, mild cognitive impairment (MCI) and MRI defined cortical infarcts at baseline were excluded for all analyses. In total, 4328 participants with baseline information on cognition, MRI and other features were included. Baseline MRI was acquired on average 0.3±0.45 years after collecting the non-imaging features. Furthermore, diffusion-MRI was acquired. However, for a subset of 680 participants in RS cohort II diffusion-MRI data was obtained on average 3.5±0.2 years later than the other baseline MRI features. Longitudinal data on global decline was available for 2542 out of 4328 participants. The follow-up cognitive assessment was on average 5.7±0.6 years after the baseline visit.

2.2.2 *Disease State Index*

Prediction was performed with DSI [28]. This classifier derives an index indicating the disease state of the participant under investigation based on the available features of that participant. DSI has two major advantages: 1) it can cope with missing data and 2) it gives an interpretable result because DSI also provides a decision tree that can be quite well explained.

DSI classifier is composed of the components: fitness and relevance [28]. Let N be the total number of negatives, P the total number of positives, $FN(x)$ the number of false negatives, and $FP(x)$ the number of false positives, when x is used as classification cut-off. Then the fitness function is estimated for each feature i as:

$$f_i(x) = \frac{FNR_i(x)}{FNR_i(x) + FPR_i(x)} = \frac{FN_i(x)}{FN_i(x) + \frac{P}{N} FP_i(x)} \tag{2.1}$$

where $FNR(x) = FN(x)/P$ is the false negative rate and $FPR(x) = FP(x)/N$ is the false positive rate in the training data when the feature value x is used as the classification cut-off. The fitness automatically accounts for the imbalance in class size making implicitly both classes equal in size, as the fraction P/N in the denominator scales the negative class (related to $FP(x)$) to correspond the size of the positive class. The fitness function is a classifier where the values < 0.5 imply negative class and > 0.5 positive class. The relevance of each feature is estimated by:

$$R = \max\{\text{sensitivity} + \text{specificity} - 1, 0\}, \tag{2.2}$$

which measures how good the feature is in differentiating the two classes. The lower the overlap between the distributions of positives and negatives, the higher R. Finally, DSI is computed from the equation:

$$\mathrm{DSI} = \frac{\sum_i R_i f_i}{\sum_i R_i} \tag{2.3}$$

DSI is a value between zero and one; somebody is classified as positive if $\mathrm{DSI} > 0.5$ and as negative if $\mathrm{DSI} < 0.5$. DSI is an ensemble classifier, meaning that it combines multiple independent classifiers (fitness functions) defined for each feature separately. Because of that, DSI can tolerate missing data. Features can be grouped in a hierarchical manner. The final DSI is a combination of the levels in the hierarchy. The fitness, relevance and their combination as a composite DSI are repeated recursively by grouping the data until a single DSI value is obtained [36]. Therefore, the final DSI, which is used for the classification, depends on the hierarchy structure, as a different structure leads to a different averaging of the feature combinations. The top-level part of the hierarchy defined for this study is shown in Figure 2.1.

2.2.3 *Baseline features*

Figure 2.1 shows the used categories of features in hierarchical manner. Please note that not all individual features are shown in this figure. The sections below describe all the used features (indicated in bold font) in detail.

2.2.3.1 *MRI features*

Multi-sequence MR imaging was performed on a 1.5 Tesla MRI scanner (GE Signa Excite). The imaging protocol and sequence details were described extensively elsewhere [38]. Morphological imaging was performed with T1-weighted, proton density-weighted and fluid-attenuated inversion recovery (FLAIR) sequences. These sequences were used for an automated tissue segmentation approach to segment scans into grey matter, white matter, cerebrospinal fluid (CSF) and background tissue [40]. **Intracranial volume (ICV)** (excluding the cerebellum and surrounding CSF cerebellar) was estimated by summing total grey and white matter and CSF. Brain tissue segmentation was complemented with a **white matter lesion** segmentation based on the tissue segmentation and the FLAIR image with extraction of white matter lesion voxels by intensity thresholding [41]. We obtained **(sub)cortical structure volumes**, **cortical thickness**, and **curvature** of the cortex and hippocampal volume using the publicly available FreeSurfer 5.1 software [2, 42, 43]. For **cerebral blood flow** measurements, we performed a 2D phase-contrast imaging as previously described [44]. In short, blood flow velocity (mm/sec) was calculated based on regions of interest (ROI) drawn on the phase-contrast images in the carotid arteries and basilar artery at a level just under the skull base. The value of mean signal intensity in each ROI reflected the flow velocity with the cross-sectional area of the vessel. Flow was calculated by multiplying the average velocity with the cross-sectional area of the vessel [44]. A 3D T2*-weighted gradient-recalled echo was used to image **cerebral micro-**

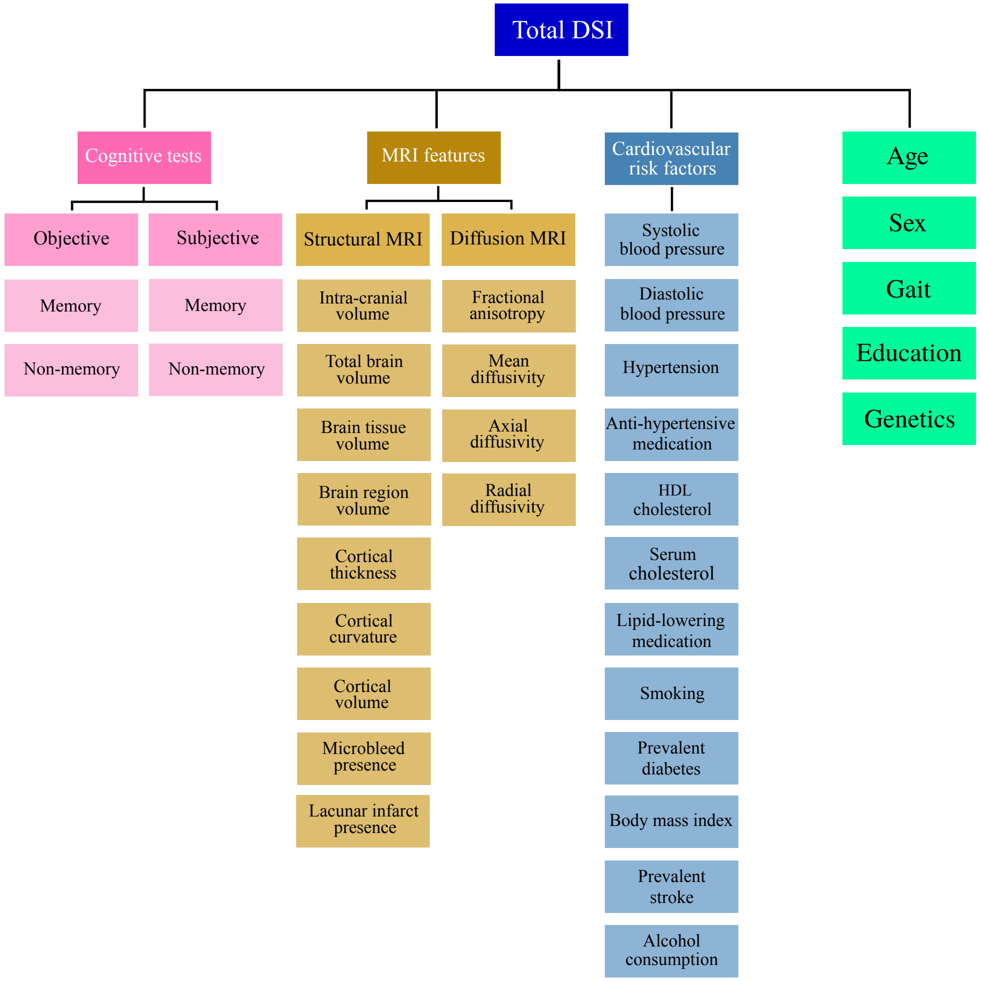


Figure 2.1: Feature categories shown in a hierarchy as used by DSI. Please note that not all individual features are included in this graph.

bleeds. Microbleeds were defined as focal areas of very low signal intensity, smaller than 10 mm in size and were rated by one of five trained raters who were blinded to other MRI sequences and to clinical data [45, 46]. **Lacunar infarcts** were defined as focal parenchymal lesions $\geq$ 3 mm and $<$ 15 mm in size with the same signal characteristics as cerebrospinal fluid on all sequences and with a hyperintense rim on the FLAIR image (supratentorially). Probabilistic tractography was used to segment 15 different white matter tracts in diffusion-weighted MR brain images, and we obtained **mean FA**, **mean MD**, **axial** and **radial diffusivity** inside each white matter tract [47].

2.2.3.2 *Cardiovascular risk factors*

Cardiovascular risk factors were based on information derived from home interviews and physical examinations during the center visit. **Blood pressure** was measured twice at the right brachial artery in sitting position using a random-zero sphygmomanometer. We used the mean of two measurements in the analyses. Information on the use of **antihypertensive medication** was obtained by using questionnaires and by checking the medication cabinets of the participants. Hypertension was defined as a systolic blood pressure $\geq$ 140 mmHg or a diastolic blood pressure $\geq$ 90 mmHg or the use of anti-hypertensive medication at baseline. **Serum total cholesterol** and **high-density lipoprotein (HDL) cholesterol** were measured in fasting serum, taking lipid-lowering medication into account. **Smoking** was assessed by interview and coded as never, former and current. **Body-mass index (BMI)** is defined as weight kilograms divided by height in meters squared. **Diabetes mellitus** status was defined as a fasting serum glucose level ($\geq$ 7.0 mmol/l) or, if unavailable, non-fasting serum glucose level ($\geq$11.1 mmol/l) or the use of anti-diabetic medication [39]. **Alcohol consumption** was acquired in a questionnaire. **Prevalent stroke** was ascertained as previously described [48]. **Educational level** was assessed during a home interview and was categorized into 7 categories, ranging from primary education only to university level [39].

2.2.3.3 *APOE-ϵ4 allele carriership*

APOE-ϵ4 allele carriership was assessed on coded genomic DNA samples [49]. APOE-genotype was in Hardy-Weinberg equilibrium. APOE-ϵ4 allele carriership was coded positive in case of one or two APOE-ϵ4 alleles.

2.2.3.4 *Gait features*

Gait was assessed by three walking tasks over a walkway: "normal walk", "turn" and "tandem walk" (heel to toe) [50]. Using a principal component analysis we obtained the following gait factors which we used by DSI: **Rhythm**, **Variability**, **Phases**, **Pace**, **Base of Support**, **Tandem**, and **Turning** [51].

2.2.3.5 *Baseline cognitive function*

We included the following **objective memory** and **non-memory** cognitive tests: **15-Word Learning Test immediate** and **delayed recall** [52], **Stroop tests** (**reading**, **color-naming** and **interference**) [53, 54], **The Letter-Digit Substitution Task** [55], **Word Fluency Test** [56], and the **Purdue Pegboard test** [57].

Subjective cognitive complaints were evaluated by interview. This interview included three questions on memory (**difficulty remembering**, **forgetting what one had planned to do**, and **difficulty finding words**), and three questions on everyday functioning (**difficulty managing finances**, **problems using a telephone**, and **difficulty getting dressed**) [58].

2.2.3.6 *Outcome: definition of cognitive decline*

A g-factor was constructed by a principal component analysis on the delayed recall score of the 15-Word Learning Test, Stroop interference Test, Letter-Digit Substitution Task, Word Fluency Test, and the Purdue Pegboard test [58]. Cognitive decline was defined by the g-factor from the follow-up visit minus the g-factor from the baseline visit resulting in a delta g-factor. Since the follow-up time was not the same for each participant, the delta g-factor was divided by the follow-up time to obtain global cognitive decline per year. Significant global cognitive decline (yes/no) was defined as belonging to the 5% of participants with the highest cognitive decline (delta g-factor) per year. In the used dataset, consisting of 2,542 participants, this resulted in 127 participants with a positive class label.

2.3 EVALUATION EXPERIMENTS

2.3.1 *Prediction performance evaluation*

The performance of DSI in predicting occurrence of global cognitive decline was evaluated using cross-validation. The area under the receiver-operator curve (AUC) was determined using 500 repetitions of 2-fold cross-validation (CV) experiments. This means that with each repetition 50% of the study dataset was used for training and the other 50% was used for testing, and vice versa, keeping the class ratio in the training and test set equal. We report the mean AUC, and the uncertainty of the mean expressed by its 95% confidence interval, derived from the 1000 resulting AUC values. The confidence interval was determined with the corrected resampled t-test for CV estimators of the generalization error [59]. AUCs were considered significantly different if the 95% confidence interval of their difference did not contain zero.

Since global cognitive decline per year is age dependent, we expect that age is an important feature for the prediction. We therefore include **age** as feature in the model. However, since other features might depend on age, correcting these features might improve the prediction performance [60]. We therefore also assessed the prediction performance using age-corrected features. We corrected the non-binary features for age using a linear regression model [61]. We evaluated four cases:

i age was included and no age-correction was performed on the non-binary features
ii age was excluded and no age-correction was performed on the non-binary features
iii age was included and non-binary features, except age, were corrected for age
iv age was excluded and non-binary features, except age, were corrected for age

To assess whether the performance of DSI was dependent on the combination of input features, we evaluated various feature combinations. In each cross-validation experiment the feature set was expanded with a feature or category of features. We analyzed four of such cumulative feature sets, differing in the order in which the feature set was expanded. Additionally, we analyzed MRI features separately and a set including all features but age.

2.3.2 *Relevance analysis*

To gain insight into the relevance weight that DSI assigns to each feature, we calculated the feature relevance distribution over the 500 repetitions of 2-fold CV, for the top-level feature categories of the hierarchy: age, sex, cognitive tests, cardiovascular risk factors, gait, education, genetics, and MRI features.

2.3.3 *Feature selection on MRI features*

In this study, hundreds of MRI features were extracted from images. It is likely that many of those features are not very efficient in detecting cognitive decline. Typically feature selection is applied to exclude poor features which may induce noise to the classifier. In DSI, weighting with relevance suppresses the effect of such features. If the number of features is high, their cumulative effect may, however, be remarkable. Previous results have shown that when including many features with a low relevance, the performance of DSI may decrease [32]. We therefore included an experiment evaluating the effect of feature selection on MRI features using their relevance. Due to averaging, feature noise reduces in higher levels of the feature hierarchy. The relevance of top-level feature categories may therefore be higher than lower-level, individual features. Therefore, due to the selection on the individual features, the top-level features may drop out, despite their high relevance. To prevent entire top-level feature categories to drop out of the model, we chose to only apply feature selection on the MRI features, which made up 80% of all input features, before selection. The relevance of the MRI features was determined on the entire dataset, before training. MRI features were selected by thresholding the relevance. Subsequently, an AUC distribution was determined in 10 repetitions of 2-fold CV. The following relevance thresholds were chosen: $t \in \{0.0, 0.01, \ldots, 0.09, 0.1\}$. For each threshold we assessed three feature sets in which the relevance-based feature selection on the MRI features was applied: 1) all features, 2) all features but age, and 3) MRI features only.

2.3.4 *Sub-group analyses*

As subjects close to the decision boundary (DSI $\sim$0.5) are more likely to be misclassified, we evaluated classification performance when only accepting/providing the classification for test subjects with low ($<$0.2) or high ($>$0.8) DSI. In this way, the subjects with $0.2 < \text{DSI} < 0.8$ are disregarded, which, in a clinical case, would mean that there is no diagnosis possible for these cases. We computed the AUC of this sub-group for DSI using all available features, both with age-correction and without age-correction.

Furthermore, we performed a sensitivity analysis in which the diffusion-MRI of 680 participants in RS cohort II were ignored, because this data was obtained on average 3.47± 0.15 years later than the other baseline MRI features.

Table 2.1: Baseline features of the study population and their relevances. The relevances were computed on the entire dataset. Continuous variables are presented as mean (standard deviation) and categorical variables as *n* (percentages).
Abbreviations: N; number of participants, HDL; high-density lipoprotein, s; seconds, FA; fractional anisotropy, MD; mean diffusivity $\times 10^{-3}$ mm^2/s. CSF; cerebrospinal fluid.
Symbols: R_{nac}; relevance when feature was not corrected for age, R_{ac}; relevance when feature was age-corrected.

Feature	R_{nac}	R_{ac}	**Positive (N=127)**	**Control (N=2415)**
Age, years	0.38	-	71.2 (10.1)	60.3 (8.7)
Sex, female	0.01	-	73 (54.5%)	1340 (55.6%)
Objective cognitive test results	0.28	0.16	-	-
Word Learning Test immediate recall	0.09	0.02	7.7 (2.2)	8.1 (2.0)
Word Learning Test delayed recall	0.05	0.04	7.9 (2.9)	8.2 (2.8)
Reading subtask of Stroop test, s	0.20	0.03	17.2 (2.7)	16.3 (2.9)
Color naming subtask of Stroop test, s	0.18	0.06	23.6 (3.6)	22.3 (4.0)
Interference subtask of Stroop test, s	0.32	0.10	53.8 (20.3)	44.0 (13.0)
Letter-Digit Substitution Task, number of correct digits	0.15	0.00	29.7 (6.7)	32.2 (6.2)
Word Fluency Test, number of animals	0.04	0.08	23.2 (5.7)	23.8 (5.7)
Purdue Pegboard test, number of pins placed	0.15	0.07	10.3 (2.1)	10.9 (1.7)
Mini-mental-state examination	0.14	0.11	27.8 (1.7)	28.4 (1.5)
Education[1]	0.07	0.07	3 (1-3)	3 (2-5)
Cardiovascular risk factors	0.34	0.27	-	-
Alcohol[1], glasses per week	0.06	0.04	3.5 (0.3-5.5)	5.5 (1.0-5.5)
Systolic blood pressure, mmHg	0.24	0.04	146.2 (20.3)	135.9 (19.6)
Diastolic blood pressure, mmHg	0.00	0.02	82.8 (9.4)	82.4 (10.6)
Blood pressure lowering medication	0.26	-	51 (38.3%)	284 (11.9%)
Body Mass Index, kg/m^2	0.07	0.07	28.2 (4.4)	27.4 (4.1)
Serum cholesterol, mmol/L	0.11	0.12	5.4 (0.9)	5.6 (1.1)
HDL-cholesterol, mmol/L	0.04	0.09	1.4 (0.4)	1.5 (0.4)
Lipid lowering medication	0.13	-	46 (34.6%)	510 (21.3%)
Smoking	0.08	0.08	-	-
Never	-	-	49 (36.6%)	746 (31.2%)
Former	-	-	54 (40.3%)	1154 (48.2%)
Current	-	-	31 (23.1%)	492 (20.6%)
Diabetes mellitus, presence	0.09	-	24 (18.2%)	220 (9.2%)
APOE-ϵ4 allele carriership	0.02	-	39 (30.2%)	639 (28.3%)
MRI features	0.41	0.25	-	-
Intra-cranial volume, mL	0.03	0.00	1137 (119)	1144 (113)
Brain tissue volume	0.38	0.08	-	-
White matter volume, mL	0.13	0.01	390 (60)	419 (57)
Gray matter volume, mL	0.10	0.01	522 (54)	537 (52)
CSF volume, mL	0.29	0.07	223 (53)	186 (46)
Brain region volume	0.35	0.12	-	-
Hippocampus volume, mL	0.23	0.09	6.4 (0.8)	6.8 (0.7)
White matter lesion volume[1], mL	0.31	0.08	4.5 (2.5-9.4)	2.4 (1.4-4.3)
Cerebral microbleeds, presence	0.09	-	33 (24.6%)	370 (15.6%)
Lacunar infarcts, presence	0.04	-	10 (7.5%)	72 (3.0%)
Global FA	0.17	0.07	0.3 (0.02)	0.3 (0.01)
Global MD,10^{-3} mm^2/s	0.33	0.07	0.8 (0.03)	0.7 (0.03)
Global cortical thickness, mm	0.08	0.01	2.4 (0.2)	2.5 (0.1)
Gait	0.19	0.17	-	-

[1] Education, alcohol and white matter lesion volume are presented as median (inter-quartile range).

2.4 RESULTS

Table 2.1 presents the characteristics of the study population. The mean age of the participants was 60.9 ± 9.1 years and 55.6% were females.

2.4.1 *Prediction performance*

Figure 2.2a shows the mean AUC (95% confidence interval) for several combinations of features in predicting global cognitive decline, without correcting the non-binary features for age. Each color represents an expanding set of used input features, where the most left set is only MRI features and the most right set is all features except age. When using only MRI features, the AUC was 0.75 (0.70 - 0.80). When using only age as baseline feature, the AUC was 0.78 (0.74 - 0.83). Using additional features on top of age resulted in an equal or slightly lower AUC (differences not statistically significant). When using all available features with DSI, the AUC was 0.77 (0.72 - 0.82). The mean AUC of DSI without age as baseline predictor was 0.75 (0.70 - 0.80).

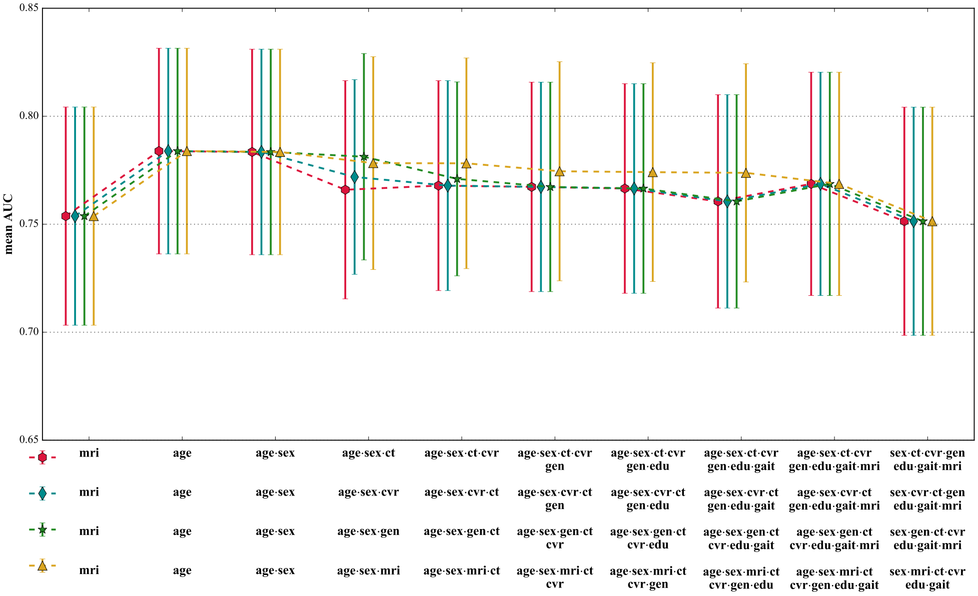
Figure 2.2b shows the mean AUC (95% confidence interval) for the same combinations of features as in Figure 2.2a, but here the non-binary features were corrected for age. The AUC for MRI features only was significantly lower with age-correction compared to without age correction, with an AUC of 0.55 (0.50 - 0.61). For the other feature sets, the AUC of the models where age correction was applied was not statistically significantly different, compared to not using age correction. When the effect of age was totally removed from the model, i.e. model iv, the AUC was 0.65 (0.58 - 0.73).

2.4.2 *Relevance analysis*

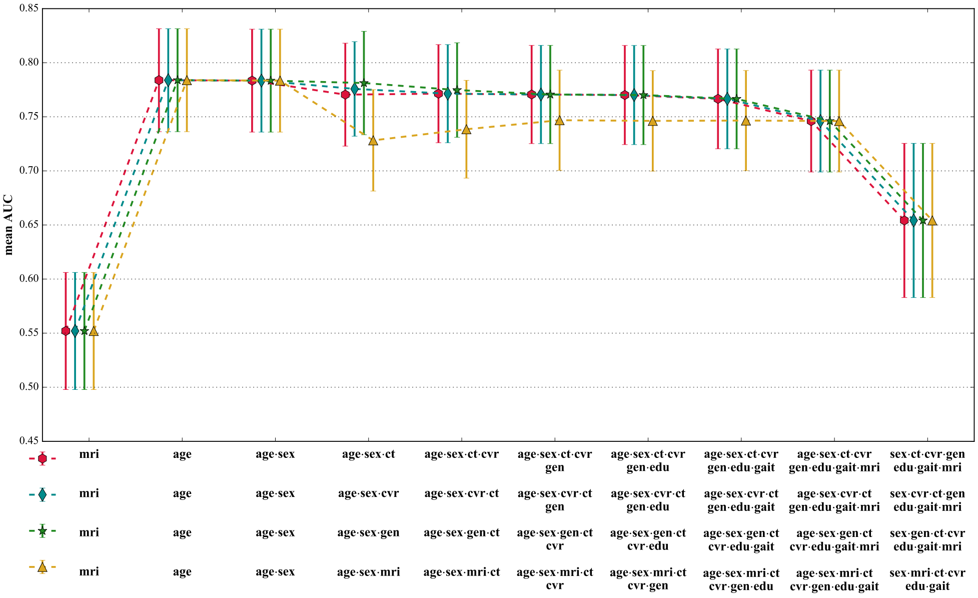
Figure 2.3 shows the relevance weight per feature category when the non-categorical features were corrected for age prior to computing DSI and without age-correction. Without age-correction, the features with the best discriminating abilities according to their relevance weights were MRI features (0.42 (0.33 - 0.51)), age (0.39 (0.27 - 0.51)), cognitive tests (0.35 (0.24 - 0.45)) and cardiovascular risk factors (0.34 (0.26 - 0.43)). When correcting the non-binary features, except age, for age, the most discriminating features were age (0.39 (0.27 - 0.51)), MRI features (0.37 (0.24 - 0.51)), and cognitive tests (0.32 (0.17 - 0.47)).

2.4.3 *Feature selection on MRI features*

Feature selection for MRI features had no effect on the AUC in any of the three feature sets, when the non-binary features were not corrected for age (see Figure 2.4a). The AUC did increase after MRI feature selection when the non-binary features, except age, had been corrected for age, with the optimal t being 0.07 (see Figure 2.4b). For $t = 0.07$, the AUC increased from 0.55 (0.50 - 0.61) to 0.62 (0.58 - 0.67) when only MRI features were included in the model. When using all features, the AUC increased from 0.75 (0.70 - 0.79) to 0.77 (0.73 - 0.82), and when using all features but age, the AUC increased from 0.65 (0.58 - 0.73) to 0.70 (0.63 - 0.76).



(a) No age correction was applied. Please note the y-axis (mean AUC) ranges from 0.65 - 0.85.



(b) Age correction was applied to the non-binary features. Please note the y-axis (mean AUC) ranges from 0.45 - 0.85.

Figure 2.2: Mean AUC for several combinations of features. Features are accumulated in four different orders, indicated by color and symbol. The bars indicate the confidence interval. Short-hand notations are used for several features: cognitive tests (ct), cardiovascular risk factors (cvr), MRI features (mri), genetics (APOE-$\epsilon4$ carrier-ship) (gen), and educational level (edu).

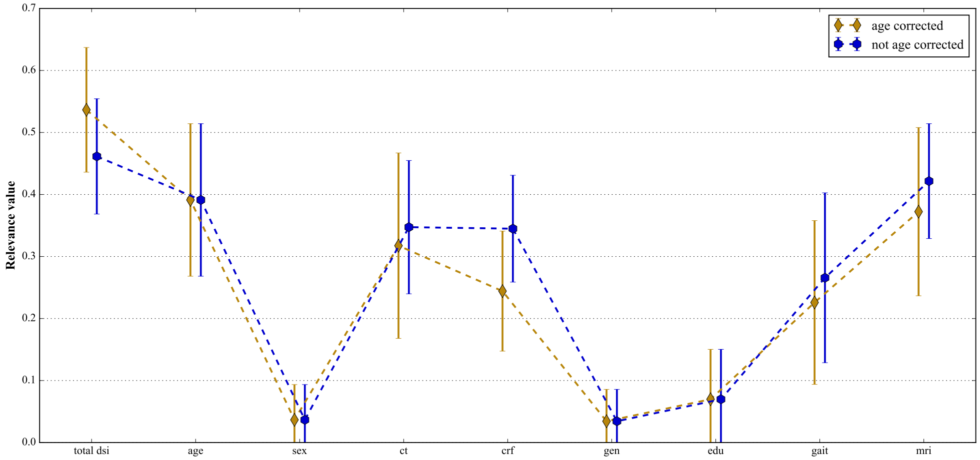


Figure 2.3: Mean relevance weight R and 95% confidence interval for the top-level features categories. The blue line shows the case where the non-binary features were non corrected for age and the golden line shows the case where the non-binary features were age-corrected.

2.4.4 *Sub-group analyses*

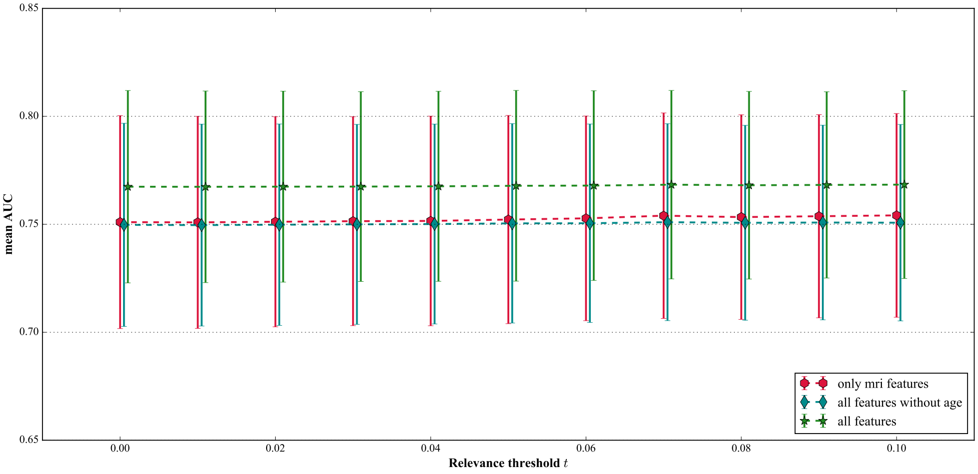
When only taking into account the extreme cases, i.e. cases for which $0.2 < \text{DSI} < 0.8$ (∼40% of the total dataset, i.e. ∼1000 subjects), the mean AUC increased to 0.82 (0.76 - 0.88) using age as input feature only. Again in this group, additional features did not significantly improve the performance of DSI (results not shown).

Ignoring the diffusion-MRI features of 680 participants of whom this data was acquired on average 3.47 ± 0.15 years later than the assessment of the other baseline MRI features did not change AUC significantly compared to the performance in the total population (results not shown).

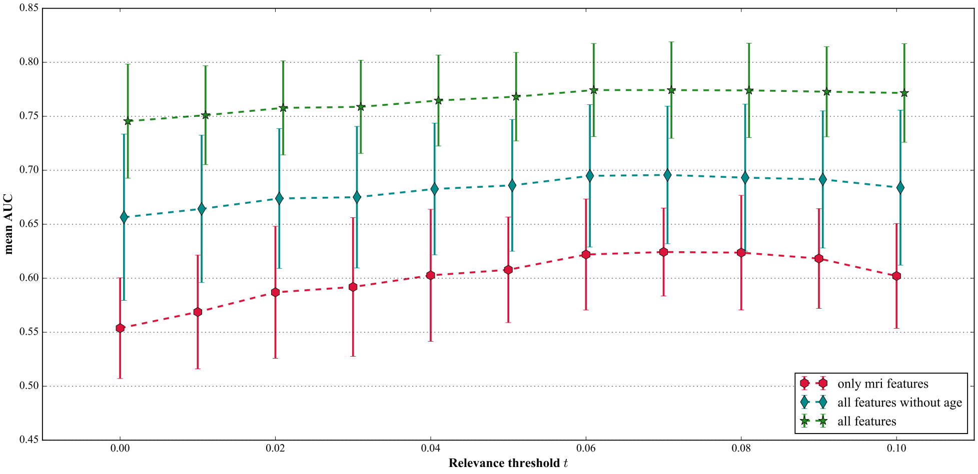
2.5 DISCUSSION

The objective of this study was to assess whether global cognitive decline can be predicted using multi-variate data with the previously proposed DSI. We found the best prediction performance, evaluated with AUC, using only age as input feature. Adding more features to DSI did not improve its performance in predicting global cognitive decline as defined in this study.

Overall performance of DSI in the prediction of global cognitive decline (mean AUC 0.78) was comparable to previously reported performances of DSI for prediction of dementia [32] in the population-based CAIDE study, consisting of 2000 participants who were randomly selected from four separate, population-based samples, originally studied in midlife (1972, 1977, 1982, or 1987) [62], and to other population-based prediction models of dementia [63]. In this study we included a large number of heterogeneous features. Age was the most



(a) No age correction was applied. Please note the y-axis (mean AUC) ranges from 0.65 - 0.85.



(b) Age correction was applied to the non-binary features. Please note the y-axis (mean AUC) ranges from 0.45 - 0.85.

Figure 2.4: Mean AUC for several combinations of features where the MRI features were selected based on their relevance. Features with $R < t$ were excluded.

important feature for predicting global cognitive decline using DSI, yielding the highest AUC. This was further supported by the observation that the performance of DSI reduced when using all features except age. Our finding that age is the single strongest predictor for cognitive decline is in line with published prediction models for dementia, that invariably assign the highest weight to age [64]. We found that the relevance R, which indicates how well a feature can discriminate between persons who will develop cognitive decline and those who will not, was highest for MRI features (0.42) followed by age (0.39). DSI, however, performed worse when using only MRI features, compared to using only age. We speculate that the high relevance of the MRI features may be explained by age-specific effects that are captured in these MRI features, which is supported by our finding that MRI feature relevance (0.37) and DSI performance dropped when adjusting MRI features for age. When the non-binary features were age-corrected and age was not included in the model, the mean AUC was 0.65, still significantly better than chance (0.5), indicating that relevant information for predicting global cognitive decline could be present in the other features. In this study, however, they did not improve the predicting performance when added to age.

To our surprise we found that APOE-ϵ4 allele carrier-ship had a low relevance weight and did not improve the performance of DSI, even though it is the best known genetic risk factor for AD. This is in contrast to a previous study focusing on the progression from MCI to AD, which found APOE-genotype to have high predictive value [65]. It may be that our study population was too young to show effect of APOE on prediction (mean age 60.9), since the risk progression effect of APOE-ϵ4 allele carriership has been described to peak between ages 70 and 75 years [66].

The relevance-based feature selection on the MRI features showed an increase in the AUC, but only when the non-binary features were corrected for age. A possible explanation is that without age correction, the AUC is strongly driven by the age-factor that is present in the MRI features. In this case, less and different features were excluded compared to the age-corrected models, causing the selection to have no effect on the prediction performance. However, after removal of these age-specific effects by age correction, performance can be increased by removal of irrelevant features. When age was totally excluded from the model iv (age was excluded and age correction was applied to the non-binary features), an AUC of 0.70 was obtained, showing the potential of the other features. One limitation of this analysis is that the relevance computation and threshold selection was done on the entire dataset, i.e. the training data was included in these computations. Therefore, AUC increase due to application of the relevance threshold might be overestimated, but can be seen as an upper limit. The overall conclusions do not change.

To our knowledge, this is the first population-based study testing the supervised machine learning DSI tool for prediction of global cognitive decline. Strengths of our study include the population-based design, large sample size and availability of an extensive set of features. However, limitations of our dataset need to be considered. We constructed a g-factor as a measure of global cognition and participants without complete cognitive data were excluded. This might have caused some selection bias towards relatively healthy subjects. Also, mortality and drop-out was not taken into account. Persons who are lost to follow-up usually have a poorer health status and are therefore more likely to develop cognitive decline or die before onset of cognitive decline. The exclusion of these assumingly more severe cases might have lowered the performance of DSI.

The result that age is the main predictor for cognitive decline indicates that the age distribution of the subjects with cognitive decline differs from the entire set of subjects. Hence age could be used to select people at risk of cognitive decline. However, when screening for significant cognitive decline, an age-dependent threshold on cognitive decline might be needed, e.g. using the 5% percentile of the cognitive decline as function of age, to detect young people at risk of developing dementia. The usage of such an age-dependent threshold will be part of future research.

Finally, it should be noted that cognitive decline is not equivalent to neurodegeneration/dementia and may result from other causes as well, due to conditions affecting the participant's cognition at the time of the cognitive assessment, normal human variability and normal aging. Nevertheless, being able to predict cognitive decline would be a step forward in selecting people for therapy or prevention.

2.6 CONCLUSION AND FUTURE WORK

Based on our results we can conclude that age is the most important predictor for cognitive decline in the general population using DSI. Other features showed having potential, but did not improve prediction performance. A next step could be to use longitudinal features in DSI, as these might improve its prediction performance. To validate whether our findings are not due to limitations of DSI, also other methods need to be evaluated in this prediction challenge. Finally, to be able to detect younger people at risk of significant global cognitive decline in future studies, thresholds for cognitive decline should be carefully chosen depending on the population, for example be age-adjusted.

2.7 FUNDING

The research leading to these results has received funding from the European Union Seventh Framework Program (FP7/2007-2013) under grant agreement no. 601055, VPH-Dare@IT (FP7-ICT-2011-9-601055).

3

DIFFERENCES BETWEEN MR BRAIN REGION SEGMENTATION METHODS: IMPACT ON SINGLE-SUBJECT ANALYSIS

Abstract. For the segmentation of magnetic resonance (MR) brain images into anatomical regions, numerous fully automated methods have been proposed and compared to reference segmentation obtained manually by clinical experts. However, there might be systematic differences between the resulting segmentations dependent on the employed segmentation method. This potentially results in differences in sensitivity to disease and can further complicate the comparison of individual patients to normative data. In this study, we aim to answer two research questions: 1) to what extent are methods interchangeable, as long as the same method is being used for computing normative volume distributions and patient-specific volumes? and 2) can different methods be used for computing normative volume distributions and assessing patient-specific volumes? To answer these questions, we compared the volumes of six brain regions calculated by five state-of-the-art brain region segmentation methods. We applied the methods on 988 non-demented (ND) subjects and computed the correlation (PCC-v) and absolute agreement (ICC-v) on the volumes. For most regions the PCC-v was good (> 0.75) indicating that volume differences between methods in ND subjects are mainly due to systematic differences. The ICC-v was generally lower, especially for the smaller regions, indicating that it is essential that the same method is used to generate normative and patient data. To evaluate the impact on single subject analysis we also applied the methods to 42 patients with Alzheimer's disease (AD). In the case where the normative distributions and the patient-specific volumes were calculated by the same method, the patient's distance to the normative distribution was assessed with the z-score. We determined the diagnostic value of this z-score, which showed to be consistent across the methods. We also determined the absolute agreement on the AD patient z-scores (ICC-z). We found that the ICC-z was high for the regions thalamus and putamen. Our results are encouraging as they indicate that methods are to some extent interchangeable for selected regions. For the regions hippocampus, amygdala, caudate nucleus and accumbens, and globus pallidus, not all method combinations showed a high ICC-z. Whether two methods are indeed interchangeable should be confirmed for the specific application and dataset of interest.

This chapter contains the content of *Differences between MR brain region segmentation methods: impact on single-subject analysis*, W. Huizinga et al., *submitted*.

3.1 INTRODUCTION

Quantitative imaging biomarkers are biological features that can be measured in medical images. They are of interest for diagnosis when changes in these features are due to disease. In the case of traumatic brain injury or neurodegenerative disease typical valuable quantitative imaging biomarkers are brain region volumes [8, 67, 68]. A well-known example is the volume of the hippocampus. A relatively low volume may indicate the presence of Alzheimer's disease (AD) [26, 69, 70]. To determine if a patient deviates significantly, one can compare it to so-called normative data [1, 71, 72]. Normative data is acquired in a reference population and it is used as a baseline distribution for a measurement, against which an individual measurement can be compared. Normative data may incorporate covariates such as age or gender, when the distribution is expected to vary significantly as a function of these variables. Well-known examples are head-circumference-for-age, height-for-age, weight-for-age, and weight-for-height norms, provided by the WHO [73] for detecting abnormal growth in children. The dependency on age is also the case for volumetric magnetic resonance (MR) brain images. Brewer *et al.* proposed using quantile curves as a function of age as normative data for volumetric MR measurements [1].

Volumetric MR measurements are acquired by segmenting the brain into its different tissue types and regions of interest. The manual segmentation of a brain image is a time-consuming task, which has to be performed by an expert and is therefore too expensive and impractical for a clinical setting [1]. To automatically obtain brain region volumes from MRI brain data, numerous fully automated brain segmentation methods have been proposed in the literature. Each method relies on different techniques to segment either the full brain or a specific region. We can subdivide the methods that are based on prior probability maps [2], statistical shape and appearance models [3–5], multi-atlas registration and labeling [6–12], deep-learning approaches [13–15], and other [16–19]. Each method aims to segment the brain as accurately as possible where manual segmentation serves as the gold standard.

Various comparison studies have been performed with regard to automated brain segmentation methods. Grimm *et al.* assessed the differences in amygdalar and hippocampal volume resulting from Freesurfer [2], VBM8 (VBM[1]) and manual segmentation. They concluded that volumes computed with VBM8 and Freesurfer V5.0 are comparable, and systematic and proportional differences were mainly due to different definitions of anatomic boundaries. They concluded that large differences can still exist even with high correlation coefficients [74]. Morey *et al.* also compared amygdalar and hippocampal volumes, but using methods FSL/FIRST[2], Freesurfer [2] and manual segmentation. They concluded that for the hippocampus Freesurfer was more similar to manual segmentation in terms of volume difference, overlap and correlation. For the amygdala, FIRST represented the shape more accurately than Freesurfer [75]. Babaola *et al.* compared four different state-of-the-art algorithms for automatic segmentation of sub-cortical structures in MR brain images and evaluated spatial

---

[1] `http://dbm.neuro.uni-jena.de/vbm/`

[2] `https://fsl.fmrib.ox.ac.uk/fsl/fslwiki`

overlap, distance and volumetric measures. They concluded that all four methods perform on par with recently published methods [76]. One of their evaluated methods, described in [77], performed significantly better than the other three methods according to their evaluation. Perlaki *et al.* compared the segmentation accuracy of the caudate nucleus and putamen between FSL/FIRST and Freesurfer by studying the Dice coefficient, absolute and relative volume difference. They also measured consistency and absolute agreement. They concluded that for caudate segmentation, FIRST and Freesurfer performed similarly, but for putaminal segmentation FIRST was superior to Freesurfer [78].

The impact, however, of using different methods on the analyses of individual patients within a normative modeling framework is still unknown. This is relevant when volumetric MR data is used to generate normative distributions for both research and clinical use. In this study, we therefore aim to answer two research questions: 1) to what extent are methods interchangeable, as long as the same method is being used for deriving normative volume distributions and patient-specific volumes? and 2) can different methods be used for deriving normative volume distributions and patient-specific volumes? To answer these questions, we evaluated five state-of-the-art segmentation methods [2, 5–8].

Different MR acquisition protocols may lead to different image contrasts, and since most automated methods are - partly or entirely - driven by the contrast in the image, this may influence the segmentation results. To rule out possible differences of the segmentation due to the acquisition protocol, the methods were applied to the same images, all acquired with the same acquisition protocol [79].

3.2 MATERIAL AND METHODS

3.2.1 *Data*

To derive the normative distributions as a function of age, we applied the brain region segmentation methods to 988 T1w MR brain images from non-demented (ND) (425 male, age=68.1±13.0 years) participants of the population-based Rotterdam Scan Study, a prospective longitudinal study among community dwelling subjects aged 45 years and over [79]. We adopted this dataset from [80]. All brain images were acquired on a single 1.5T MRI system (GE Healthcare, US). The T1w imaging protocol was a 3-dimensional fast radiofrequency spoiled gradient recalled acquisition with an inversion recovery pre-pulse sequence [79]. The images were reconstructed to a voxel size of $0.5 \times 0.5 \times 0.8$ mm^3 and the number of voxels in each dimension was $512 \times 512 \times 192$.

In addition, we used the brain images of 42 (25 male, age=81.9±4.9 years) patients with AD at the time of the MRI scan from the same imaging study.

Table 3.1: Brain regions segmented by each method.

Method	# regions	description
EMC	83	Sub-cortical regions, cortical regions, ventricles, corpus callosum, substantia nigra, lobes, brain stem, cerebellum
FS	261	Sub-cortical regions, cortical regions, ventricles, lobes, optic chiasm, ventral diencephalon, lesions, vessels, corpus callosum, choroid plexus, brain stem, cerebellum
GIF	144	Sub-cortical regions, cortical regions, ventricles, optic chiasm, ventral diencephalon, lesions, vessels, lobes, brain stem, cerebellum
MALP-EM	138	Sub-cortical regions, cortical regions, ventricles, lobes, brain stem, cerebellum
MBS	56	Sub-cortical regions, ventricles, corpus callosum, fornix, septum pellucidum, lobes, brain stem, pons, cerebellum

3.2.2 *Brain segmentation methods*

We applied five previously proposed brain segmentation methods to the imaging data. The following five segmentation methods, explained in detail below, were evaluated:

1. Multi-atlas registration combined with tissue segmentation for cortical regions, developed at Erasmus MC (EMC), the Netherlands.
2. Freesurfer (FS), developed at the Athinoula A. Martinos Center for Biomedical Imaging at Massachusetts General Hospital, United States of America.
3. Geodesic information flows (GIF), developed at University College London, United Kingdom.
4. Multi-Atlas Label Propagation with Expectation-Maximisation based refinement (MALP-EM), developed at Imperial College London, United Kingdom.
5. Model-based brain segmentation (MBS), developed at Philips Research Hamburg, Germany.

The regions segmented by each method are shown in Table 3.1. Below, a short description of each method is given.

3.2.2.1 *EMC*

This method combines multi-atlas registration and voxel-wise tissue segmentation for cortical regions, and hippocampus and amygdala. Probabilistic tissue segmentations are obtained on the image to be segmented using the unified tissue segmentation method [81] of SPM8 (Statistical Parametric Mapping, London, UK). Thirty labeled T1-weighted MR brain images are used as atlas images [82, 83]. The atlas images are registered to the subjects' image using a rigid, affine, and non-rigid transformation model consecutively, and a mutual information-based similarity measure. The subjects' images are corrected for inhomogeneities to improve registrations using the N3 algorithm [84]. Labels are fused using a majority voting algorithm [85]. For the cortical regions, as well as hippocampus and amygdala, the label-map is combined with the tissue map such that the brain region volumes are determined on gray matter voxels only. For sub-cortical regions, the volumes are determined with a multi-atlas segmentation only, as the probabilistic tissue segmentation for these regions is inaccurate. A more detailed description of this method can be found in [6].

3.2.2.2 *FS*

Freesurfer is widely used neuroimaging software developed by the Laboratory for Computational Neuroimaging at the Athinoula A. Martinos Center for Biomedical Imaging at Massachusetts General Hospital. It has many applications, but in this work we use the brain region segmentation method described in [2]. The method defines the problem of segmentation using a Bayesian approach in which the probability is estimated of a segmentation given the observed image. First, the image is transformed into the atlas space with an affine transformation. Manually labeled atlas images provide the prior spatial information of the brain regions. The final segmentation is estimated by combining this spatial information with the intensity distribution of each brain region in the individual image. For more detailed information about this method we refer the reader to [2]. In our experiments we used FS version 5.1. The user is able to use his own atlas, however, we used the atlas provided by FS. This method is publicly available[3].

3.2.2.3 *GIF*

This method is atlas-based and uses the geodesic path of a spatially-variant graph to propagate the atlas labels. The atlas image database contains 130 T1-weighted MR brain images of cognitively normal participants from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study and 35 T1-weighted MR brain images from 30 young controls of the OASIS database [86]. The labeled images are made publicly available by Neuromorphometrics[4] under academic subscription, as part of the MICCAI 2012 Grand Challenge on label fusion. First, each atlas image is registered to the individual image using a non-rigid transformation. A morphological distance of this image to each atlas image is estimated using the displacement field resulting from the image registration and the intensity similarity. The segmentation is estimated by fusing the labels of the morphologically closest atlas images. For more details about this method we refer the reader to [7]. This method is publicly available[5].

3.2.2.4 *MALP-EM*

Like EMC, this method also combines multi-atlas registration and voxel-wise tissue segmentation. The atlas database of this method consists of 35 manually annotated T1-weighted MR brain images of 30 subjects of the OASIS database, which are also part of the atlas images of the GIF method (see Section 3.2.2.3). The atlas images of these 30 subjects are transformed to the space of the image that is to be segmented. These transformations are obtained via a non-rigid image registration approach [87]. The subjects' brains are extracted using the method

---

[3] `http://freesurfer.net/`

[4] `http://neuromorphometrics.com/`

[5] `http://cmictig.cs.ucl.ac.uk/niftyweb/program.php?p=GIF`

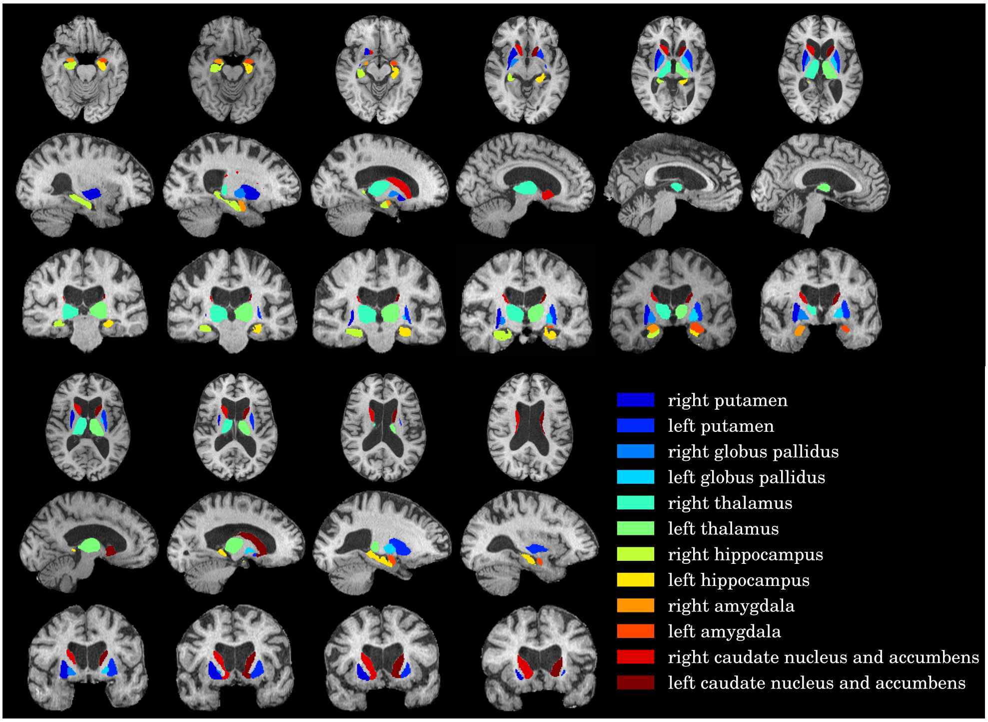


Figure 3.1: A T1w MR brain image from one of the subjects in ND, with a colored overlay of the brain regions analyzed in this work, segmented with MBS, one of the evaluated methods. Slices in the axial direction are shown in the top row, slices in the sagittal direction are shown in the middle row, and slices in the coronal direction are shown in the bottom row. The legend on the right side shows the regions and their corresponding colors in the overlay.

in [88]. The resulting 30 label images are fused and a probabilistic map of each brain region is obtained. The labels are refined through expectation-maximization (EM) [89], a brain tissue segmentation technique based on the image intensities. More details can be found in [8]. In our experiments we used MALP-EM version 1.2. This method is publicly available[6].

3.2.2.5 *MBS*

The MBS method is based on the model-based brain segmentation presented in [5]. The model is shape-constrained and represented by a triangulated mesh of fixed topology. Shape variations are modeled by principal component analysis of manually annotated meshes of a set of training images, resulting in a point distribution model (PDM) with a mean mesh and shape modes [90]. To segment a new image, the mean mesh is placed within the image by a Generalized Hough Transform compensating global translation and translation. Subsequently, the mean mesh is adapted by a global affine transformation, and then region-specific affine transformations, by adding weighted shape modes. The global and local affine transform parameters and the mode weights are estimated using a boundary detection based, e.g., on the

---

[6] `https://github.com/ledigchr/MALPEM`

Table 3.2: Number of outliers or rejected segmentations. As the outliers of the methods may overlap, the last column of the tables indicates the number of subjects included in the statistical analysis.

(a) Number of outliers in the ND subjects per method for each brain region. The outliers were defined as having an absolute z-score > 5.0, derived with the population mean and standard deviation. The ten subjects that failed in the postprocessing were not included.

	EMC	FS	GIF	MALP-EM	MBS	N
Hippocampus	0	0	0	0	0	978
Amygdala	0	1	1	0	0	976
Caudate nucleus and accumbens	2	1	0	2	0	975
Thalamus	0	1	0	0	0	977
Putamen	0	2	0	1	0	976
Globus pallidus	0	0	0	0	0	978

(b) Number of rejected segmentations in the AD subjects per method for each brain region, determined by visual inspection.

	EMC	FS	GIF	MALP-EM	MBS	N
Hippocampus	0	0	0	0	0	40
Amygdala	0	0	0	0	0	40
Caudate nucleus and accumbens	1	1	1	1	1	39
Thalamus	0	0	0	0	0	40
Putamen	0	0	0	0	1	39
Globus pallidus	0	0	0	0	1	39

local intensity gradient and a penalization component regularizing the mesh shape, including the PDM. Finally, in a deformable deformation step, triangles can adapt individually, leading to a close match of the model surface with the image boundaries.

3.2.3 *Regions of interest*

The set of brain regions in which each image is segmented differs per method. In this study we focus on the following $S = 6$ regions: hippocampus, amygdala, caudate nucleus and accumbens, putamen, thalamus, and globus pallidus. Figure 3.1 shows an example image of an ND subject with the analyzed brain regions in colored overlay. In the analysis, the volumes of the regions in the left hemisphere and the right hemisphere were summed.

For all methods except MBS, the volume of the caudate nucleus was added to the accumbens volume, because MBS already segments these as a single region.

Table 3.3: Mean (standard deviation) of brain region volumes in mm^3 for the ND subjects.

	Hippocampus	Amygdala	Caudate nucleus and accumbens	Thalamus	Putamen	Globus pallidus
EMC	3652 (494)	2289 (320)	8428 (1265)	11926 (1637)	8049 (1139)	1897 (281)
FS	7533 (1166)	2664 (402)	7995 (1154)	12328 (1614)	9008 (1338)	2834 (480)
GIF	8766 (906)	2284 (269)	7882 (1059)	12581 (1333)	9014 (1090)	1735 (207)
MALP-EM	5723 (862)	1887 (299)	7640 (1568)	13678 (1654)	7427 (1218)	2472 (349)
MBS	6052 (782)	1775 (243)	7280 (895)	12422 (1451)	7746 (977)	2561 (304)

3.2.4 *Outlier detection*

Segmentation errors may occur due to bad image quality, pathology or other method related problems. These errors could lead to outliers in the volume data and may influence the statistics excessively. We therefore remove them from the volume data prior to the statistical analyses.

The segmentations of the ND subjects were not visually inspected, as this would be too time consuming. Method failures, i.e. when the software pipeline did not result in a segmentation of the image, were excluded. On the remaining images, outliers were defined as having an absolute z-score higher than 5.0, derived with the population mean and standard deviation. Note that a z-score > 5.0 does not necessarily imply a failed segmentation.

The segmentations of the AD patients were visually inspected and obviously failed regions were excluded.

3.2.5 *Statistical analyses*

In the analyses two scenarios are considered: 1) both the normative volume distribution and the patient-specific volumes are calculated by the same method, and 2) the normative volume distribution and the patient-specific volumes are calculated by different methods. The requirements for the application of scenario 1) are:

1. a high correlation on the absolute volumes;
2. a high absolute agreement on the patient's distances relative to the normative distribution, i.e. a high absolute agreement on the patients' z-scores.

The requirements for application of scenario 2) are:

1. a high absolute agreement on the absolute volumes;
2. a high absolute agreement on the patients' z-scores.

For scenario 2) requirement i naturally results in requirement ii. The requirements for scenario 2) are more strict than those for scenario 1). The next sections describe how the normative distribution was established, how the correlation and absolute agreement are measured, and, in the case of scenario 1), how the diagnostic value of the z-scores was assessed.

3.2.5.1 *Normative distribution fitting*

We fit an age-dependent normative distribution with the previously proposed LMS method [91]. This method assumes that the data is standard normally distributed after applying the Yeo-Johnson transformation [92]. The method estimates the $\lambda-$parameter of this transformation (L), the median (M) and coefficient of variation (S) for the appropriate volume at each age. With these three parameters z-scores can be computed at each age. The smoothness of the resulting iso-z-score curves is influenced by the degrees of freedom δ, a user-defined parameter. In our experiments, we set the smoothness parameter δ to a value of

2. We used the R-package VGAM for fitting these iso-z-score curves [93].

The value of the brain region volume may also be influenced by other covariates than age, e.g. gender and height. We correct for these covariates in the fitting procedure. Figure 3.2 shows an example of a fitted distribution, represented by six iso-z-score curves.

3.2.5.2 *Correlation and absolute agreement*

To verify if scenario 1) is applicable, we first measure the correlation of the volumes calculated by the methods, with the Pearson's correlation coefficient (PCC). We refer to these correlations as PCC-v. This coefficient is invariant for an offset and scaling of the data.

To verify if scenario 2) is applicable, we compute the absolute agreement on the volumes, which was measured with the intraclass correlation coefficient (ICC). The type of ICC to be chosen depends on the problem at hand. McGraw *et al.* give an overview of the possible ICCs [94]. For the presented experiments, ICC(A,1) [94] is the appropriate absolute agreement measure. We refer to the absolute agreement on the volumes as ICC-v. The absolute agreement is maximal (1.0) when the measurements are exactly the same. When one or more measures deviate, the absolute agreement is no longer 1.0 and drops according to how large the deviation is. A systematic error, causing an offset in the measurements with a magnitude of e.g. the population standard deviation, would lower the absolute agreement to $\sim$0.67. Or a scaling of the data by a factor of 1.2 would lower the absolute agreement to $\sim$0.7. The higher the ICC-v the more reasonable it is to interchange methods.

We report all possible pairwise method combinations of PCC-v and ICC-v for $M = 5$ methods for each of the S brain regions. Since the correlation and absolute agreement are determined with symmetric measures, we present PCC-v and ICC-v of the methods in a single 5×5 table, for each of the analyzed brain regions.

3.2.5.3 *Absolute z-score agreement*

To further assess the applicability of scenario 1), we also computed the absolute agreement on the AD patient z-scores with ICC(A,1). We indicated these values with ICC-z. We present ICC-z on AD subjects with PCC-v for ND subjects (see Section 3.2.5.2) in the same table, to facilitate their comparison.

3.2.6 *AUC*

To estimate how well the AD patient z-scores discriminate between normative volumes and patient-specific volumes in scenario 1), we determine the area under the receiver operating characteristic curve (AUC) of the z-score. The z-score was computed as described in Section 3.2.5.1. The AUC is the probability that a randomly chosen ND subject will have a higher z-score than a randomly chosen AD patient. The higher the AUC, the better the discrimination between AD patients and ND subjects. Since not every region is a known discriminative

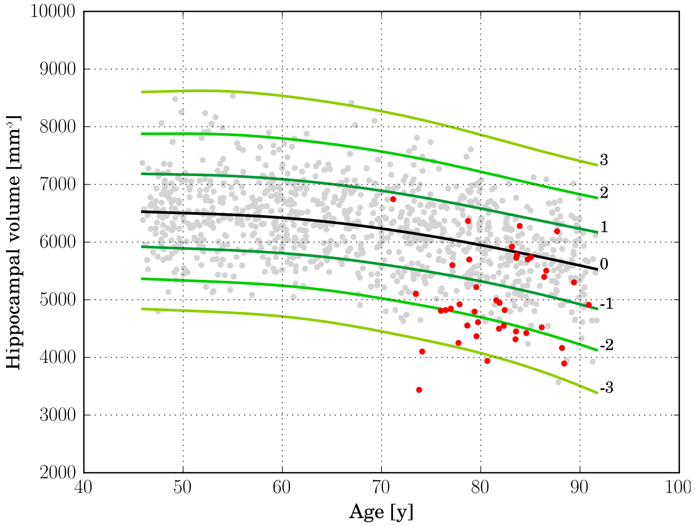


Figure 3.2: Example of a normative hippocampal volume distribution fitted on 978 ND subjects, visualized in iso-z-score lines from -3 to 3. The light gray scatters show the hippocampal volumes of the ND subjects and the red scatters show the volumes of the 40 AD patients. The volumes were generated with method MBS. The distribution was corrected for sex and height and is shown here for males of height 170cm.

biomarker for AD, it is not necessarily expected that the AUC is high for each region. Hippocampus and amygdala are known to be discriminative biomarkers for AD, so for these regions a high AUC is expected. For the computation of the AUC only ND subjects within the age range of the AD patients, [71, 91] years, were included. A 95% confidence interval was computed by bootstrapping the z-scores 1000 times.

3.3 RESULTS

We used the following rating scale for PCC-v, ICC-v and ICC-z, adopted from the rules of thumb in [95]:

- poor: < 0.5
- fair: $0.5 - 0.7$
- good: $0.7 - 0.9$
- excellent: > 0.9

3.3.1 *Outlier detection*

Method FS failed for nine ND subjects, either by not finishing the segmentation pipeline or by giving a zero volume output for some of the analyzed brain regions. The method EMC failed for one ND subject, which was due to the failure of the brain extraction tool [96], which is used at the beginning of the pipeline. The remainder of the methods provided a segmentation of all images. The number of outliers per region and method on the remaining 978 subjects is reported in Table 3.2a.

Two T1w images of AD patients were excluded due to large scanning or motion artefacts. The number of failed segmentations per region and method in the remaining 40 images are shown in Table 3.2b. In one image there was a large lesion in the frontal lobe, affecting the segmentation of the caudate nucleus and accumbens of all methods. In one other image, method MBS failed to segment the putamen and globus pallidus correctly.

3.3.2 *Volume distributions*

Table 3.3 shows the mean and standard deviation of the volumes of the ND subjects for each method and region. The hippocampus volume of methods EMC and GIF deviate substantially from the other methods. Method EMC method deviates due to a different definition of the hippocampus in the atlases that are used by the methods. The Hammers' atlas [82, 83], used by method EMC, defines the posterior border of the hippocampus such that the hippocampus tail is not included in the definition, whereas the other methods include the hippocampus tail. Method GIF deviates because it generally delineates the hippocampus in a larger volume. These same methods have a smaller average globus pallidus volume, compared to the other methods. Visual inspection on a representative subset showed that these methods delineated a smaller globus pallidus. Methods MALP-EM and MBS calculated a smaller amygdala, compared to the other methods.

3.3.3 *Correlation and absolute agreement*

Table 3.4 presents PCC-v and ICC-v for each pairwise combination of the five methods. For most regions PCC-v was good (≥ 0.75) and was excellent for region thalamus ($0.91 - 0.97$) and good to excellent for putamen ($0.88 - 0.96$).

For the three smallest structures, hippocampus, amygdala and globus palliuds, the ICC-v was generally poor, with some exceptions. Combination MALP-EM – MBS scored relatively high on ICC-v compared to the other method combinations. Visual inspection on a representative subset showed that the delineated hippocampus, amygdala and globus pallidus for MALPEM and MBS was similar in shape, explaining the good ICC-v. For amygdala, combination GIF – EMC also showed a good ICC-v. The three larger structures, caudate nucleus and accumbens, thalamus and putamen, showed generally a higher ICC-v. Visual inspection showed that their shape was on average more similar, possibly due to the less irregular shape of these regions compared to the smaller regions. Some method combinations showed poor ICC-v values for these larger regions, e.g. MBS – EMC and MBS – MALP-EM for the caudate nucleus and accumbens, and GIF – MALP-EM for the putamen. MALP-EM – MBS also had a fair PCC-v for region caudate nucleus and accumbens, however, the other combinations showed a good PCC-v, indicating that the low ICC-v can mainly be explained by a volume offset and/or scaling.

Table 3.4: PCC-v (upper-right triangle) and ICC-v (lower-left triangle) of ND volumes.

(a) Hippocampus

PCC-v ICC-v	EMC	FS	GIF	MALP-EM	MBS
EMC	1.00	0.77	0.85	0.82	0.85
FS	0.05	1.00	0.77	0.80	0.76
GIF	0.03	0.44	1.00	0.83	0.87
MALP-EM	0.13	0.30	0.12	1.00	0.79
MBS	0.10	0.33	0.14	0.73	1.00

(b) Amygdala

PCC-v ICC-v	EMC	FS	GIF	MALP-EM	MBS
EMC	1.00	0.80	0.77	0.79	0.80
FS	0.51	1.00	0.82	0.73	0.80
GIF	0.76	0.47	1.00	0.74	0.81
MALP-EM	0.43	0.20	0.37	1.00	0.75
MBS	0.29	0.15	0.27	0.68	1.00

(c) Caudate nucleus and accumbens

PCC-v ICC-v	EMC	FS	GIF	MALP-EM	MBS
EMC	1.00	0.93	0.89	0.92	0.72
FS	0.87	1.00	0.93	0.86	0.80
GIF	0.79	0.92	1.00	0.82	0.85
MALP-EM	0.78	0.80	0.75	1.00	0.55
MBS	0.44	0.62	0.71	0.46	1.00

(d) Thalamus

PCC-v ICC-v	EMC	FS	GIF	MALP-EM	MBS
EMC	1.00	0.92	0.93	0.96	0.91
FS	0.89	1.00	0.93	0.95	0.93
GIF	0.83	0.90	1.00	0.95	0.97
MALP-EM	0.61	0.71	0.73	1.00	0.95
MBS	0.86	0.92	0.96	0.71	1.00

(e) Putamen

PCC-v ICC-v	EMC	FS	GIF	MALP-EM	MBS
EMC	1.00	0.91	0.94	0.93	0.93
FS	0.69	1.00	0.91	0.88	0.90
GIF	0.69	0.89	1.00	0.88	0.96
MALP-EM	0.81	0.50	0.45	1.00	0.87
MBS	0.88	0.54	0.55	0.82	1.00

(f) Globus pallidus

PCC-v ICC-v	EMC	FS	GIF	MALP-EM	MBS
EMC	1.00	0.59	0.76	0.74	0.80
FS	0.13	1.00	0.71	0.66	0.75
GIF	0.61	0.10	1.00	0.77	0.83
MALP-EM	0.27	0.46	0.16	1.00	0.75
MBS	0.22	0.56	0.13	0.72	1.00

Table 3.5: PCC-v of the ND volumes (upper-right triangle) and ICC-z of AD volume z-scores (lower-left triangle).

(a) Hippocampus

PCC-v ICC-z	EMC	FS	GIF	MALP-EM	MBS
EMC	1.00	0.77	0.85	0.82	0.85
FS	0.61	1.00	0.77	0.80	0.76
GIF	0.69	0.56	1.00	0.83	0.87
MALP-EM	0.72	0.61	0.69	1.00	0.79
MBS	0.79	0.57	0.78	0.81	1.00

(b) Amygdala

PCC-v ICC-z	EMC	FS	GIF	MALP-EM	MBS
EMC	1.00	0.80	0.77	0.79	0.80
FS	0.80	1.00	0.82	0.73	0.80
GIF	0.85	0.70	1.00	0.74	0.81
MALP-EM	0.78	0.65	0.80	1.00	0.75
MBS	0.82	0.67	0.88	0.71	1.00

(c) Caudate nucleus and accumbens

PCC-v ICC-z	EMC	FS	GIF	MALP-EM	MBS
EMC	1.00	0.93	0.89	0.92	0.72
FS	0.85	1.00	0.93	0.86	0.80
GIF	0.87	0.90	1.00	0.82	0.85
MALP-EM	0.96	0.83	0.85	1.00	0.55
MBS	0.58	0.78	0.69	0.51	1.00

(d) Thalamus

PCC-v ICC-z	EMC	FS	GIF	MALP-EM	MBS
EMC	1.00	0.92	0.93	0.96	0.91
FS	0.85	1.00	0.93	0.95	0.93
GIF	0.88	0.75	1.00	0.95	0.97
MALP-EM	0.93	0.91	0.88	1.00	0.95
MBS	0.88	0.88	0.88	0.94	1.00

(e) Putamen

PCC-v ICC-z	EMC	FS	GIF	MALP-EM	MBS
EMC	1.00	0.91	0.94	0.93	0.93
FS	0.86	1.00	0.91	0.88	0.90
GIF	0.93	0.95	1.00	0.88	0.96
MALP-EM	0.90	0.85	0.88	1.00	0.87
MBS	0.89	0.89	0.96	0.83	1.00

(f) Globus pallidus

PCC-v ICC-z	EMC	FS	GIF	MALP-EM	MBS
EMC	1.00	0.59	0.76	0.74	0.80
FS	0.58	1.00	0.71	0.66	0.75
GIF	0.60	0.79	1.00	0.77	0.83
MALP-EM	0.69	0.50	0.62	1.00	0.75
MBS	0.72	0.69	0.71	0.60	1.00

Table 3.6: The AUC (95% confidence interval) for all regions, where the volumes of the normative distribution and the AD patients were generated by the same method (scenario 1).

	EMC	FS	GIF	MALP-EM	MBS
Hippocampus	0.78 (0.68,0.87)	0.83 (0.75,0.89)	0.73 (0.64,0.82)	0.80 (0.72,0.88)	0.80 (0.71,0.88)
Amygdala	0.77 (0.67,0.85)	0.81 (0.73,0.88)	0.76 (0.67,0.85)	0.82 (0.74,0.89)	0.73 (0.63,0.83)
Caudate nucleus and accumbens	0.52 (0.42,0.62)	0.56 (0.46,0.66)	0.47 (0.37,0.57)	0.49 (0.40,0.60)	0.63 (0.54,0.73)
Thalamus	0.68 (0.58,0.77)	0.63 (0.54,0.71)	0.76 (0.66,0.84)	0.69 (0.60,0.78)	0.66 (0.56,0.75)
Putamen	0.62 (0.52,0.72)	0.61 (0.51,0.71)	0.62 (0.52,0.72)	0.63 (0.53,0.73)	0.61 (0.51,0.71)
Globus pallidus	0.50 (0.41,0.60)	0.63 (0.52,0.74)	0.71 (0.61,0.81)	0.47 (0.38,0.56)	0.58 (0.49,0.68)

3.3.4 *Absolute z-score agreement*

Table 3.5 shows ICC-z in the lower left triangle. In the upper-right triangle, PCC-v of the ND subjects is showed again, for easy comparison. ICC-z was good to excellent for regions thalamus ($0.75 - 0.94$) and putamen ($0.83 - 0.96$), fair to good for regions hippocampus ($0.56 - 0.81$), amygdala ($0.65 - 0.88$) and globus pallidus ($0.50 - 0.72$), and fair to excellent for caudate nucleus and accumbens ($0.51 - 0.96$).

The two method combinations with the lowest PCC-v of the caudate nucleus and accumbens, MBS – EMC and MBS – MALP-EM, also have the lowest ICC-z. This is also the case for the globus pallidus, where combinations EMC – FS and MALP-EM – FS have the lowest PCC-v and the lowest ICC-v.

3.3.5 *AUC*

Table 3.6 shows the AUC for each method and brain region. The highest AUC was achieved for hippocampus (on average 0.79) and amygdala (on average 0.78), demonstrating their involvement in AD. For the thalamus and putamen, AUC > 0.5 for all methods, indicating that these regions are also affected by AD. For method GIF the AUC of regions thalamus and globus pallidus were high compared to the other methods. Methods FS, MBS and GIF had comparable thalamus volumes for the ND subjects, but the AD thalamus volumes segmented by GIF were on average 120 mm^3 lower than those segmented by MBS and 50 mm^3 lower than those segmented by FS. Methods EMC and GIF had comparable globus pallidus volumes for the ND subjects, but for AD subjects the volumes segmented by GIF were on average 320 mm^3 lower than those segmented by GIF.

3.3.6 *Computational efficiency*

All methods were executed on a Linux Sun Grid Engine (SGE) computing cluster with eight computing nodes, each having multiple cores. All methods except FS provide an option for using multiple cores. This is especially efficient for methods that use multi-atlas registration, where the registrations of the subjects in the atlas database can run in parallel. In practice, method GIF had the longest computation time, despite the usage of multiple cores. This was mainly due to the non-rigid image registrations of the 165 images in the atlas database. Method MBS was most efficient, needing only a few minutes to segment all 56 regions in a brain image on a single core. Except for MALP-EM, needing 33 GB of RAM per brain image, the memory usage of the methods was modest ($\leq$ 8GB) for the hardware in modern computers.

3.4 DISCUSSION

We evaluated the correlation and absolute agreement on regional volumes computed with different automated brain segmentation methods, and the impact of the volume differences between these methods on single-subject analysis in a normative modeling framework. We evaluated two scenarios: 1) The normative volume distributions and the patient-specific volumes were calculated by the same method, and 2) the normative volume distributions was calculated by a different method than the patient-specific volumes. To this end, we applied five state-of-the-art automated brain segmentation methods on the T1w MR brain images of 988 ND subjects and 42 AD patients acquired with the same MR acquisition protocol.

The PCC-v showed that the volumes of all regions correlated well, indicating that volume differences between methods in ND subjects are mainly due to systematic differences, such as the usage of different atlases and region definitions. The ICC-v, however, was generally low, especially for the smaller regions, hippocampus, amygdala and globus pallidus. The low ICC-v indicates that the methods cannot be interchanged in a normative modeling framework and scenario 2) is not applicable.

The ICC-z, with which the agreement on the AD patient position relative to the normative distribution was measured in case of scenario 1), was good to excellent for the thalamus and putamen, who also showed a good to excellent PCC-v. The other four regions showed lower ICC-z, indicating that different methods would result in different AD patient positions relative to the normative distribution, even when the normative distribution was computed by the same method as the patient data. A low PCC-v seemed to also result in a low ICC-z. A high PCC-v, however, does not necessarily result in a high ICC-z. This can be explained by the fact that the methods may be affected differently by the brain morphology change due to AD.

The AUC, with which the z-score discrimination between patient and normative volumes was measured in case of scenario 1), was relatively high for regions hippocampus and amygdala for all methods, demonstrating the involvement of these regions in AD. For method GIF, the thalamus volume showed to be a better discriminator for AD than the hippocampus volume, which is unexpected, as this region is not known for its involvement in AD and the other methods did not show such a high AUC for the thalamus. A possible explanation is that method GIF is more affected than the other methods by the brain morphology change due to

AD, such as larger ventricles.

Several limitations of this study should be considered. First, the segmented results rely strongly on the atlas that was used by the method. As was shown with the hippocampus, differences in volume may be largely explained by the atlas and how the region was defined. In this study, however, we considered the atlas a part of the method and we did not study specific atlas-related volume differences. Second, the number of AD patients was limited, which limits the generalization of the conclusions drawn from these results. In future studies, a higher number of AD patients should be used to generalize the study results. Third, we used images that were acquired on a single 1.5 Tesla scanner with the same acquisition protocol. This allowed us to study the effect of differences in segmentation methods, while not considering the confounding effect of differences in acquisition protocols. Future research should investigate how differences in acquisition protocols influence the comparison of individual patients to normative data, and to study the generalizability of our results in more heterogeneous data sets. Finally, this study was limited to five automated brain segmentation methods, but many more have been proposed previously and are currently being developed. Especially techniques that use deep learning showed promising results [13, 14]. These methods may achieve higher accuracy and precision and therefore the AUC of the AD patient z-scores may increase. Future studies should therefore also include deep-learning based approaches.

3.5 CONCLUSION

In this study, we aimed to answer two research questions: 1) to what extent are methods interchangeable, as long as the same method is being used for computing normative volume distributions and patient-specific volumes? and 2) can different methods be used for generating normative volume distributions and patient-specific volumes? Based on the absolute agreement results on the volume data of 988 non-demented subjects, we conclude that it is essential that the same method is used to generate normative volume distributions and patient-specific volumes. For most regions the correlation was good (> 0.75) indicating that volume differences between methods in ND subjects are mainly due to systematic differences. When the same method is being used for generating normative and patient data, we found that the agreement on the AD patient's position relative to the normative distribution (ICC-z) was high for the regions thalamus and putamen. Our results are encouraging as they indicate that methods are to some extent interchangeable for selected regions. For the regions hippocampus, amygdala, caudate nucleus and accumbens, and globus pallidus, not all method combinations showed a high ICC-z. Whether two methods are indeed interchangeable should be confirmed for the specific application and dataset of interest.

3.6 FUNDING

The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007 - 2013, project VPH-DARE@IT (grant agreement no:

601055) and from the European Union’s Horizon 2020 research and innovation programme, project EuroPOND (grant agreement no: 666992).

4

A SPATIO-TEMPORAL REFERENCE MODEL OF THE AGING BRAIN

Abstract. Both normal aging and neurodegenerative disorders such as Alzheimer's disease (AD) cause morphological changes of the brain. It is generally difficult to distinguish these two causes of morphological change by visual inspection of magnetic resonance (MR) images. To facilitate making this distinction and thus aid the diagnosis of neurodegenerative disorders, we propose a method for developing a spatio-temporal model of morphological differences in the brain due to normal aging. The method utilizes groupwise image registration to characterize morphological variation across brain scans of people with different ages. To extract the deformations that are due to normal aging we use partial least squares regression, which yields modes of deformations highly correlated with age, and corresponding scores for each input subject. Subsequently, we determine a distribution of morphologies as a function of age by fitting smooth percentile curves to these scores. This distribution is used as a reference to which a person's morphology score can be compared. We validate our method on two different datasets, using images from both cognitively normal subjects and patients with Alzheimer disease (AD). Results show that the proposed framework extracts the expected atrophy patterns. Moreover, the morphology scores of cognitively normal subjects are on average lower than the scores of AD subjects, indicating that morphology differences between AD subjects and healthy subjects can be partly explained by accelerated aging. With our methods we are able to assess accelerated brain aging on both population and individual level. A spatio-temporal aging brain model derived from 988 T1-weighted MR brain scans from a large population imaging study (age range 45.9 - 91.7y, mean age 68.3y) is made publicly available at `www.agingbrain.nl`.

This chapter contains the content of *A spatio-temporal reference model of the aging brain*, W. Huizinga et al., NeuroImage 169, pp 11–22 (2018).

4.1 INTRODUCTION

Magnetic Resonance (MR) imaging plays an important role in diagnosing neurodegenerative diseases due to its depiction of the brain morphology in vivo [97]. Interpretation of MR images in the context of dementia diagnosis can be challenging, as early brain abnormalities may be difficult to distinguish from those related to normal aging, especially in the early stages of the disease. Quantitative methods that can distinguish brain morphology due to healthy aging from morphology due to accelerated aging or pathology can therefore aid and possibly improve the diagnosis of neurodegenerative diseases [1].

Quantitative information on brain morphology is usually obtained by measuring e.g. tissue volumes and regional volumes [1]. However, these measures do not provide fully detailed information about shape differences, since volume is a quantity of an enclosed surface, and shape is a description or outline and therefore potentially much more informative [98]. Results on hippocampal shape studies suggest that shape may have additional predictive value over volume when used in the prediction of Alzheimer's disease [99, 100]. Therefore, there is an emerging interest in methods for quantifying shape differences and variations in shape in the human brain.

In literature, several methods for estimating models quantifying these shape differences and changes have been proposed. Davis *et al.* [98] proposed a kernel regression on image dissimilarities to estimate a brain image representative for each age. Both Serag *et al.* and Dittrich *et al.* use a similar or more advanced kernel regression to build a spatio-temporal atlas for neonatal and fetal brain development respectively. The latter two methods are especially suited for fetal and neonatal brain development, where the brain rapidly grows with increasing age [101, 102]. Fishbaugh *et al.* developed a geodesic shape regression method which uses a sparse representation of diffeomorphisms, describing complex nonlinear changes over time with a small number of model parameters defined by the user [103]. The mentioned methods estimate change in mean morphology of the population with age, but do not model the statistical distribution; the mean, but not the variance of the morphology at a certain age, is modelled. This concern was addressed by Ziegler *et al.*, who presented numerous approaches that relate aging to differences in brain morphometry. They considered generative models (in which brain morphology is predicted from age) and recognition models (in which age is predicted from brain morphology) in cross-sectional data, and models that estimate individual decline and explain inter-individual variability in aging in longitudinal data [104].

Rather than predicting age or brain morphology, or classifying healthy and diseased subjects based on brain morphology, we propose a method that generates a reference distribution of healthy brain morphologies as a function of age to which an individual brain can be compared. Like Marquand *et al.*, Ziegler *et al.* and Brewer we use a normative modeling approach, in which we aim to quantify the variation within a population and assess deviations from that population [1, 105, 106]. Marquand *et al.* applied normative modeling to assess deviations of brain structure or function as a function of clinical covariates (e.g. cognitive test scores) [105]. Brewer assesses deviations of volumetric MR imaging measures as a function

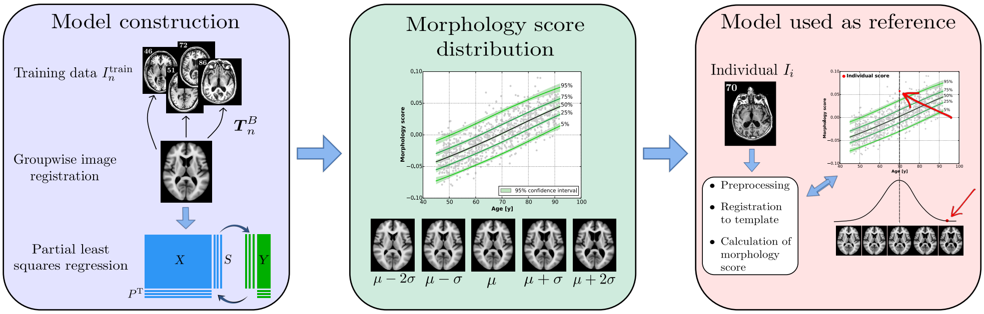


Figure 4.1: Illustration of the proposed concept. Using training data I_n^{train} a model of brain morphology as a function of age is constructed. This model is represented by a set of percentile curves and a mode of deformation. Subsequently, an individual I_i is compared to the model that is used as reference.

of age [1]. Ziegler *et al.* provide normative voxelwise maps of local gray matter abnormalities and global tissue volume z-scores [106]. In our approach we aim to assess deviations of typical aging patterns of the brain morphology, measured with voxelwise deformations. To generate a distribution of brain morphologies we aim to find brain deformation patterns that are highly correlated with age. To achieve this we choose a data-driven approach using structural MR brain scans of elderly people in a wide age range (46y - 92y). We determine the morphology distribution by applying a regression model to the morphological variation within these brains. The morphological variation is characterized by deformation fields that map each brain image to a common space, which is a standard approach in computational anatomy [107]. We compute these deformation fields with a groupwise image registration technique. Since we are only interested in the deformations due to aging, we employ a regression technique called partial least squares regression (PLSR). PLSR is especially suitable when there are more predictors than samples and when the predictors are highly collinear or linearly dependent [108, 109]. It was first evaluated for neuroimaging by Krishnan *et al.* to relate brain function to behavior [110]. Ziegler *et al.* used multivariate PLS correlation to explore the relationship between cognitive ability patterns and differences in local brain anatomy in the maturing brain [111]. Singh *et al.* used PLSR to quantify anatomical shape variation in the brain. They used kernel PLSR to find the relationship between the manifold of diffeomorphisms from atlas to subject domain and global cognitive and functional assessment test scores [112]. Whereas Singh *et al.* were interested in the PLSR regression coefficient, we aim to find deformation patterns that are most correlated with age, i.e. the PLSR loadings. With each of these loadings comes a corresponding score, and we use these scores to quantify the distribution of brain morphologies due to aging. The density of a distribution can be indicated with percentile values: measures specifying the value below which a given percentage of observations in a group of observations fall. The morphology distribution, however, is a function of age, and therefore we fit percentile *curves* to quantify how this distribution varies with age. To assess if an individual suffers from accelerated brain aging, its score can be compared to these percentile curves.

To validate our method, we use 988 structural MR brain scans from the population-based

Rotterdam Scan Study (RSS), a prospective study among community dwelling subjects aged 45 years and over [37]. The method's robustness against scanning protocol and its diagnostic value is evaluated using the 988 scans from the RSS and a selection of 509 scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset which is adopted from the study of Cuignet *et al.* [24]. The morphology score distribution and the corresponding 4D atlases are made publicly available through a web-based application at `www.agingbrain.nl`.

4.1.1 *Extension of preliminary results*

Preliminary results of this method were presented at SPIE Medical Imaging, San Diego 2016 [113]. In the current work, we considerably expand the previous study. First, in this version we use the displacement field inside the brain instead of the transformation parameters, in order to exclude deformation outside the brain. The control points of the B-spline transformation model that exist outside the brain region influence the deformation field both outside and inside the brain. We do not want deformations in the background to be part of the modelling, which is why we chose the displacement field inside the brain. This voxel-based approach also makes the method more generalizable to other nonparametric registration methods and other voxelwise tissue property maps. Second, we correct for the subject's head position in the scanner by removing rigid body motion inside the brain mask from the deformation field. Third, we introduce a method to determine the number of PLSR components. Fourth, we analyze the residuals that contain deformations due to factors other than aging, e.g. unexpected pathologies. Fifth, we take into account other covariates such as sex and height. Sixth, we added an evaluation of our method on the ADNI dataset.

4.2 METHODS

We propose a method for modeling brain morphology and its distribution over the population as a function of age. The model is constructed using N training images I_n^{train}, $n \in \{1 \dots N\}$, from a population-based cross-sectional data collection. This population-based model is then used as a reference to which an individual brain image I_i can be compared. The concept of the proposed method is shown in Figure 4.1. The sections below explain all the steps of the proposed framework in detail: 4.2.1) preprocessing, 4.2.2) non-rigid groupwise image registration, 4.2.3) elimination of translation and rotation, 4.2.4) partial least squares regression, 4.2.5) percentile curve fitting, 4.2.6) spatio-temporal atlas construction, and 4.2.7) individual subject assessment.

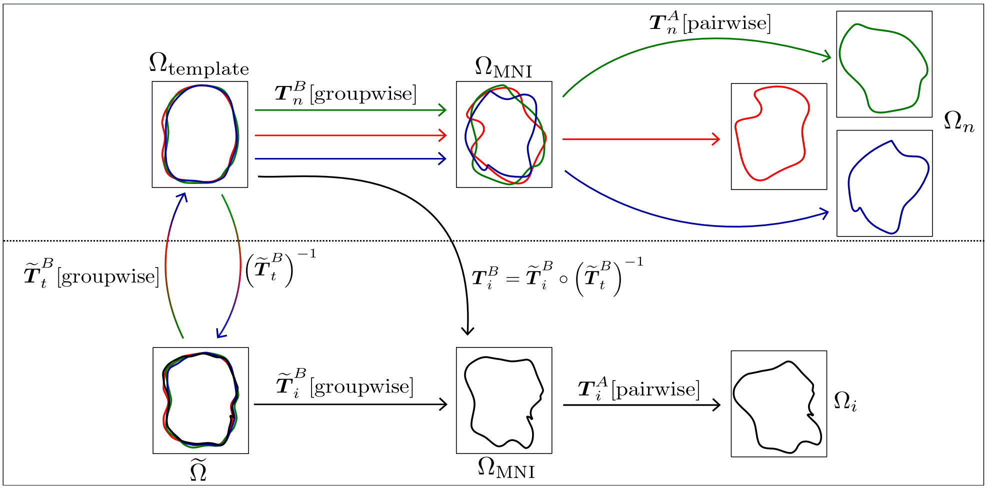


Figure 4.2: Scheme showing the different transformations and domains of the registration framework of the proposed method. The top part shows three different training images in in red, green and blue in their domains Ω_n, and how the template domain is constructed from the training data. The bottom half shows how an individual I_i in Ω_i is registered to the template domain. After each transformation is indicated if it was obtained using a pairwise or a groupwise registration.

4.2.1 *Preprocessing*

The preprocessing steps are 1) non-uniformity correction of the images using the N3 algorithm [84] and 2) brain extraction using a multi-atlas method described in Bron *et al.* [6], with a set of 30 atlases [82, 83].

4.2.2 *Non-rigid groupwise image registration*

The morphological variation is characterized by deformation fields that map each brain image to a common domain, the template domain Ω_{template}. Let $\boldsymbol{x}$ be an image coordinate in $\mathbb{R}^3$. The function $I_n^{\text{train}}(\boldsymbol{x})$ gives the intensity of image n at $\boldsymbol{x}$, i.e. $I_n^{\text{train}}(\boldsymbol{x}) : \Omega_n \subset \mathbb{R}^3 \to \mathbb{R}$. Let N be the total number of images in the training set. The aim of the groupwise registration is to find a set of coordinate transformations $\boldsymbol{T}_n(\boldsymbol{x}) : \Omega_{\text{template}} \to \Omega_n$, $n \in \{1 \ldots N\}$, such that the warped images $I_n^{\text{train}}\left(\boldsymbol{T}_n\left(\boldsymbol{x};\boldsymbol{\mu}_n\right)\right)$ are aligned with each other in the template domain. The image registration is performed with a parametric approach. The degrees of freedom of the transformation is limited by introducing a parameterization to the transformation: $\boldsymbol{T}_n\left(\boldsymbol{x};\boldsymbol{\mu}_n\right)$, where $\boldsymbol{\mu}_n$ is a vector containing the transformation parameters of subject n. In our method, these transformations are determined in two steps. First, we obtain a coarse alignment of all images in the training set via transformations $\boldsymbol{T}_n^A(\boldsymbol{x};\boldsymbol{\mu}_n^A)$. Then we use a non-rigid transformation model, $\boldsymbol{T}_n^B(\boldsymbol{x};\boldsymbol{\mu}_n^B)$, for a more precise alignment of the images. Figure 4.2 shows an overview of all transformations. This section explains the transformations and domains shown in the top-half of this figure. As shown in the top-right part of Figure 4.2, transformations $\boldsymbol{T}_n^A(\boldsymbol{x};\boldsymbol{\mu}_n^A)$

map $\boldsymbol{x}$ from MNI domain Ω_{MNI} [114] to the subject-specific domains Ω_n. Second, as shown in the top-left part of Figure 4.2, we obtain $\boldsymbol{T}_n^B : \Omega_{\text{template}} \to \Omega_{\text{MNI}}$, where $\boldsymbol{T}_n^B$ is parameterized by $\boldsymbol{\mu}_n^B$. The total transformation from Ω_{template} to Ω_n is a composition of $\boldsymbol{T}_n^A$ and $\boldsymbol{T}_n^B$.

The $\boldsymbol{T}_n^A$ are parameterized by an affine transformation model. The parameters $\boldsymbol{\mu}_n^A$ are found by performing a pairwise registration of the brain mask of each subject to the reference brain mask in the MNI domain.

The $\boldsymbol{T}_n^B$ are parameterized by cubic B-splines [115]. We chose cubic B-splines, because their compact support property makes the computation efficient, which is relevant in our large-scale groupwise image registration problems. The spacing of the control points of the B-splines is a setting with which the degrees of freedom of the transformation can be controlled. The parameters of the B-spline transformation model, $\boldsymbol{\mu}_n^B$, are B-spline control point coefficients. To obtain $\boldsymbol{\mu}_n^B$, we use groupwise image registration. During such a registration, Ω_{template} is implicitly defined by constraining the sum of all deformations from the template to each subject to be zero, an approach proposed by Bhatia *et al.* [116] and Balcie *et al.* [117]. To achieve this, $\boldsymbol{\mu}_n^B$ must be optimized simultaneously for all $n \in \{1 \ldots N\}$. Advantages of groupwise registration are that the information of all images is taken into account during the registration and, as opposed to pairwise registration, the result is not biased towards any chosen reference image.

For the non-rigid groupwise image registration we use the method of Huizinga *et al.* [118]. This method was designed for intra-subject registration of images originating from a quantitative MRI experiment. It is therefore robust against arbitrary intensity scaling between the aligned images. In this work we investigate if the method is applicable to inter-subject registration as well. The method assumes that, when images are registered, the intensities can be mapped to a low-dimensional subspace. The dimension of this low-dimensional subspace depends on the model describing the intensity variation in the aligned images. When this subspace is assumed to be one-dimensional, the intensity may vary due to a global scale or shift, which is the case for our application. In this case, the method could be considered as an extension of normalized cross-correlation from pairwise ($N = 2$) to groupwise ($N > 2$) settings.

In our experiments, the image registration was performed with Elastix [119]. We used a multi-resolution strategy with four resolutions in which the control point spacing of the B-spline transformation model is halved with each resolution step, until a final spacing of 10 mm. The final spacing was determined heuristically.

4.2.2.1 *Individual registration to the template domain*

To be able to compare the brain image of an individual, I_i, to the reference model, we need a deformation field that maps I_i to Ω_{template}. Similarly as during the template construction from training images, this is done in two steps. This section explains the transformations and domains shown in the bottom-half of Figure 4.2. First, we register I_i to Ω_{MNI} yielding an affine transformation $\boldsymbol{T}_i^A$. Second, we seek a non-rigid transformation $\boldsymbol{T}_i^B$ that maps I_i from Ω_{MNI} to Ω_{template}.

A possible approach to finding $\boldsymbol{T}_i^B$ would be a pairwise registration of I_i with one of the registered training images in Ω_{template} chosen as a reference. However, this could introduce a bias towards this chosen reference. Another possiblity would be to perform a pairwise registration of I_i to the mean of the registered training images, $\bar{I}(\boldsymbol{x}) = \frac{1}{N}\sum_n I_n^{\text{train}}\left(\boldsymbol{T}_n\left(\boldsymbol{x}\right)\right)$ for $\boldsymbol{x} \in \Omega_{\text{template}}$, however, the disadvantage of such an approach is that $\bar{I}$ is blurry at the cortex edges which hampers accurate registration at those locations. Therefore, we use a different approach that is visualized in Figure 4.2 and exlained below.

We propose to formulate the non-rigid registration of I_i to the template domain as an additional non-rigid groupwise registration, involving $N+1$ images, namely the N previously registered training images I_n^{train} and the individual image I_i. As shown in the bottom-left part of Figure 4.2, this procedure leads to N transformations $\widetilde{\boldsymbol{T}}_n^B(\boldsymbol{x};\widetilde{\boldsymbol{\mu}}_n^B) : \widetilde{\Omega} \rightarrow \Omega_{\text{template}}$ and one transformation $\widetilde{\boldsymbol{T}}_i^B(\boldsymbol{x};\widetilde{\boldsymbol{\mu}}_i^B) : \widetilde{\Omega} \rightarrow \Omega_{\text{MNI}}$, where $\widetilde{\Omega}$ is the common domain of I_n^{train} and I_i. Since the training images I_n^{train} had already been aligned before the registration, it is safe to assume that their transform parameters $\widetilde{\boldsymbol{\mu}}_n^B$ associated with $\widetilde{\Omega}$ are approximately equal. A single, unbiased transformation, from $\widetilde{\Omega}$ to Ω_{template} $\widetilde{\boldsymbol{T}}_t^B : \widetilde{\Omega} \rightarrow \Omega_{\text{template}}$, is obtained by averaging $\widetilde{\boldsymbol{\mu}}_n^B$ over all n, obtaining a single transform parameter vector $\widetilde{\boldsymbol{\mu}}_t^B$:

$$\widetilde{\boldsymbol{\mu}}_t^B = \frac{1}{N}\sum_n \widetilde{\boldsymbol{\mu}}_n^B \tag{4.1}$$

To bring I_i to Ω_{template} we finally compute $\boldsymbol{T}_i^B : \Omega_{\text{template}} \rightarrow \Omega_{\text{MNI}}$ as:

$$\boldsymbol{T}_i^B = \widetilde{\boldsymbol{T}}_i^B\left(\left(\widetilde{\boldsymbol{T}}_t^B\right)^{-1}\left(\boldsymbol{x};\widetilde{\boldsymbol{\mu}}_t^B\right);\widetilde{\boldsymbol{\mu}}_i^B\right) \tag{4.2}$$

where $\boldsymbol{T}^{-1}\left(\boldsymbol{x};\boldsymbol{\mu}\right)$ is the inverse of $\boldsymbol{T}\left(\boldsymbol{x};\boldsymbol{\mu}\right)$, obtained using the procedure described in [120]. In total one pairwise registration, the affine registration to MNI space, and one groupwise registration, the non-rigid registration to Ω_{template}, are required to analyze a new image I_i.

4.2.3 *Elimination of translation and rotation*

As global brain shrinkage could be (partially) captured by the affine transformation $\boldsymbol{T}^A$, we consider the composition of the affine and non-rigid transformations, $\boldsymbol{T}^A\left(\boldsymbol{T}^B\left(\boldsymbol{x}\right)\right)$. To focus on brain morphology only, we propose to extract the rigid body motion, like the arbitrary orientation of the subject's head in the scanner, from this composition. In this way, scale and skew transformations are preserved, and any rigid body motion present in $\boldsymbol{T}^B$ is also eliminated.

For any rigid transformation $\boldsymbol{T}_n^R$ parameterized by $\boldsymbol{\mu}_n^R$, the residual deformation is defined by:

$$\boldsymbol{d}_n^*\left(\boldsymbol{x};\boldsymbol{\mu}_n^R\right) = \boldsymbol{T}_n^R\left(\boldsymbol{T}_n^A\left(\boldsymbol{T}_n^B\left(\boldsymbol{x};\boldsymbol{\mu}_n^B\right);\boldsymbol{\mu}_n^A\right);\boldsymbol{\mu}_n^R\right) - \boldsymbol{x}. \tag{4.3}$$

The rigid transformation parameters of each subject are estimated by:

$$\hat{\boldsymbol{\mu}}_n^R = \arg\min_{\boldsymbol{\mu}_n^R} \frac{1}{|\Omega_{\text{mask}}|} \sum_{\boldsymbol{x}\in\Omega_{\text{mask}}} \left\| \boldsymbol{d}_n^* \left(\boldsymbol{x}; \boldsymbol{\mu}_n^R\right) \right\|^2, \tag{4.4}$$

where Ω_{mask} is the domain containing voxels inside the brain mask in Ω_{template} to ensure that we only evaluate the deformation inside the brain. In the statistical analysis that follows we will be using $\boldsymbol{d}_n^*(\boldsymbol{x}) \hat{=} \boldsymbol{d}_n^* \left(\boldsymbol{x}; \hat{\boldsymbol{\mu}}_n^R\right)$ for $\boldsymbol{x} \in \Omega_{\text{mask}}$. The same approach is applied to the individual deformation field $\boldsymbol{d}_i$.

4.2.4 *Partial least squares regression*

We aim to correlate the deformations obtained by image registration with age. Let N be the number of subjects in the training set and let $|\Omega_{\text{mask}}| = M$, then $\boldsymbol{X}$ is the $N \times 3M$ matrix of which each row contains $\boldsymbol{d}_n^*(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Omega_{\text{mask}}$. Let $\boldsymbol{Y}$ be the $N \times 1$ vector containing the corresponding ages of each subject, then PLSR finds the directions in the deformation space that explain the maximum covariance with age. Let $\boldsymbol{X}_0 = \left(\boldsymbol{X} - \overline{\boldsymbol{X}}\right)$ and let $\boldsymbol{Y}_0 = \left(\boldsymbol{Y} - \overline{\boldsymbol{Y}}\right)$, where $\overline{\boldsymbol{X}}$ and $\overline{\boldsymbol{Y}}$ are the column-wise data means replicated to all N rows. In PLSR $\boldsymbol{X}_0$ is decomposed into:

$$\boldsymbol{X}_0 = \boldsymbol{S}\boldsymbol{P}^{\mathrm{T}} + \boldsymbol{E}. \tag{4.5}$$

Let L be the number of components used in the PLSR, then $\boldsymbol{S}$ are the $N \times L$ scores and $\boldsymbol{P}$ are the $3M \times L$ orthonormal loadings. Matrix $\boldsymbol{E}$ is the $N \times 3M$ residual matrix. The decomposition of $\boldsymbol{X}_0$ is made by maximizing the covariance between $\boldsymbol{Y}_0$ and a weighted sum of $\boldsymbol{X}_0$. The weight vector $\boldsymbol{W}^j$ for each component $1 \leq j \leq L$ is estimated by solving the following optimization problem:

$$\max_{\boldsymbol{W}^j} \left[\left(\boldsymbol{X}_0 \boldsymbol{W}^j\right)^{\mathrm{T}} \boldsymbol{Y}_0 \right] \tag{4.6}$$

subject to $\left(\boldsymbol{X}_0 \boldsymbol{W}^j\right)^{\mathrm{T}} \boldsymbol{X}_0 \boldsymbol{W}^j = 1$. The weight vector is found by a singular value decomposition on $\boldsymbol{R} = \boldsymbol{X}_0^{\mathrm{T}} \boldsymbol{Y}_0$ [109, 121]. The PLSR scores for component j are defined as $\boldsymbol{S}^j = \boldsymbol{X}_0 \boldsymbol{W}^j$, and the loadings of component j are defined as $\boldsymbol{P}^j = \boldsymbol{X}_0^{\mathrm{T}} \boldsymbol{S}^j$. Covariance $\boldsymbol{R}$ is deflated with each iteration of the algorithm to obtain $\boldsymbol{W}^j$, $\boldsymbol{S}^j$ and $\boldsymbol{P}^j$ for $j \geq 2$.

The number of components used in PLSR is a tradeoff between overfitting (too many components) or losing valuable information (not enough components). To determine the optimal number of components we propose to use the randomization test of Wiklund *et al.* [122]. In this test a null-distribution of $\left(\boldsymbol{S}^j\right)^{\mathrm{T}} \boldsymbol{Y}_0$ is determined by randomly permuting $\boldsymbol{Y}$. When the probability of finding the observed $\left(\boldsymbol{S}^j\right)^{\mathrm{T}} \boldsymbol{Y}_0$ is smaller than α, the component is significant. Since PLSR is prone to overfitting, we chose a conservative significance level of $\alpha = 0.01$. Results of the randomization test showed that only the first PLSR component was significant (see Sec. 4.4.1). We therefore only used the scores of the first PLSR component, $\boldsymbol{S}^1$, to describe the brain morphology distribution as a function of age, and from here on, we omit the superscript

1.

To compare an individual subject to the reference population, the morphology score of this individual subject, S_i, is required. To this end, the individual deformation field $\boldsymbol{d}_i^*$ is projected on $\boldsymbol{W}$ to obtain S_i:

$$S_i = \left(\boldsymbol{d}_i^* - \overline{\boldsymbol{X}}\right) \boldsymbol{W}. \tag{4.7}$$

where $\overline{\boldsymbol{X}}$ is a vector of size $3M$ with the column-wise means of $\boldsymbol{X}$.

Since an individual's brain morphology may not only change due to aging but also due to pathology we also propose to compare the individual's residual to that of the model. Each row of the residual matrix $\boldsymbol{E}$ contains the non-age-related deformations for a subject in I_n^{train}. The residual norm of I_n^{train} is defined as:

$$\|\boldsymbol{E}\|_n = \sqrt{\sum_{m=1}^{3M} E_{nm}^2} \tag{4.8}$$

where E_{nm} is element (n, m) of matrix $\boldsymbol{E}$. The residual of an individual is defined as:

$$\boldsymbol{E}_i = \boldsymbol{d}_i^* - S_i \boldsymbol{P}^{\mathrm{T}} \tag{4.9}$$

Inspection of $\boldsymbol{E}_i$ is important since an individual could have a different morphology due to other factors than aging. If this is the case, it would not be visible by the individual's morphology score. We therefore propose to compare the individual residual norm $\|\boldsymbol{E}\|_i$ to the distribution of $\|\boldsymbol{E}\|_n$ in I_n^{train}. If $\|\boldsymbol{E}\|_i$ is significantly different from the distribution of $\|\boldsymbol{E}\|_n$ further inspection of $\boldsymbol{d}_i^*$ is necessary.

To perform the PLSR, we implemented the SIMPLS algorithm [109] in Python.

4.2.5 *Percentile curve fitting*

The score of subject n in I_n^{train} is referred to as S_n. To visualize the distribution of scores S_n as function of age we fit percentile curves. These curves show both the distribution of the morphology scores and how they vary with age. We refer to the $p^{\text{th}}\%$ percentile curve at age a as $s(a, p)$.

For fitting of percentile curves to the morphology score data, we use the LMS method [123]. The LMS method assumes that the data is standard normally distributed after applying the Yeo-Johnson transformation, which is an extension of the Box-Cox transformation proposed by Cole and Green [123]. This method estimates the $\lambda-$parameter of the Yeo-Johnson transformation [124] (L), the median (M) and coefficient of variation (S) for the appropriate morphology score at each age. With the parameters L, M, and S, percentiles can be computed at each age to obtain a smooth curve. The smoothness of the fitted curves is influenced by the degrees of freedom δ, a user-defined parameter. In our experiments, we set the smoothness parameter δ to a value of 2 and we deployed the R-package VGAM [125] for the percentile curve fitting.

The value of the morphology score may also be influenced by other covariates than age, e.g. sex or height, since those covariates may influence the head size and may therefore affect brain scaling in the deformation fields that are used in the PLSR. It is therefore desirable to correct the reference distribution for these covariates. We model the correction for these two covariates as a linear shift in the morphology score distribution.

The precision of the estimated percentile curves depends on the number of data points in the appropriate age range. If the data is non-uniformly distributed over age, it could be that the curve estimation is not precise in the part where there are very few data points. To assess the precision of the fitted curves, we use a bootstrapping procedure, by random sampling subjects with replacement and re-estimating the percentile curves. A distribution of possible curves is collected, from which confidence intervals at any significance level can be estimated [126].

4.2.6 *Spatiotemporal atlas construction*

The deformation having the highest covariance with age is contained in the loading vector $\boldsymbol{P}$. To be able to interpret the morphology score it is necessary to know what this deformation looks like, and therefore we aim to visualize $\boldsymbol{P}$ as a spatio-temporal atlas. An estimate of the age-related morphology of subject n is obtained by multiplying score S_n with $\boldsymbol{P}$. Instead of multiplying with just one specific score S_n, we multiply $\boldsymbol{P}$ with $s(a,p)$ for a chosen range of a and a specific p. We choose to show the aging trajectory of $\overline{I}$. We convert the deformation field in $s(a,p)\boldsymbol{P}$ to B-spline transformations and invert these to obtain $\boldsymbol{T}^{a,p}_{\text{atlas}}(\boldsymbol{x})$. The spatio-temporal atlas for percentile p and a chosen range of a is then constructed by warping $\overline{I}$ for each a:

$$\overline{I}^{a,p}(\boldsymbol{x}) = \overline{I}\left(\boldsymbol{T}^{a,p}_{\text{atlas}}\left(\boldsymbol{x};\boldsymbol{\mu}\right)\right), \tag{4.10}$$

4.2.7 *Individual subject assessment*

Given the distribution of morphology scores and residual magnitudes of the reference data, outliers from these distributions can be assessed in terms of percentiles, which is analogous to the use of growth charts to map child development in terms of height and weight as a function of age [73]. Let p_i be the percentile at which the individual morphology score S_i at age a_i can be found in the reference morphology score distribution, after correction for available covariates. Let $p_{\|\boldsymbol{E}\|_i}$ be the percentile at which the individual residual magnitude can be found in the reference residual magnitude distribution, then the status of the brain morphology of patient i could, for example, be assessed according to the following set of rules:

1. No accelerated brain aging: $0.05 < p_i < 0.95$ and $0.05 < p_{\|\boldsymbol{E}\|_i} < 0.95$
2. At risk of accelerated brain aging: $p_i > 0.95$ and $0.05 < p_{\|\boldsymbol{E}\|_i} < 0.95$
3. At risk of unknown pathology: $p_{\|\boldsymbol{E}\|_i} > 0.95$

In addition, the individual scan I_i should be compared to the spatio-temporal atlas $\overline{I}^{a_i,p_i}$, for a qualitative assessment of the morphology due to aging. In the third case, further investigation of I_i, $\boldsymbol{E}_i$, and the individual deformation $\boldsymbol{d}_i^*$ is necessary.

4.3 EXPERIMENTS

In our experiments we used two image databases:

1. The Rotterdam Scan Study (RSS), containing brain scans of non-demented, asymptomatic subjects,
2. The Alzheimer's Disease Neuroimaging Initiative (ADNI) database, containing both asymptomatic and symptomatic subjects.

First, the model was applied and validated on the RSS database. Next, the difference in scores between asymptomatic and symptomatic subjects was evaluated in the ADNI dataset. Finally, the robustness of the scores across databases was evaluated.

4.3.1 *Data*

4.3.1.1 *Rotterdam Scan Study*

We used 988 T1w scans (433 male, age=68.3±13.0 (mean±SD)) from the population-based RSS, a prospective longitudinal study among community dwelling subjects aged 45 years and over [37]. Participants with dementia at the time of MRI were excluded [127]. All brain scans were acquired on a single 1.5T MRI system (GE Healthcare, US). The T1w imaging protocol was a 3-dimensional fast radiofrequency spoiled gradient recalled acquisition with an inversion recovery pre-pulse sequence [37]. The voxel size was $0.5\times0.5\times0.8$ mm^3. Besides age and sex, height information of the participants was available as well. For six participants, the height variable was missing, for which we substituted the average height (= 170 cm). We will refer to this dataset as RSS988.

4.3.1.2 *ADNI*

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see `www.adni-info.org`.

The ADNI cohort used in this article is adopted from the study of [24], consisting of an AD patient group, an MCInc group (mildly cognitive impaired but not converted to AD within 18 months), an MCIc (mildly cognitive impaired and converted to AD within 18 months), and a cognitive normal group (CN). The inclusion criteria for participants were defined in the ADNI-GO protocol[1]. The AD group consisted of 137 patients (67 male, age=76.0±7.3 years) (AD137), the MCInc group of 134 participants (84 male, age=74.4±7.2 years) (MCInc134), the MCIc group of 76 participants (43 male, age=74.7±7.4 years) (MCIc76), and the CN group of 162 participants (76 male, age=76.2±5.4 years) (CN162). Acquisition had been performed according to the ADNI acquisition protocol [26]. The brain scans were acquired on 1.5T MRI systems (GE Healthcare, Philips Medical Systems, Siemens Medical Solutions) and the T1w imaging protocol was a 3-dimensional magnetization prepared rapid acquisition gradient echo sequence. The voxel size was approximately 1 mm^3, with a maximum of 1.5 mm in any direction. Besides age and sex, height information of the partipants was available as well. For one participant, the height variable was missing, for which we substituted the average height of the subjects in RSS988 (= 170 cm). We will refer to the entire dataset as ADNI509.

4.3.2 *Morphology distribution RSS988*

We trained the model on RSS988 to visualize the main age-related deformations in a healthy reference population. Sex and height were used as covariates.

After the preprocessing step, the brain images were cropped to the bounding box of the mask and resampled to 1.5×1.5×1.5 mm^3 spacing. The 988 obtained deformation fields $\boldsymbol{d}^*$, which are used as input of the PLSR, were downsampled to a 3×3×3 mm^3 spacing and their size was 52×66×55 voxels. Since the number of parameters with which the deformation field was generated was much smaller than the number of voxels in the downsampled field, the downsampling will not influence the result.

The number of PLSR components was determined by applying the randomization test to RSS988, using 1000 randomizations of $\boldsymbol{Y}$. In addition, we inspected the score distributions, the explained variance in age, and deformation modes (PLSR loadings) of the first ten components.

4.3.3 *Morphology distribution ADNI509*

We trained the model ADNI509 and fitted percentile curves on the scores from each class separately to see if the morphology distribution is different for the various classes in the ADNI database. Sex and height were used as covariates.

After the preprocessing step, the brain images were cropped to the bounding box of the mask and resampled to 1.5×1.5×1.5 mm^3 spacing, to reduce computation time and memory consumption of the groupwise registration. The obtained 509 deformation fields, which are used as input for the PLSR, were downsampled to a 3×3×3 mm^3 spacing and their size was 59×66×55 voxels.

---

[1] `http://adni.loni.usc.edu/wp-content/themes/freshnews-dev-v2/documents/clinical/ADNI_Go_Protocol.pdf`

4.3.4 *Model validation*

We validated the model using data from the RSS database, in which we trained the model on a subset of 888 images, RSS888, and tested on the remaining 100 images, RSS100. We also validated the model on data from different databases, in which we trained the model on RSS988and tested on the AD137 and CN162.

4.3.4.1 *Leave-100-out validation (1)*

The groupwise image registration on RSS988 resulted in 988 brain deformation fields. In this experiment we evaluate the generalizability of the PLSR outside the training data. Therefore we randomly selected 888 deformation fields as input of the PLSR. The resulting PLSR weights are used to compute individual scores on the remaining 100 deformation fields using Equation (4.7).

4.3.4.2 *Leave-100-out validation (2)*

In this experiment we trained the entire model only on RSS888: I_n^{train} = RSS888. The remaining 100 subjects were treated as entirely new individuals, I_i = RSS100. RSS100 was registered to Ω_{template}, which was constructed from RSS888, and 100 S_i were computed. We evaluated if the scores from the leave-100-out experiment (1) could be reproduced.

4.3.4.3 *Individual subject comparison*

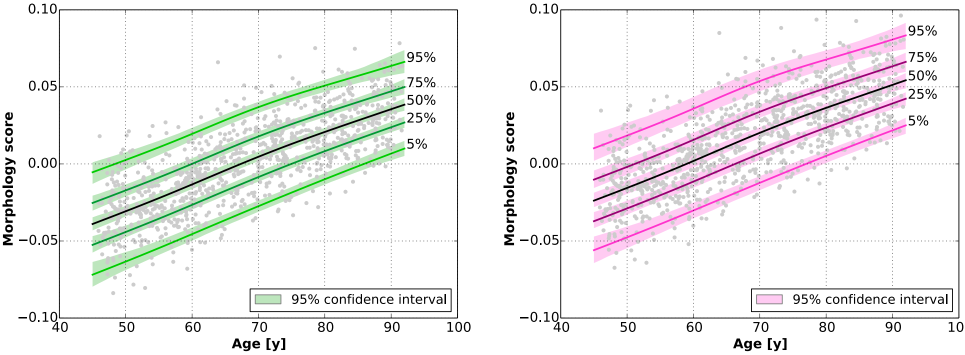
In this experiment the training dataset was I_n^{train} = RSS988, and the individual subjects were I_i^{CN} = CN162 and I_i^{AD} = AD137, respectively. We compared the individual scores S_i^{CN} and the S_i^{AD} to the morphology score distribution of RSS988, while taking into account the covariates sex and height. The goals of this experiment were:

- To evaluate if the individual comparison can be performed when individual subjects are scanned on different scanners with different scanning protocols.
- To evaluate if healthy subjects from different populations have the same brain morphology distribution.
- To evaluate if the AD subjects have different morphology scores than the healthy subjects from a different population.

4.4 RESULTS

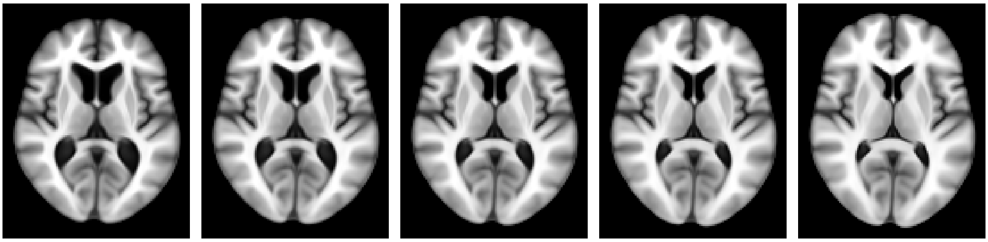
4.4.1 *Morphology distribution RSS988*

Results from the randomization test indicated that only the first component was significantly different from the null-distribution ($p = 2.6 \cdot 10^{-10}$). The second component was not significant ($p = 1.4 \cdot 10^{-2}$). Figure 4.4 shows the explained variance in age of the first ten PLSR



(a) Morphology score distribution for RSS1000 (male, 170 cm).

(b) Morphology score distribution for RSS1000 (female, 150 cm).



(c) -0.063 (d) -0.032 (e) 0 (f) 0.032 (g) 0.063

Figure 4.3: Morphology scores S_n for all n and $s(a, p)$ for $p \in \{5, 25, 50, 75, 95\}$ for (b) male subjects of height 170 cm of RSS988, and (b) female subjects of height 150 cm of RSS988. The 95% confidence intervals were determined with 1000 bootstraps. (c)-(g) The main mode of deformation where S equals -2σ, $-\sigma$, 0, σ, and 2σ, with σ the standard deviation of S_n for $n \in \{1 \ldots 988\}$.

components. This figure shows that the first component explains most variance in age (∼60%) and that the following components do not add much information, which is in agreement with the results of the randomization test. The score distributions of components two to ten showed very little to no relation with age and the deformation modes did not contain clear patterns that can be expected in aging. We therefore only used the scores of the first PLSR component to describe the brain morphology distribution as a function of age.

Figure 4.3(a) shows S_n of the first PLSR component and the fitted percentile curves $s(a, p)$ for $p \in \{5, 25, 50, 75, 95\}$, for male subjects of height 170 cm, the average height of all subjects in RSS988. Figure 4.3(b) shows S_n of the first PLSR component and the fitted percentile curves for female subjects of height 150 cm, to show the effect of the covariates on the morphology score distribution. The distributions have a clear relation with age. As expected, the scores of short, female subjects are higher than the scores of male subjects of average height. Figure 4.3(c)-(g) shows the main mode of deformation in $\boldsymbol{P}$ applied to $\bar{I}$ for RSS988. This mode shows that the higher S_n, the larger the ventricles, (cortical) atrophy, and brain shrinkage.

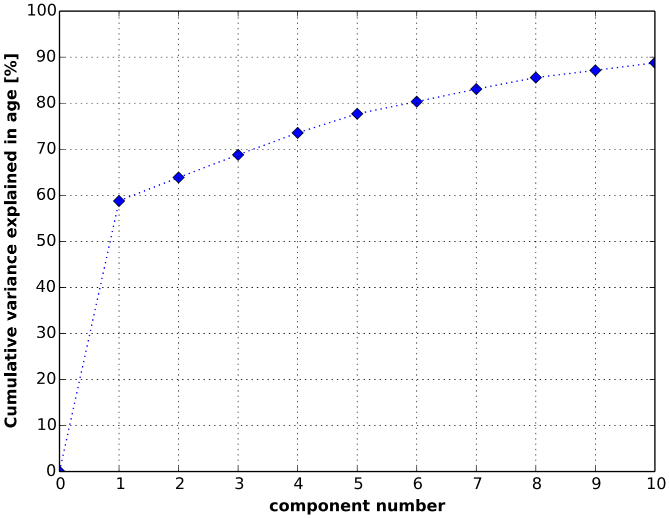


Figure 4.4: Cumulative explained variance [%] of the first ten PLSR components.

4.4.2 *Morphology distribution ADNI509*

Figure 4.5 shows S_n and $s(a, p)$ for $p \in \{5, 25, 50, 75, 95\}$, adjusted for covariates sex and height, of the model trained on ADNI509. The scores of CN162 are the lowest, and MCIc76 and AD137 have the highest scores. Interestingly, a morphology difference between MCInc134 and MCIc76 is visible. The confidence bounds of the median curves of AD137 and CN162 do not overlap, indicating that the median curve of the two groups are significantly different. The confidence bounds of the median curves of MCInc134 and MCIc76 group overlap slightly, possibly due to the lower number of subjects in MCIc76.

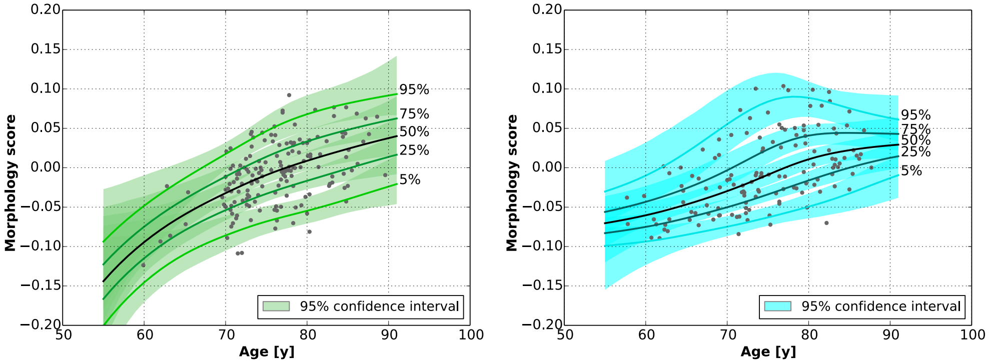
4.4.2.1 *Distributions of $\|\boldsymbol{E}\|$ in ADNI509*

The distributions of $\|\boldsymbol{E}\|$ are shown in Figure 4.5(f). The Welch's two-sample t-test was performed to test if the distributions of $\|\boldsymbol{E}\|$ are significantly different. Tests were performed between all possible group pairs. All p-values were higher than the significance level of 0.05, and therefore no significant difference between the distributions was observed.

4.4.3 *Model validation*

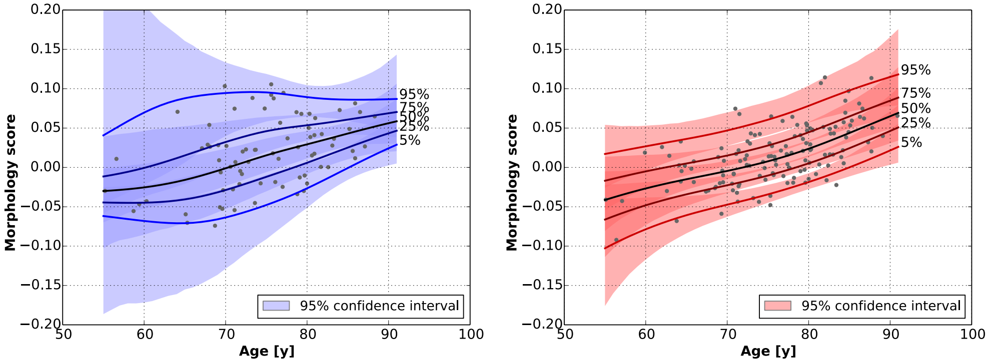
4.4.3.1 *Leave-100-out validation (1)*

Figure 4.6a shows the morphology scores for RSS888 grey and RSS100 test subjects in red. The morphology scores of the test subjects fall within the morphology score distribution of the training subjects.



(a) Morphology score distribution of CN162 (male, 170 cm).

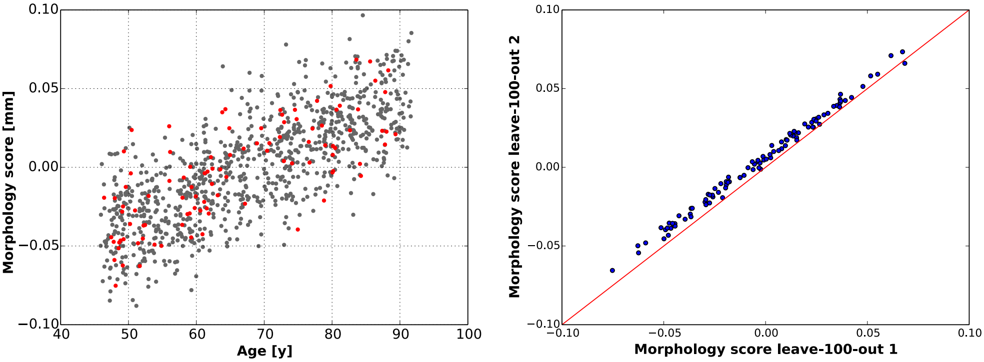
(b) Morphology score distribution of MCInc134 (male, 170 cm).



(c) Morphology score distribution of MCIc76 (male, 170 cm).

(d) Morphology score distribution of AD137 (male, 170 cm).

Figure 4.5: In figures 4.5a - 4.5d we see the morphology score distribution of each subgroup in ADNI509. The 95% confidence intervals are estimated using 1000 bootstraps. At the age extremities not many data points are available resulting in wide confidence intervals.



(a) Morphology scores for 900 training subjects in gray and 100 test subjects in red, obtained in leave-100-out validation (1).

(b) Morphology scores of the 100 test subjects from leave-100-out validation (1) vs the scores from leave-100-out validation (2).

Figure 4.6: Morphology scores from leave-100-out validation experiments.

4.4.3.2 *Leave-100-out validation (2)*

Figure 4.6b shows a scatterplot of the morphology scores of the leave-100-out (1) experiment versus the leave-100-out (2) experiment. The morphology scores show a high correlation (Pearson's $r = 0.996$), but a small bias is present. The distribution of $\|\boldsymbol{E}\|$ of RSS888 is not significantly different from the distribution of $\|\boldsymbol{E}\|$ of RSS988 according to the Welch's two-sample t-test, as shown in Figure 4.7.

4.4.3.3 *Individual subject comparison*

Figure 4.8 shows the S_i^{CN} and S_i^{AD} projected onto the model of RSS988 (male, 170 cm). Figure 4.8(a) shows that 94% of the CN162 lie below the 95% percentile lines of the RSS distribution.

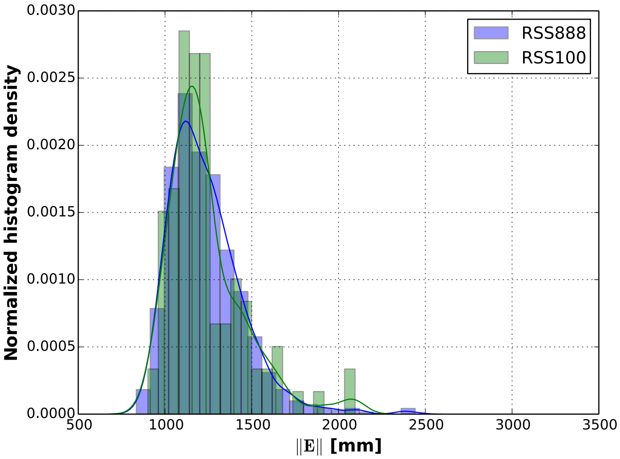
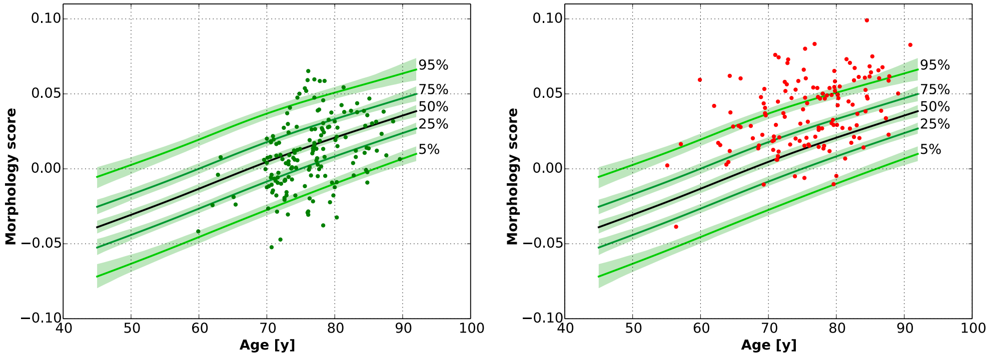


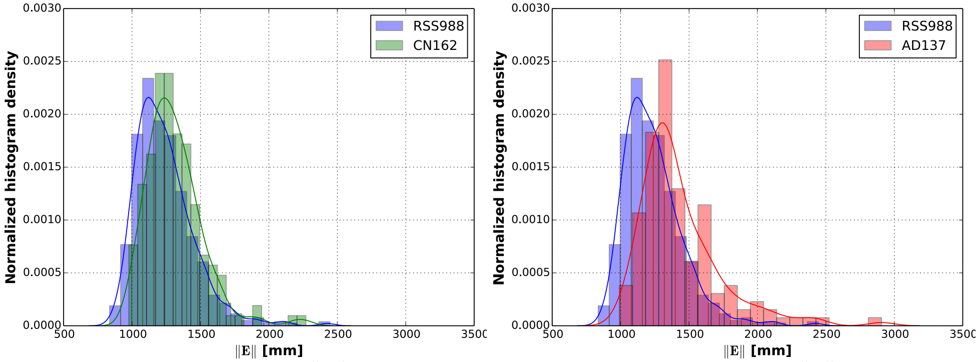
Figure 4.7: Distributions of $\|\boldsymbol{E}\|$ of training set RSS888 (blue) and test set RSS100 (green).



(a) S_i^{CN} are shown in green.

(b) S_i^{AD} are shown in red.

Figure 4.8: Morphology score distribution (male, 170 cm) of RSS1000 in five percentile curves and S_i^{CN} (a) and S_i^{AD} (b). The S_i^{CN} and S_i^{AD} are corrected and projected onto the reference distribution.



(a) Normalized histogram of $\|\boldsymbol{E}\|$ for RSS1000 and CN162.

(b) Normalized histogram of $\|\boldsymbol{E}\|$ for RSS1000 and AD137.

Figure 4.9: Distributions of $\|\boldsymbol{E}\|$ where the RSS images are I_n^{train} for $n \in \{1, \ldots, 1000\}$ and the ADNI images were I_i^{CN} for $i \in \{1, \ldots, 162\}$ and I_i^{AD} for $i \in \{1, \ldots, 137\}$.

From Figure 4.8(b) we see that 34% of AD137 lie above the 95% percentile, and 76% above the 75% percentile, indicating that their morphology shows more atrophy than that of cognitive normals at the same age.

Figure 4.9 shows the distributions of $\|\boldsymbol{E}\|$ of RSS988 subjects and CN162 (a) and AD137 (b). The distributions have overlap, but are significantly different according to the Welch's two-sample t-test.

4.5 DISCUSSION

We proposed a method for developing a spatio-temporal model of morphological differences in the brain due to normal aging, to which an individual's brain morphology can be compared. We applied the framework to a set of 988 images of non-demented aging subjects from a large population imaging study and on a set of 509 subjects from the case-control study ADNI. The main mode of deformation due to aging, $\boldsymbol{P}$, shows the expected deformation patterns due to aging: larger ventricles, (cortical) atrophy, and brain shrinkage. We performed various experiments to validate the proposed method. The results of these experiments are encouraging and show clinical potential.

The leave-100-out experiment (1) showed that the PLSR generalizes for subjects that were not part of the training set. The leave-100-out experiment (2) showed that we can reproduce the morphology scores from leave-100-out experiment (1). This indicates that the deformations resulting from the individual subject registration to the template are similar to the deformations following from the groupwise registration, and that the method can be used to compute individual morphology scores. We did observe a slight bias when comparing the scores of leave-100-out experiment (1) and leave-100-out experiment (2). This is due to the fact that the starting point of the registration in leave-100-out experiment (1) is different from the starting point of the registration in leave-100-out experiment (2), leading to slightly different deformation fields.

The percentile curves computed on the ADNI database showed a difference between the morphology score distributions of the CN and AD subgroups (median difference significant, i.e. no overlapping confidence bands, for the age range 60-72y), and, interestingly, also between MCInc and MCIc (median difference not significant, i.e. confidence bands overlap). Although it was not specifically trained to separate the groups, the model showed that the AD and MCIc patients have, on average, higher morphology scores than the CN and MCInc groups. The distributions of the norm of the PLSR residual, $\|\boldsymbol{E}\|$, between the four groups was not significantly different, indicating that the non-age related deformations were of similar magnitude.

To evaluate the reproducibility across datasets we computed scores of both AD and cognitive normal individuals from the ADNI database using the RSS subjects as reference data. The cognitive normal subjects from ADNI fell nicely within the morphology score distribution of the cognitive normal subjects from RSS. The AD subjects had on average higher morphology scores than the cognitively normal subjects. These results suggest that the proposed method is able to compute individual morphology scores of subjects from one population when trained on subjects from another population. We did observe that $\|\boldsymbol{E}\|$ is different for the two databases of scans, which we did not see when registering subjects from the same population. Possible causes for this are the presence of non-age-related morphology differences between the two populations, e.g. differences in scanner type or protocol, inclusion criteria, demographics, environmental factors, etc.

4.5.1 *Suggestions for future work*

Several improvements can be made to the proposed model to increase sensitivity and performance. Also, the proposed method can be used in a variety of applications.

The uncertainty in the percentile curves was quantified by bootstrapping, obtaining confidence intervals for each curve. The uncertainty on individual morphology scores due to registration errors was not taken into account in the presented method. However, this could be estimated by perturbation of the individual deformation field, according to estimations of registration uncertainty. In this way, a distribution of possible morphology scores for this individual could be obtained. Obtaining reliable estimates of registration uncertainty is, however, still an active topic of research [128–130].

Overall, the experiments showed that the proposed method is valid when comparing individual brain morphologies to a (cognitively) healthy reference population. Diagnostic value of the morphology score alone, however, is limited due to the high variability between individuals. In this study, we applied our method to T1-weighted MRI brain scans, but in principle, it could be applied to scans of other sequences or modalities. Future work may investigate age and other demographics related changes in other voxelwise maps, such as diffusion or perfusion imaging derived maps. This can be accomplished by replacing the deformation field $\boldsymbol{d}^*$ with such a voxelwise map in the template domain. This could possibly improve discriminative ability between people at risk of accelerated brain aging and people without risk.

Besides application of this method to voxelwise maps, it can also be applied to selected points on the surface of a segmented anatomical structure of interest, for example the hippocampus. In that case, a spatio-temporal model and its variation in the population of that specific structure can be studied in more detail and individual anatomical structures can be compared to a reference shape distribution.

It is clinically relevant to follow how a person's brain morphology changes with age. The proposed method can be used to compute morphology scores of individual scans at baseline and at follow-up. Plotting the baseline and follow-up scores in the percentile curves allows comparing the aging trajectory of an individual to the (cross-sectional) reference population. Following individual morphology scores over time allows estimation of the physiological variation in subject specific aging trajectories. On an individual level, inspection of these aging trajectories could be relevant in clinical trials to, for example, investigate if changes in e.g. medication or lifestyle have any effect. A clinical evaluation of the proposed deformation-based framework would be necessary to prove its value in the clinic.

4.5.2 *Limitations*

A limitation of using cross-sectional data, which was also mentioned by [104], is that effects of different birth cohorts are not excluded.

Because PLSR tries to maximize the covariance between the morphology scores and age, and the number of predictor variables is high ($= 3M \gg N$), it is very likely that it will find a linear relationship. Therefore, when modeling a population, and the relation with age is not expected to be linear, this may be a limitation of PLSR.

To compute the score of an individual brain image an additional groupwise registration of all images in the training set plus the individual brain image(s) has to be performed. This additional groupwise registration is time-consuming and therefore this is a practical limitation of our proposed framework. In the current implementation, the registration takes about two days and has to be performed on a machine capable of reserving 100GB of RAM to complete the task. In the future we will search for possibilities to register an individual brain image to the template domain without having to perform an additional groupwise registration with the already registered training data. In the presented work, however, the groupwise registration is preferred over a pairwise registration for obtaining an unbiased result. This is supported by preliminary experiments which are not shown due to lack of space.

4.6 CONCLUSION

We developed a spatio-temporal model of morphological differences in the brain due to normal aging. The method provides a representative distribution of brain morphologies as a function of age instead of a single population mean morphology. Our method reduces high dimensional morphology to a single score. This score can be interpreted using the spatio-temporal atlas showing which deformation due to aging belongs to that score.

The framework was tested using data from two different datasets. Experiments showed that the proposed method extracts the expected deformation patterns due to aging and they showed that on a group level there is a morphology difference between cognitively normal and AD subjects, which manifests as accelerated aging, indicating the potential at least for clinical group studies. The spatio-temporal model can be used to compare an individual's brain morphology to a cognitively healthy reference population. Smooth percentile curves showing the brain morphology changes as a function of age as well as spatio-temporal atlases derived from the cognitively healthy reference population (RSS988) are publicly available via an interactive web application at `www.agingbrain.nl`. We believe that this framework has the potential to be used clinically as an indicator of accelerated brain aging.

4.7 FUNDING

The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007 - 2013, project VPH-DARE@IT (grant agreement no: 601055) and from the European Union's Horizon 2020 research and innovation programme, project DynaHEALTH (grant agreement no: 633595), project EuroPOND (grant agreement no: 666992), and from the European Research Council (ERC), project ORACLE (grant agreement no: 678543).

This research is supported by the Dutch Technology Foundation STW (12723), which is part of the NWO, and which is partly funded by the Ministry of Economic Affairs.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen;

Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (`www.fnih.org`). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Part II

Efficient non-rigid groupwise image registration

5

PCA-BASED GROUPWISE IMAGE REGISTRATION FOR QUANTITATIVE MRI

Abstract. Quantitative magnetic resonance imaging (qMRI) is a technique for estimating quantitative tissue properties, such as the T_1 and T_2 relaxation times, apparent diffusion coefficient (ADC), and various perfusion measures. This estimation is achieved by acquiring multiple images with different acquisition parameters (or at multiple time points after injection of a contrast agent) and by fitting a qMRI signal model to the image intensities. Image registration is often necessary to compensate for misalignments due to subject motion and/or geometric distortions caused by the acquisition. However, large differences in image appearance make accurate image registration challenging. In this work, we propose a groupwise image registration method for compensating misalignment in qMRI. The groupwise formulation of the method eliminates the requirement of choosing a reference image, thus avoiding a registration bias. The method minimizes a cost function that is based on principal component analysis (PCA), exploiting the fact that intensity changes in qMRI can be described by a low-dimensional signal model, but not requiring knowledge on the specific acquisition model. The method was evaluated on 4D CT data of the lungs, and both real and synthetic images of five different qMRI applications: T_1 mapping in a porcine heart, combined T_1 and T_2 mapping in carotid arteries, ADC mapping in the abdomen, diffusion tensor mapping in the brain, and dynamic contrast-enhanced mapping in the abdomen. Each application is based on a different acquisition model. The method is compared to a mutual information-based pairwise registration method and four other state-of-the-art groupwise registration methods. Registration accuracy is evaluated in terms of the precision of the estimated qMRI parameters, overlap of segmented structures, distance between corresponding landmarks, and smoothness of the deformation. In all qMRI applications the proposed method performed better than or equally well as competing methods, while avoiding the need to choose a reference image. It is also shown that the results of the conventional pairwise approach does depend on the choice of this reference image. We therefore conclude that our groupwise registration method with a similarity measure based on PCA is the preferred technique for compensating misalignments in qMRI.

This chapter contains the content of *PCA-based groupwise image registration for quantitative MRI*, W. Huizinga et al., Medical Image Analysis 29, pp 65 - 78 (2016).

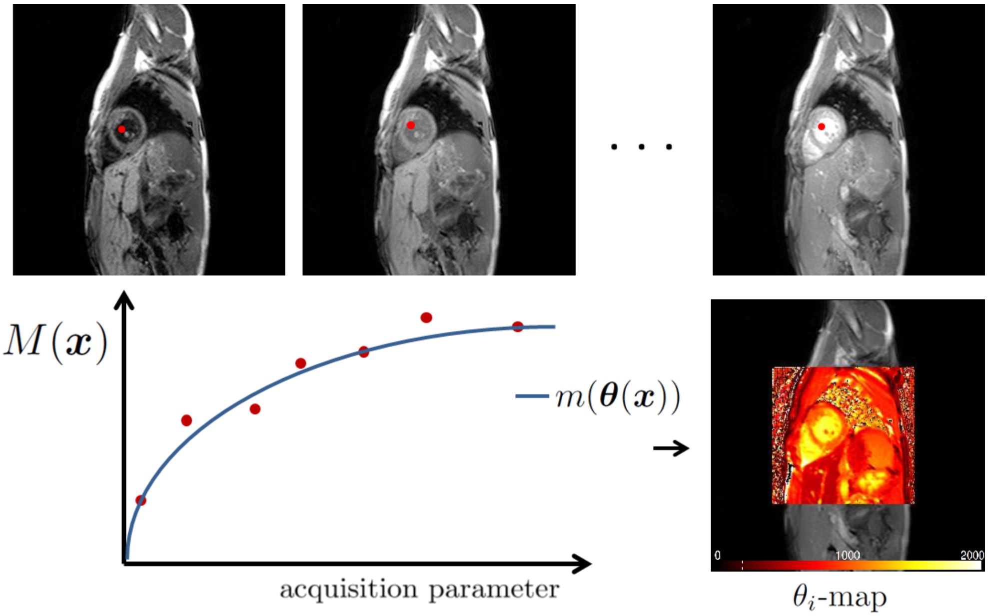


Figure 5.1: A scheme showing the concept of qMRI. On the top there are three contrast-varying images of a heart. The red dots indicate one specific pixel location $\boldsymbol{x}$ in each of the images and their intensities are plotted in the graph below the images. A model $m(\boldsymbol{\theta})$ is fitted to the pixel intensities, where $\boldsymbol{\theta}$ contains the tissue properties of interest. The model is fitted for each pixel in the image so that a map of the element(s) in $\boldsymbol{\theta}$ can be created. Such a map, showing T_1, is shown on the right of the graph.

5.1 INTRODUCTION

Quantitative magnetic resonance imaging (qMRI) is a technique that enables the estimation of tissue properties from a series of images acquired with different imaging parameters or acquired at multiple time points after injection of a contrast agent. Because these tissue properties can be indicators of the biological state of the tissue and their change during disease, their precise and accurate estimation is important. Examples of such tissue properties are the relaxation parameters T_1 and T_2, the mean diffusivity (MD), the apparent diffusion coefficient (ADC) and K^{trans}, a measure of capillary permeability. They are estimated by fitting a low-dimensional signal model (qMRI model) to the acquired MR images, see e.g. [131]. Typically five to over a hundred images are acquired, depending on the tissue properties of interest. Figure 5.1 illustrates the concept of qMRI.

The fitting procedure assumes an anatomical correspondence between the images in the acquired series. However, due to patient motion and/or geometric distortions caused by the acquisition this correspondence can be lost, which may lead to erroneous parameter estimation, especially at tissue boundaries. Corrections during acquisition, such as gating or breath-holding, do not always give the desired effect and can significantly increase the acquisition time. Another solution is to align the images prior to fitting the qMRI model. This alignment can be achieved with image registration techniques. However, image registration for qMRI imposes two main challenges: firstly, the contrast is different for each of the images in the acquired series, complicating registration based on the image intensities. Secondly, often more than two images need to be registered in the context of qMRI. In this second case one can choose a pairwise registration approach in which all images are registered to a chosen reference image. To deal with contrast changes that occur in a series of qMRI images the pairwise approach is commonly used with a metric based on mutual information (MI), because this metric is robust against intensity changes in the images [6, 132, 133]. However, a major disadvantage of this approach is that the choice of reference image will influence the result of the registration, which we will demonstrate in this paper. To circumvent the need to choose a reference image one can use a so-called groupwise registration approach. In such an approach all images are simultaneously registered to a mean space. Moreover, with this approach the information of all images is taken into account during the registration. This improves consistency over a pairwise registration approach, as shown in [120] for groupwise registration of dynamic CT images.

We distinguish two categories of registration methods for qMRI data, discussed in the following two paragraphs: a) model-based methods that use the qMRI model to register the images, [134–138] and b) data-driven methods, which do not rely on the qMRI model [139, 140]. The model-based registration method proposed in [134] for the registration of T_1 data, uses the fitted parameters of the qMRI model to generate reference images for all images in the dataset. The registration is done in a pairwise fashion using a cross-correlation similarity metric. A similar approach was used by Buonaccorsi *et al.* for dynamic contrast-enhanced (DCE) images [135]. Andersson *et al.* and Hallack *et al.* directly minimize the residual error of the qMRI model fit [136, 137]. For the registration of DCE images Bhushan *et al.* propose to maximize the joint posterior probability between the intensities estimated by the model and the true data [138]. All these approaches eliminate the requirement to choose a reference image and are robust to the appearance differences in the image. However, these model-based methods assume that the images adhere exactly to the qMRI model, which is not always true in all structures that are present in the images due to noise and acquisition artefacts or when the model is too simple to represent the image intensities. For example, the DT model may not fully describe the signal in all voxels of the brain due to the presence of multiple fiber orientations, as shown in [141] and [142].

Data-driven methods for qMRI registration have been proposed [139, 140]. For the registration of DCE data, Hamy *et al.* proposed robust data decomposition. With the assumption that the low rank components are free from local contrast changes or artefacts, they use a sparse and low-rank decomposition and register the low-rank components to the mean of all low-rank components in the current resolution [139]. Melbourne *et al.* proposed a progressive principal component registration for DCE data. This method registers the series of images to an artificial series, which is generated from the first principal components of the original images. The registration is repeated and at each iteration a new series is generated and more principal components are added in the reconstruction [140].

Groupwise registration methods, not specifically developed for the registration of qMRI data, have been proposed in [143], [144], [120], [145], and [116]. Liu *et al.* use a sparse/low-rank decomposition [143], similar to the method proposed in [139], to register brain images between subjects, with the assumption that the low-rank components are free from lesions and pathologies. Miller *et al.* proposed a method based on voxel-wise entropy [144]. While this method is robust against intensity variation, qMRI acquisitions often have few measurements, making it difficult to estimate a well-defined probability density function, which is needed for entropy calculation. Metz *et al.* proposed a groupwise dissimilarity metric based on voxel-wise variance, implicitly assuming small intensity differences between the images [120]. Due to the large intensity variations in qMRI it is not expected that this method is suitable for qMRI applications, however we investigate if this is indeed the case. Wachinger *et al.* proposed a sum of accumulated pairwise estimates (APE), such as the sum of normalized correlation coefficients of all possible image pairs in the series, as a similarity metric [145]. Even though this method has not yet been validated for qMRI data, it is designed to align images with different contrasts. Bhatia *et al.* proposed a groupwise extension of the pairwise mutual information registration approach, in which the sum of the mutual informations between the voxel-wise mean of all images and each image is maximized [116]. A downside of such a metric for qMRI is that contrast at edges may be reduced after computing the mean image, due to the large intensity differences among the images.

In this paper we propose a generic data-driven groupwise registration approach which by design is suitable for a wide range of qMRI applications without explicitly requiring the applicable qMRI model. In our approach we exploit the fact that in qMRI the intensity changes according to a low-dimensional acquisition model. When the images are not aligned, the complexity of the data is increased, i.e. the data can no longer be described by the acquisition model. We propose two dissimilarity metrics based on this principle. These metrics use PCA to quantify the amount of misalignment in the qMRI series. Because of the groupwise formulation of the registration, the need to choose a reference image is eliminated. We evaluated the generic applicability of the method on five different challenging applications of qMRI: T_1 mapping in a porcine heart, combined T_1 and T_2 mapping in carotid arteries, ADC mapping in the abdomen, diffusion tensor mapping in the brain and dynamic contrast-enhanced mapping in the abdomen. As the groupwise approaches in [145], [116], [120] have potential for qMRI data we use them as reference methods. The model-based approach in e.g. [137] is implemented for T_1 mapping and will be used as a reference method for this application. Finally we also compare our method to the commonly used MI-based pairwise registration approach.

5.2.1 *Quantitative magnetic resonance imaging*

In a qMRI examination multiple images are acquired in the same subject. Let $M_g(\boldsymbol{x})$ for $g \in \{1 \dots G\}$ be a series of G images and $\boldsymbol{x}$ a 2D or 3D spatial coordinate. The intensity at $\boldsymbol{x}$ for each image M_g is predicted by a low-dimensional qMRI model m_g:

$$M_g(\boldsymbol{x}) = m_g(\boldsymbol{\theta}(\boldsymbol{x})) + \epsilon(\boldsymbol{x}), \tag{5.1}$$

where $\boldsymbol{\theta}$ is a $\mathit{\Gamma}$-dimensional vector with tissue properties and ϵ is the noise at coordinate $\boldsymbol{x}$. In our applications, the number of tissue properties $\mathit{\Gamma}$ ranges from three to seven and $\mathit{\Gamma} < G$, which is why we call the qMRI model low-dimensional. Each qMRI application follows a different model m_g. An example of such a qMRI model m_g is the modified Look-Locker inversion recovery model proposed by [146]:

$$m_g(\boldsymbol{\theta}) = \left| A \left(1 - Be^{-TI_g/T_1^*} \right) \right|, \tag{5.2}$$

where $\boldsymbol{\theta} = \left(A, B, T_1^*\right)$ and TI_g the inversion time for image M_g. The parameter of interest, T_1, is calculated using $T_1 = T_1^*(B-1)$. For this specific model $\mathit{\Gamma} = 3$, as $\boldsymbol{\theta}$ contains three parameters. The functions m_g that are used for the experiments in this paper are presented in Section 5.3.

5.2.2 *Registration frameworks: pairwise and groupwise*

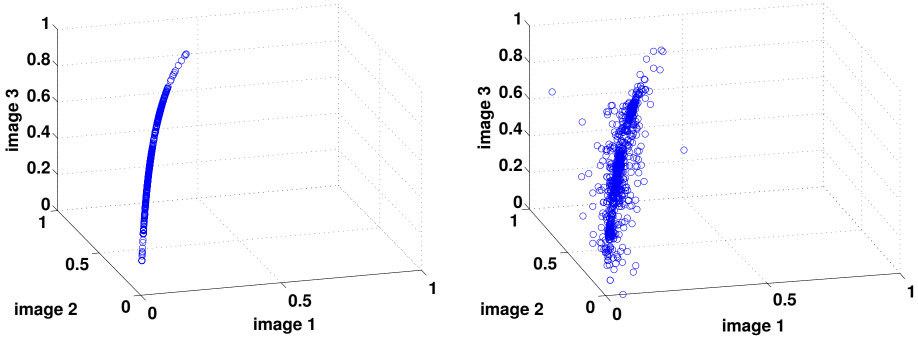
Prior to estimating $\boldsymbol{\theta}$ the images M_g need to be registered. Due to motion and/or geometric distortions caused by the acquisition correspondence between the images is lost, i.e. a spatial coordinate $\boldsymbol{x}$ does not correspond to the same anatomical location in each of the images. In the pairwise registration approach, one reference image $M_R(\boldsymbol{x})$ is chosen and all other images $M_g(\boldsymbol{x})$ for $g \neq R$ are registered to $M_R(\boldsymbol{x})$. In our registration method, the transformation is modeled by a set of transform parameters $\boldsymbol{\mu}$. For each image $M_g(\boldsymbol{x})$ there is a transformation $\boldsymbol{T}_g(\boldsymbol{x}; \boldsymbol{\mu}_g)$. The pairwise registration is formulated as the minimization of a dissimilarity metric $\mathcal{D}$ with respect to $\boldsymbol{\mu}_g$:

$$\hat{\boldsymbol{\mu}}_g = \underset{\boldsymbol{\mu}_g}{\arg\min}\, \mathcal{D}(\boldsymbol{\mu}_g), \tag{5.3}$$

which is repeated for all $g \neq R$. $\mathcal{D}$ measures the dissimilarity of $M_R(\boldsymbol{x})$ and $M_g(\boldsymbol{T}_g(\boldsymbol{x}; \boldsymbol{\mu}_g))$.

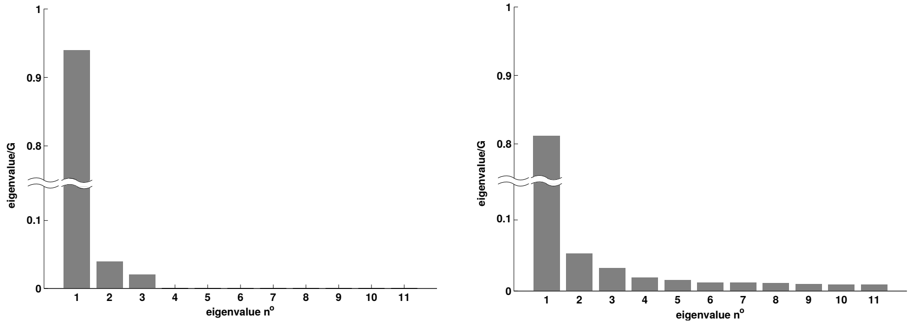
In the groupwise registration framework the images $M_g(\boldsymbol{x})$ for all g are registered simultaneously to a mean space. We formulate groupwise registration as the minimization of a dissimilarity metric $\mathcal{D}$ with respect to $\boldsymbol{\mu}$:

$$\hat{\boldsymbol{\mu}} = \underset{\boldsymbol{\mu}}{\arg\min}\, \mathcal{D}(\boldsymbol{\mu}), \tag{5.4}$$



(a) Aligned (b) Misaligned

Figure 5.2: Scatter plots of three aligned images (5.2a) and three misaligned images (5.2b) that are generated with a one-dimensional non-linear model.



(a) Aligned (b) Misaligned

Figure 5.3: Eigenvalue spectra of aligned (5.3a) and misaligned (5.3b) set of images. Note that $\Gamma = 3$, so we expect three dominant eigenvalues in the aligned case. Note that in (a) eigenvalues 4 - 11 are not exactly zero but very small ($< 10^{-3}$) and therefore not visible in the plot.

where $\boldsymbol{\mu}$ is a vector containing all $\boldsymbol{\mu}_g$. Here, $\mathcal{D}$ measures the dissimilarity of all transformed images $M_g(\boldsymbol{T}_g(\boldsymbol{x};\boldsymbol{\mu}_g))$ with respect to each other. The parameters $\boldsymbol{\mu}_g$ are simultaneously optimized for all g.

5.2.3 *Proposed dissimilarity metrics*

We present two novel groupwise dissimilarity metrics. Let the images M_g be represented as columns of an $N \times G$ matrix $\mathbf{M}$, where N is the number of voxels in one image M_g. A row of $\mathbf{M}$ can be considered as a data point in a G-dimensional space. Note that when M_g is noise-free for all g, these datapoints lie in a - possibly non-linear - Γ-dimensional subspace, where Γ is the number of free qMRI model parameters. Figure 5.2a shows an intensity scatter plot of three images generated using a one-dimensional ($\Gamma = 1$) non-linear model: $m_g(\boldsymbol{\theta}) = 1 - e^{-at}$, where $\boldsymbol{\theta} = a$. We see that the points lie on a curved line, i.e. a non-linear Γ-dimensional subspace. Figure 5.2b illustrates the effect of a small misalignment: the intensity scatter plot becomes more dispersed.

The correlation matrix of the data points in $\mathbf{M}$ is defined as:

$$\mathbf{K} = \frac{1}{N-1}\mathbf{\Sigma}^{-1}\left(\mathbf{M} - \overline{\mathbf{M}}\right)^{\mathrm{T}}\left(\mathbf{M} - \overline{\mathbf{M}}\right)\mathbf{\Sigma}^{-1}, \tag{5.5}$$

where $\mathbf{\Sigma}$ is a diagonal matrix with the standard deviations of each column of $\mathbf{M}$ as diagonal elements and $\overline{\mathbf{M}}$ is a matrix with in each element of that column the column-wise average of $\mathbf{M}$. The dimension of the subspace can be estimated by an eigenvalue decomposition of $\mathbf{K}$ i.e., by a PCA. The key idea behind the proposed dissimilarity metric is that, when motion is present in the images, the data no longer adheres to the presumed qMRI model and the eigenvalue spectrum of $\mathbf{K}$ changes. We choose to perform PCA on the correlation matrix $\mathbf{K}$, instead of the covariance matrix, to be insensitive to arbitrary intensity scaling between images. Figure 5.3a shows the eigenvalue spectrum of $\mathbf{K}$ for an aligned set of synthetic, noise-free images, created with the qMRI model of Eq. (5.2), and the deformed set of images (Figure 5.3b). We can observe that once the images are deformed, the higher eigenvalues increase. Our method aims to transform the images M_g such that the eigenvalue spectrum of $\mathbf{K}$ approaches the spectrum of an aligned set of images.

Let λ_j be the j^{th} eigenvalue of $\mathbf{K}$, with $\lambda_j > \lambda_{j+1}$, i.e. the eigenvalues with a lower index have a higher value. The first dissimilarity metric we propose is defined as the difference between the sum of all eigenvalues (which is equal to the trace of $\mathbf{K}$, which is equal to the constant G) and the sum of the L highest eigenvalues:

$$\mathcal{D}_{\mathrm{PCA}}(\boldsymbol{\mu}) = \sum_{j=1}^{G} \mathrm{K}_{jj}(\boldsymbol{\mu}) - \sum_{j=1}^{L} \lambda_j(\boldsymbol{\mu}) = G - \sum_{j=1}^{L} \lambda_j(\boldsymbol{\mu}), \tag{5.6}$$

where the dependence on $\boldsymbol{\mu}$ has been made explicit to clarify that $\mathbf{K}$ (and thus λ_j) is computed based on the deformed images $M_g(\boldsymbol{T}_g(\boldsymbol{x};\boldsymbol{\mu}_g))$. The constant $1 \leq L \leq G$ is a user-defined parameter. For different qMRI models, a different value of L must be chosen. A good initial guess is $L = \Gamma$, assuming that the non-linear Γ-dimensional subspace can be approximated by a Γ-dimensional hyperplane. [147] also assume that well-aligned images are linearly correlated but they use a sparse and low-rank decomposition to directly minimize the rank of $\mathbf{M}$ with a trade-off parameter for sparsity. Since the rank of $\mathbf{M}$ is equal to the number of non-zero eigenvalues of $\mathbf{K}$, the methods are related, but in our method the rank is not minimized but can be controlled by L, which is preferable in qMRI, since the rank depends on the qMRI model. The dimension of the subspace in qMRI may be less than Γ when the parameters in $\boldsymbol{\theta}$ are correlated, or higher than Γ due to the non-linearity of the acquisition models. That is why we propose a second dissimilarity metric that circumvents the need to choose L:

$$\mathcal{D}_{\mathrm{PCA2}}(\boldsymbol{\mu}) = \sum_{j=1}^{G} j\lambda_j(\boldsymbol{\mu}), \tag{5.7}$$

In $\mathcal{D}_{\mathrm{PCA2}}$ the eigenvalues with the highest values (and the lowest indices) have the lowest weight. Given that $\lambda_1 + \lambda_2 + \ldots + \lambda_j = G$, $\mathcal{D}_{\mathrm{PCA2}}$ promotes that as much variance as possible is explained by a few large eigenvectors.

5.2.3.1 *Metric derivatives*

Minimization of the dissimilarity metric with gradient based optimizers requires the derivative of the metric with respect to $\boldsymbol{\mu}$. To differentiate equations (5.6) and (5.7) with respect to $\boldsymbol{\mu}$ we use the approach of [148]:

$$\frac{\partial \mathcal{D}_{\mathrm{PCA}}}{\partial \boldsymbol{\mu}} = -\sum_{j=1}^{L} \frac{\partial \lambda_j}{\partial \boldsymbol{\mu}} = -\sum_{j=1}^{L} \boldsymbol{v}_j^{\mathrm{T}} \frac{\partial \mathbf{K}}{\partial \boldsymbol{\mu}} \boldsymbol{v}_j, \tag{5.8}$$

and

$$\frac{\partial \mathcal{D}_{\mathrm{PCA2}}}{\partial \boldsymbol{\mu}} = \sum_{j=1}^{G} j \frac{\partial \lambda_j}{\partial \boldsymbol{\mu}} = \sum_{j=1}^{G} j \boldsymbol{v}_j^{\mathrm{T}} \frac{\partial \mathbf{K}}{\partial \boldsymbol{\mu}} \boldsymbol{v}_j, \tag{5.9}$$

where $\boldsymbol{v}_j^T$ is the j^{th} eigenvector of $\mathbf{K}$. Similarly to [148] we ignore the unlikely repetition of eigenvalues. For repeated eigenvalues, linear combinations of eigenvectors are also an eigenvector, which invalidates the above. Using equation (5.5) and (5.8) we obtain the derivative of $\mathcal{D}_{\mathrm{PCA}}$ with respect to an element μ_p:

$$\begin{aligned} \frac{\partial \mathcal{D}_{\mathrm{PCA}}}{\partial \mu_p} = -\frac{2}{\mathrm{N}-1} \sum_{i=1}^{L} \Bigg[& \boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \left(\mathbf{M} - \overline{\mathbf{M}}\right)^{\mathrm{T}} \left(\frac{\partial \mathbf{M}}{\partial \mu_p} - \frac{\partial \overline{\mathbf{M}}}{\partial \mu_p} \right) \boldsymbol{\Sigma}^{-1} \boldsymbol{v}_i \\ & + \boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \left(\mathbf{M} - \overline{\mathbf{M}}\right)^{\mathrm{T}} \left(\mathbf{M} - \overline{\mathbf{M}}\right) \frac{\boldsymbol{\Sigma}^{-1}}{\partial \mu_p} \boldsymbol{v}_i \Bigg]. \end{aligned} \tag{5.10}$$

The above expression is obtained after simplifications and using the fact that

$$\boldsymbol{v}^{\mathrm{T}} \mathbf{B}^{\mathrm{T}} \mathbf{E} \boldsymbol{v} = \boldsymbol{v}^{\mathrm{T}} \mathbf{E}^{\mathrm{T}} \mathbf{B} \boldsymbol{v} \tag{5.11}$$

for two matrices $\mathbf{B}$ and $\mathbf{E}$ and vector $\boldsymbol{v}$. The derivative of $\boldsymbol{\Sigma}^{-1}$ with respect to μ_p is equal to

$$\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \mu_p} = -\frac{\boldsymbol{\Sigma}^{-3}}{N-1} \operatorname{diag} \left[\left(\mathbf{M} - \overline{\mathbf{M}}\right)^{T} \left(\frac{\partial \mathbf{M}}{\partial \mu_p} - \frac{\partial \overline{\mathbf{M}}}{\partial \mu_p} \right) \right] \tag{5.12}$$

and $\partial \mathbf{M} / \partial \mu_p$ and $\partial \overline{\mathbf{M}} / \partial \mu_p$ are computed using

$$\frac{\partial M_g \left(\boldsymbol{T}_g \left(\boldsymbol{x}; \boldsymbol{\mu}_g \right) \right)}{\partial \mu_p} = \left(\frac{\partial M_g}{\partial \boldsymbol{x}} \right)^{\mathrm{T}} \Bigg|_{\boldsymbol{T}_g \left(\boldsymbol{x}; \boldsymbol{\mu}_g \right)} \left(\frac{\partial \boldsymbol{T}_g}{\partial \mu_p} \right) \Bigg|_{\left(\boldsymbol{x}; \boldsymbol{\mu}_g \right)}. \tag{5.13}$$

The same steps are used to calculate $\partial\mathcal{D}_{\mathrm{PCA2}}/\partial\mu_p$. It was verified that the value of $\partial\overline{\mathbf{M}}/\partial\mu_p$, the derivative of the mean intensities, was negligibly small and could therefore be ignored in the actual implementation.

5.2.4 *Transformation models*

In our experiments (see Section 5.4) we use two different transformation models: a non-rigid transformation model in which deformations are modeled by cubic B-splines, proposed by [115], and an affine transformation model. Similar to [145] we used an exponential mapping of the affine matrix for parametrization.

5.2.5 *Optimization*

An adaptive stochastic gradient descent (ASGD) optimization method, proposed by [149], was used, which randomly samples positions in image space at each iteration in order to reduce computation time. Sampling was done off the voxel grid, which was shown to be necessary to reduce interpolation artifacts, as proposed by [119]. A multi-resolution strategy was used. In such a strategy the image is Gaussian-blurred with a certain standard deviation and at each level the standard deviation is decreased, such that large deformations are corrected first and finer deformations are corrected at higher levels. When a B-spline transformation was used, the control point spacing was also decreased at each resolution level. The number of random samples, the number of resolution levels, and the number of iterations per resolution level are user-defined parameters. Linear interpolation was used to interpolate the images during registration, to limit computation time. Cubic B-spline interpolation was used to produce the final motion-compensated images.

For the groupwise framework, the average deformation of the images was constrained to be zero by the approach of [117]: the average derivative of the dissimilarity metric with respect to its parameters $\boldsymbol{\mu}_g$ is subtracted from each derivative to $\boldsymbol{\mu}_g$, i.e. the derivatives are centered to zero mean.

$$\frac{\partial\mathcal{D}^*}{\partial\boldsymbol{\mu}_g} = \frac{\partial\mathcal{D}}{\partial\boldsymbol{\mu}_g} - \frac{1}{G}\sum_{g'}\frac{\partial\mathcal{D}}{\partial\boldsymbol{\mu}_{g'}}, \tag{5.14}$$

where $\partial\mathcal{D}^*/\partial\boldsymbol{\mu}_g$ is the zero-centered derivative.

5.2.6 *Reference dissimilarity metrics*

We compared the proposed method with three other methods: a pairwise method that uses a MI-based dissimilarity metric, and four groupwise methods. The pairwise MI dissimilarity used was proposed by [150]. The number of histogram bins used to calculate the probability functions is set to 32 in all experiments.

Wachinger *et al.* proposed accumulated pairwise estimates (APE) as a family of metrics [145]. One of the metrics they propose is the sum of squared normalized correlation coefficients. This can be written as the sum of the squared elements of the correlation matrix **K**. We implemented this metric as follows:

$$\mathcal{D}_{\text{APE}}(\boldsymbol{\mu}) = 1 - \frac{1}{G}\sqrt{\sum_i \sum_j K_{ij}(\boldsymbol{\mu})^2}. \tag{5.15}$$

Metz *et al.* proposed the sum of the variances, assuming no intensity changes between images [120]. The metric is defined as:

$$\mathcal{D}_{\text{VAR}}(\boldsymbol{\mu}) = \frac{1}{NG}\sum_{i=1}^{N}\sum_{g=1}^{G}\left[M_g\left(\boldsymbol{T}_g\left(\boldsymbol{x}_i;\boldsymbol{\mu}_g\right)\right) - \widetilde{M}\left(\boldsymbol{x}_i;\boldsymbol{\mu}\right)\right]^2, \tag{5.16}$$

where $\widetilde{M}\left(\boldsymbol{x}_i;\boldsymbol{\mu}\right) = \frac{1}{G}\sum_g M_g\left(\boldsymbol{T}_g\left(\boldsymbol{x}_i;\boldsymbol{\mu}_g\right)\right)$.

The groupwise mutual information based method proposed by [116] was implemented as:

$$\begin{aligned}\mathcal{D}_{\text{G-MI}}(\boldsymbol{\mu}) = -\sum_{g=1}^{G}\Bigg[&H\left(\widetilde{M}\left(\cdot;\boldsymbol{\mu}\right)\right) + H\left(M_g\left(\boldsymbol{T}_g\left(\cdot;\boldsymbol{\mu}_g\right)\right)\right)\\ &-H\left(\widetilde{M}\left(\cdot;\boldsymbol{\mu}\right), M_g\left(\boldsymbol{T}_g\left(\cdot;\boldsymbol{\mu}_g\right)\right)\right)\Bigg],\end{aligned} \tag{5.17}$$

where $H(\cdot)$ represents the marginal entropy function and $H(\cdot,\cdot)$ the joint entropy function. For this method the number of histogram bins was also set to 32 in all experiments.

Andersson *et al.* and Hallack *et al.* proposed to minimize the residual error of the model fit [136, 137]:

$$\mathcal{D}_{T_1}(\boldsymbol{\mu}) = \frac{1}{N}\sum_{i=1}^{N}\min_{\boldsymbol{\theta}}\sum_{g=1}^{G}\left(M_g\left(\boldsymbol{T}_g\left(\boldsymbol{x}_i;\boldsymbol{\mu}_g\right)\right) - m_g\left(\boldsymbol{\theta}\left(\boldsymbol{x}_i\right)\right)\right)^2 \tag{5.18}$$

We implemented this method for the model given by Eq. (5.2).

5.2.7 *qMRI fitting method*

Having registered the images M_g, the qMRI parameters $\boldsymbol{\theta}$ can be estimated by fitting the model m_g at each voxel. The qMRI model is fitted using a maximum-likelihood (ML) estimator that takes into account the Rician noise of MRI data. The estimation procedure is defined as [151]:

$$\hat{\boldsymbol{\theta}}, \hat{\varepsilon} = \underset{\boldsymbol{\theta},\varepsilon}{\arg\max}\left(\ln p\left(\mathbf{M}\middle| m(\boldsymbol{\theta}), \varepsilon\right)\right), \tag{5.19}$$

where p is the probability density function of the Rician distribution, $\mathbf{M}$ are the images, m is the qMRI model with parameters $\boldsymbol{\theta}$, and ε is the noise level, which is estimated per voxel. To get an accurate estimate of the noise level, the bias caused by the reduced degrees of freedom of the residual is corrected according to [151]:

$$\hat{\varepsilon}^{\text{modif}} = \hat{\varepsilon}\sqrt{G/(G-\Gamma)} \tag{5.20}$$

When $G - \Gamma$ is low, there are not enough degrees of freedom left in a voxel to estimate the noise level precisely. In these cases we therefore regularized the noise level with a log prior $S(\varepsilon)$ promoting a spatially smooth noise level field [151]:

$$\hat{\boldsymbol{\theta}}, \hat{\varepsilon} = \underset{\boldsymbol{\theta},\varepsilon}{\arg\max} \left(\ln p \left(\mathbf{M} \middle| m(\boldsymbol{\theta}), \varepsilon \right) - S(\varepsilon) \right) \tag{5.21}$$

The qMRI fitting method is publicly available at `fitMRI.bigr.nl`.

5.3 DATA

We evaluated the registration methods on a purely synthetic dataset, real CT lung datasets, five real qMRI datasets and five synthetic datasets derived from the real data. The subsections below describe all datasets.

5.3.1 *SYNTH-MODEL*

A purely synthetic image was generated to demonstrate the advantages of the proposed PCA-based methods when registering qMRI data. The generated images are linear combinations of five basis images, representing five independent 'tissue' properties in a qMRI experiment:

$$m_g(\boldsymbol{\theta}) = \boldsymbol{w}_g^{\mathrm{T}}\boldsymbol{\theta} \tag{5.22}$$

where $\boldsymbol{\theta}$ contains five 'tissue' properties. Each property is spatially correlated and contains large and small structures. An example 'tissue' property map is shown in Figure 5.4a. The weights are optimized such that the sum of the pairwise mutual information of the images is minimized. The intrinsic dimension of the synthetic data can however be perfectly resolved by a principal component analysis. We expect that especially the proposed PCA-based methods can successfully register the images. An example of an image M_g is shown in Figure 5.4b. The code to generate the SYNTH-MODEL is included as supplementary material.

5.3.2 *Synthetic qMRI*

To evaluate how the methods perform with different qMRI models, in a setting with known ground truth, we created synthetic data based on real qMRI data. To save computation time,

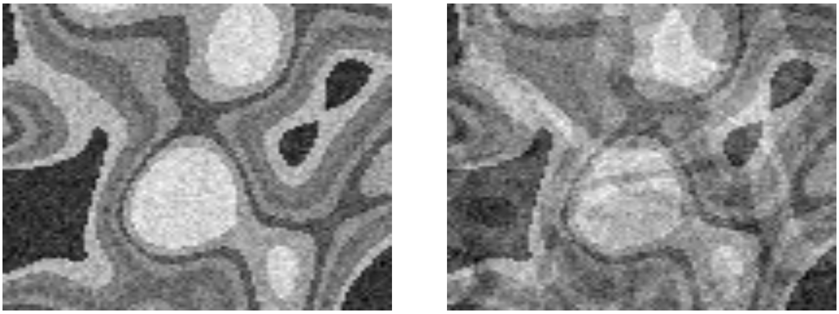

(a) An element of $\boldsymbol{\theta}$ (b) An example image M_g

Figure 5.4: (a) One of the five 'tissue' properties, i.e. an element of $\boldsymbol{\theta}$, and (b) one of the G weighted images.

we extracted a 2D slice from a single subject, for each of the qMRI applications studied in this paper (see Sections 5.3.4 - 5.3.8). These slices were fitted and the obtained $\boldsymbol{\theta}$ and the acquisition parameters were used to simulate the contrast with the qMRI model that belongs to the data. Rician noise was added to the synthetic qMRI data such that the signal-to-noise (SNR) ratio was equal to 10, where the SNR is defined as the mean signal intensity over all images divided by the σ-parameter of the Rician distribution. The pixel spacings in the first two dimensions and G for the synthetic datasets are equal to their values in the corresponding real datasets. All synthetic qMRI datasets are written with a prefix 's-'.

5.3.3 *CT-LUNG*

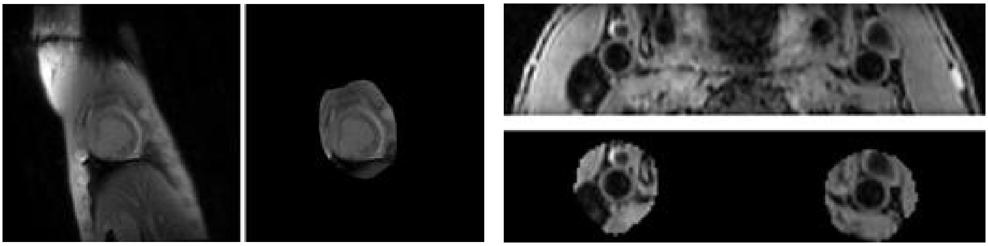
Ten 4D CT lung datasets from the DIR-LAB database [152] were used to demonstrate the results of the methods on datasets with only slight intensity changes. Note that this can be seen as a special case of quantitative imaging where a 1-parameter model ($\Gamma = 1$) describes the intensity with a constant:

$$m_g(\boldsymbol{\theta}) = c, \tag{5.23}$$

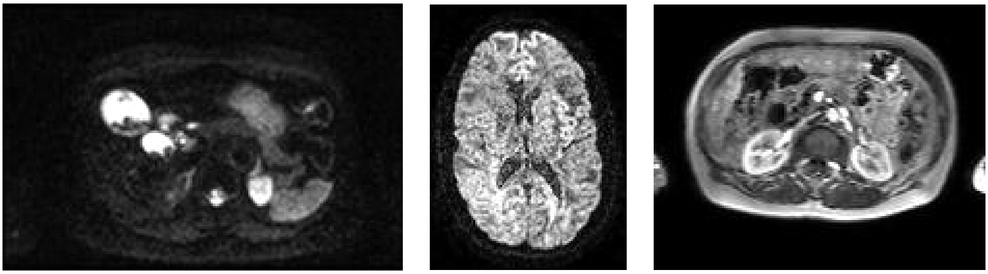
with $\boldsymbol{\theta} = c$, wich is trivially estimated by computing the mean over M_g. The voxel size was approximately $1.1\times1.1\times2.5$ mm^3. The size of the first five images was around $256\times256\times100$ and of the last five images around $512\times512\times128$. For all datasets $G = 10$.

5.3.4 *T1MOLLI-HEART*

Quantification of T_1 relaxation is important for the characterization of myocardial tissue, which is useful to assess both ischemic and non-ischemic heart muscle diseases [134]. A popular cardiac T_1 mapping method uses the modified Look-Locker inversion recovery (MOLLI) sequence, described in [146]. The qMRI model for MOLLI is given by Eq. (5.2). Nine T1MOLLI-HEART datasets from porcine hearts with a transmural myocardial infarction of the lateral wall were acquired using single-slice acquisition. For each subject $G = 11$



(a) T1MOLLI-HEART and registration mask (b) T1VFA-CAROTID and registration masks



(c) ADC-ABDOMEN (d) DTI-BRAIN (e) DCE-ABDOMEN

Figure 5.5: Examples of real datasets and, if used, corresponding registration masks.

two-dimensional images of size 512×512 were acquired. The pixel size was around 0.7×0.7 mm^2. Inversion times TI_g ranged from 82 to 3866 ms. Figure 5.5a shows an example image.

5.3.5 *T1VFA-CAROTID*

Quantifying the relaxation parameters T_1 and T_2 can help to assess the composition of atherosclerotic plaque in the carotid arteries [153]. Carotid plaque composition has been shown to be associated with the occurrence of cerebrovascular events [154]. A quantitative carotid plaque imaging method was proposed based on a 3D improved motion-sensitized driven-equilibrium prepared black-blood turbo field echo sequence [153]. The qMRI model of this variable flip-angle approach is given by:

$$m_g(\boldsymbol{\theta}) = \left| A \sin(\alpha_g) \frac{1 - e^{-TR/T_1}}{1 - \cos(\alpha_g) e^{-TR/T_1}} e^{-TE_g/T_2} \right|, \tag{5.24}$$

with $\boldsymbol{\theta} = (A, T_1, T_2)$, α_g the flip-angle and TE_g the T_2 preparation time per image M_g and TR the repetition time [153]. Values for the acquisition parameters were $\alpha_g \in [6, 4, 15, 15, 15]$ degrees, $TE_g \in [11.5, 11.5, 11.5, 26, 45]$ ms and $TR = 10$ ms. Eight T1VFA-CAROTID datasets were acquired. For each subject, $G = 5$ three-dimensional images of size 224×223×36 with voxel size 0.7×0.7×0.7 mm^3 were acquired. One of these five images was an anatomical reference scan. The other four scans were used for the combined T_1 and T_2 fitting. Although this method gave accurate T_2 mapping results when analyzed in a ROI [153], the acquisitions were not optimal yet for true voxelwise T_2 fitting, possibly due to the rather low maximum TE_g. In our experiments, it proved difficult to perform accurate voxelwise T_2 fits on the current data, regardless of the registration strategy. Therefore, we only report evaluations of T1 values. More acquisition details can be found in [153] and an example image is shown in Figure 5.5b. Please note that although the T1MOLLI-HEART and the T1VFA-CAROTID experiments both result in a T_1 estimate, a different model m_g is underlying the acquisition.

5.3.6 *ADC-ABDOMEN*

The ADC provides quantitative information related to the diffusion of water molecules in biological tissues. Pathophysiological processes such as cancer are known to have an impact on cell density, which translates into different diffusion properties and therefore the ADC is an interesting biomarker to assess these diseases [133]. In this work, we evaluate ADC mapping in the abdomen. The ADC acquisition model is the following:

$$m_g(\boldsymbol{\theta}) = B_0 e^{-b_g \boldsymbol{u}_g^{\mathrm{T}} \mathbf{D} \boldsymbol{u}_g}, \tag{5.25}$$

with $\boldsymbol{\theta} = (B_0, D_{11}, D_{22}, D_{33})$ [133]. The vector in the direction of the applied gradient is given by $\boldsymbol{u}_g$, $\mathbf{D}$ is a 3×3 symmetric diffusion tensor and b_g is the so-called b-value. The ADC is given by

$$\mathrm{ADC} = \mathrm{trace}(\mathbf{D})/3, \tag{5.26}$$

hence only the diagonal elements of $\mathbf{D}$ are required. For acquiring the ADC data, an interleaved multi-slice diffusion weighted acquisition sequence was used. Five ADC-ABDOMEN datasets were acquired. For each subject $G = 19$ images of size 256×224×40 were acquired with a voxel size of 1.48×1.48×5 mm^3 with b-values of 0, 100, 150, 200, 300, 500 and 900 $\mathrm{s/mm}^2$. Diffusion weighting was applied in three orthogonal directions aligned with the read, phase and slice directions. More details about the acquisition settings of these datasets can be found in [133]. Figure 5.5c shows an example of a diffusion weighted image.

5.3.7 *DTI-BRAIN*

Measuring the diffusion tensor of water in tissues can provide parameters that help to characterize tissue composition, the physical properties of tissue constituents, tissue microstructure and its architectural organization. It is often used to map the white matter tissue structure [155–157]. The qMRI model belonging to DT imaging equals:

$$m_g(\boldsymbol{\theta}) = B_0 e^{-b \boldsymbol{u}_g^{\mathrm{T}} \mathbf{D} \boldsymbol{u}_g} \tag{5.27}$$

Table 5.1: Acquisition details of the DTI-BRAIN data.

Dataset	# b=0 $\mathrm{s/mm^2}$	# DWIs	G	b-value $[\mathrm{s/mm^2}]$	voxel size $[\mathrm{mm^3}]$
1	10	60	70	700	$2.0\times2.0\times2.0$
2	1	60	61	1200	$1.75\times1.75\times2.0$
3	1	32	33	800	$1.75\times1.75\times2.0$
4	1	32	33	800	$1.75\times1.75\times2.0$
5	1	45	46	1200	$1.72\times1.72\times2.0$

with $\boldsymbol{\theta} = (B_0, D_{11}, D_{12}, D_{13}, D_{22}, D_{23}D_{33})$ for the DT mapping. The vector in the direction of the applied gradient is given by $\boldsymbol{u}_g$, $\mathbf{D}$ is a 3×3 symmetric diffusion tensor and b is the so-called b-value. In this paper we calculate the mean diffusivity (MD), which is given by:

$$\mathrm{MD} = \mathrm{trace}(\mathbf{D})/3. \tag{5.28}$$

Five DTI-BRAIN datasets were acquired with a diffusion weighted EPI sequence. Details about the five datasets are provided in Table 5.1 and details of the acquisition settings can be found in, respectively, [158–162]. Figure 5.5d shows an example of a diffusion weighted image.

5.3.8 *DCE-ABDOMEN*

DCE imaging is an established method for assessing microvascular changes associated with disease in tissues. Examples of tissue diseases where DCE imaging is used are cancer, inflammatory conditions, cerebral ischemia, and cardiac ischemia [163]. In DCE imaging the contrast in the images varies due to the injected contrast agent. The acquisition parameters T_R and α are not varied. In this work, we study DCE imaging in the abdomen. The qMRI model belonging to the DCE acquisition equals:

$$m_g(\boldsymbol{\theta}) = S_0 \frac{(1-\cos\alpha)\, e^{-T_R/T_{10}}}{1 - e^{-T_R/\left(T_{10} + C(\boldsymbol{\theta}, t_g) T_1\right)}} \tag{5.29}$$

with $C(\boldsymbol{\theta}, t_g)$ the contrast agent concentration in the tissue, described by the extended Tofts model [164]:

$$C(\boldsymbol{\theta}, t_g) = C_p(t_g - \Delta) v_p + C_p(t_g - \Delta) * K^{\mathrm{trans}} e^{-\frac{K^{\mathrm{trans}}}{v_e} t_g} \tag{5.30}$$

where $\boldsymbol{\theta} = \left(K^{\mathrm{trans}}, v_e, v_p, \Delta\right)$, $C_p(t)$ is the blood plasma concentration described by the general arterial input function by [163], and $*$ denotes the convolution operator. The time delay Δ was estimated separately, prior to fitting θ. K^{trans} is a measure of capillary permeability, v_p is the blood volume fraction, v_e is the volume fraction of extracellular, extravascular space and S_0 is the signal without contrast agent. Known variables were the acquisition time t_g, the flip-angle α, the repetition time T_R, the longitudinal relaxation parameter without contrast T_{10}, and the longitudinal relaxation parameter of the contrast agent T_1. The values for the known acquisition parameters were $\alpha = 20$ degrees, $T_R = 3.19$ ms, $T_{10} = 725$ ms, and $T_1 = 200$ ms. Five DCE-ABDOMEN datasets were acquired with a spoiled gradient echo sequence. Per subject $G = 160$ images of size $160 \times 160 \times 30$ were acquired with a voxel size of $2.5 \times 2.5 \times 2.5$ mm^3. More acquisition details can be found in [165]. An example image is shown in Figure 5.5e.

5.4 EXPERIMENTS

The proposed registration methods were implemented in the publicly available registration package Elastix [119]. All datasets were registered with six intensity-based dissimilarity metrics and, in addition, both the real and synthetic T1MOLLI-HEART datasets were registered with a model-based metric.

1. $\mathcal{D}_{\mathrm{MI}}$ (pairwise)
2. $\mathcal{D}_{\mathrm{VAR}}$ (groupwise)
3. $\mathcal{D}_{\mathrm{APE}}$ (groupwise)
4. $\mathcal{D}_{\mathrm{G\text{-}MI}}$ (groupwise)
5. $\mathcal{D}_{T_1}$ (groupwise) (T1MOLLI model)
6. $\mathcal{D}_{\mathrm{PCA}}$ (groupwise)
7. $\mathcal{D}_{\mathrm{PCA2}}$ (groupwise)

To account for deformations caused by heart-pulsations and breathing we used a B-spline transformation model for the CT-LUNG data, T1MOLLI-HEART, T1VFA-CAROTID, ADC-ABDOMEN and DCE-ABDOMEN experiments. For ADC-ABDOMEN and DCE-ABDOMEN datasets, results are also reported for an affine transformation. To account for deformations caused by head motion and eddy current distortions we used an affine transformation model for application DTI-BRAIN. Two types of experiments were done:

1. The synthetic images were deformed by applying a known transformation and registered with all methods;
2. The real CT-LUNG and qMRI data were registered with all methods.

The parameters of the transformation applied to the SYNTH-MODEL data were drawn from a normal distribution, such that the initial deformation was approximately six pixels for the affine transformation and three pixels for the B-spline transformation. The parameters of the transformations applied to the synthetic qMRI datasets were drawn from normal distributions, such that the initial deformation was approximately one to two pixels. For the cases with non-rigid deformations, the control point spacing $\boldsymbol{\sigma}$ for the initial and estimated B-spline transformation was equal in all dimensions: $\boldsymbol{\sigma} = [\nu, \nu]$, with the following values for ν:

- SYNTH-MODEL: $\nu = 16$ mm
- s-T1MOLLI: $\nu = 64$ mm
- s-T1VFA: $\nu = 16$ mm

Table 5.2: Summary of the settings used in the image registration experiments, for all datasets.

qMRI application	G	Γ	L	ν [mm]
SYNTH-MODEL	20	5	5	16
CT-LUNG	10	1	1	6, 13 and 20
T1MOLLI-HEART	11	3	3	32, 64 and 128
T1VFA-CAROTID	5	3	1	8, 16 and 32
ADC-ABDOMEN	19	4	4	32, 64 and 128
DTI-BRAIN	see Table 5.1	7	7	-
DCE-ABDOMEN	160	4	4	32, 64 and 128

Table 5.2: Summary of the settings used in the image registration experiments, for all datasets (continued).

qMRI application	reference image	registration mask	evaluation ROI
SYNTH-MODEL	first image of the series	-	-
CT-LUNG	first image of the series	lungs	red
T1MOLLI-HEART	slice 1, 4, 7 and 11	see Figure 5.5a	myocardium
T1VFA-CAROTID	anatomical scan	see Figure 5.5b	carotid artery wall
ADC-ABDOMEN	$b = 0\ \mathrm{s/mm^2}$ image	-	spleen
DTI-BRAIN	$b = 0\ \mathrm{s/mm^2}$ image	-	parenchyma
DCE-ABDOMEN	non-contrast weighted image	-	pancreas

- s-ADC: $\nu = 64$ mm
- s-DCE: $\nu = 64$ mm

The values of ν were chosen such that the simulated deformations were realistic. For the T1MOLLI-HEART, T1VFA-CAROTID experiments and their synthetic datasets we used registration masks, loosely drawn around the heart and the carotid arteries, to reduce the influence of surrounding organs. Examples are shown in Figure 5.5a and Figure 5.5b. For T1VFA-CAROTID, registration was performed separately for the left and right carotid. Lung masks were used in the CT-LUNG experiment, similar to the study performed in [120]. For the cases with non-rigid deformations, different values for the control point spacing σ for the B-spline transformation were used. The values for ν are provided in Table 5.2. Before further processing of the ADC-ABDOMEN data, within-image motion artifacts due to interleaved acquisition were corrected with the methods of [133].

For all registrations we used the following settings:

- two resolutions
- 1000 iterations per resolution
- 2048 random coordinate samples per resolution

In the pairwise registration framework, a reference image must be chosen. The images that are used as reference for each application are shown in Table 5.2. For the T1MOLLI-HEART data there is no obvious reference image, so in this case we used multiple images of each dataset as a reference image, which enables us to evaluate the effect of the reference image on the results.

The $\mathcal{D}_{\mathrm{PCA}}$ method requires the user to set the parameter L. This parameter is shown in Table 5.2 for all qMRI applications.

5.4.1 *Evaluation measures*

As the synthetic data are aligned by design, they can be used as a ground truth to evaluate registration accuracy. A known transformation $\boldsymbol{T}_g(\boldsymbol{x};\boldsymbol{\mu}_{\text{init},g})$ was applied to the data and corrected using the registration methods. The accuracy of the registration method can be evaluated by calculating the resulting residual deformation field $\boldsymbol{d}(\boldsymbol{x})$:

$$\boldsymbol{d}_g(\boldsymbol{x}) = \boldsymbol{T}_g\left(\boldsymbol{T}_g(\boldsymbol{x};\boldsymbol{\mu}_{\text{init},g});\hat{\boldsymbol{\mu}}_g\right) - \boldsymbol{x}, \tag{5.31}$$

with $\boldsymbol{T}_g(\boldsymbol{x};\hat{\boldsymbol{\mu}}_g)$ the transformation estimated by the registration method. The lower $\boldsymbol{d}(\boldsymbol{x})$ for all $\boldsymbol{x}$, the more accurate the registration method. However, the constraint, equation (5.14), was not applied to the initial transformation, so we subtract the mean of the deformation field: $\boldsymbol{d}^*_g(\boldsymbol{x}) = \boldsymbol{d}_g(\boldsymbol{x}) - \frac{1}{G}\sum_{g'=1}^{G}\boldsymbol{d}_{g'}(\boldsymbol{x})$. We report the mean and standard deviation of $\|\boldsymbol{d}^*_g(\boldsymbol{x})\|$ over all $\boldsymbol{x}$ and g. Secondly, we measured how the uncertainty of the fitted parameter of interest of the qMRI model was altered by the registration method (see Section 5.4.1.4).

For the real data, no ground truth was available. We therefore used, besides a visual inspection, the following evaluation measures, which can be summarized as:

1. how well manually defined landmarks correspond in the images (CT-LUNG, T1VFA-CAROTID and DCE-ABDOMEN);
2. how well anatomical regions of interest (ROI) overlap in all images (T1MOLLI-HEART and ADC-ABDOMEN);
3. the smoothness of the deformation field, evaluated within an ROI;
4. the effect of the registration method on the uncertainty of the fitted parameter of the qMRI model, evaluated within an ROI.

To facilitate a fair comparison of all registration methods for the different evaluation measures, the images resulting from the groupwise registration metrics $\mathcal{D}_{\text{VAR}}$, $\mathcal{D}_{\text{APE}}$, $\mathcal{D}_{\text{G-MI}}$, $\mathcal{D}_{T_1}$, $\mathcal{D}_{\text{PCA}}$ and $\mathcal{D}_{\text{PCA2}}$ were transformed to the space of the reference image $M_R(\boldsymbol{x})$ as used by $\mathcal{D}_{\text{MI}}$. To that purpose, the inverse transformation $\boldsymbol{T}_R^{-1}(\boldsymbol{x};\boldsymbol{\mu}_R)$ was computed. The manually outlined structures that are used as ROIs for all qMRI applications are mentioned in Table 5.2. For the DTI-BRAIN data, neither landmarks nor structures could be reliably identified on each of the diffusion weighted images, so for this application no overlap or point correspondence was calculated. In the following sections the evaluation methods are explained in more detail.

5.4.1.1 *Point correspondence*

The correspondence between landmarks was evaluated with the mean target registration error (mTRE):

$$\text{mTRE}_l = \frac{1}{G-1}\sum_{g=1, g\neq R}^{G} \|\boldsymbol{p}_g - \boldsymbol{T}_g\left(\boldsymbol{T}_R^{-1}\left(\boldsymbol{p}_R;\boldsymbol{\mu}_R\right);\boldsymbol{\mu}_g\right)\|, \tag{5.32}$$

where G is the number of images in which the landmark l is annotated, $\boldsymbol{p}_R$ the landmark in the reference image, $\boldsymbol{p}_g$ the landmark annotated in image g, and $\boldsymbol{T}_g\left(\boldsymbol{T}_R^{-1}\left(\boldsymbol{p}_R;\boldsymbol{\mu}_R\right);\boldsymbol{\mu}_g\right)$ the transformation of point $\boldsymbol{p}_R$ from the reference space to the space of image g. The transformations from reference space were used to facilitate the comparison with the non-groupwise $\mathcal{D}_{\text{MI}}$ approach.

For the CT-LUNG data anatomical landmarks are publicly available. For the DCE-ABDOMEN and T1VFA-CAROTID data the landmarks were manually annotated. Multiple landmarks were annotated and therefore we report the mean over the mTRE_l for all l. For CT-LUNG, 75 landmarks were annotated in the first five time points and 300 landmarks were annotated in the first and fifth time point. The mTRE is calculated for the 75 landmarks in the first five time points and for the 300 landmarks in inhale and exhale phase. The mean of these two mTREs is reported. In the T1VFA-CAROTID data four landmarks per dataset were determined, two per registration mask, and they were defined in all five images. In the DCE-ABDOMEN five landmarks per dataset were determined and they were defined in 50 of the 160 images. Due to sliding motion, low resolution and high contrast differences the five landmarks were annotated in 50 different images, which results in an unequal number of landmarks per image of the DCE-ABDOMEN data. Note that instead of an mTRE_l, i.e. an mTRE per landmark l, we could have computed an mTRE_g, i.e. an mTRE per image g. For datasets CT-LUNG and T1VFA-CAROTID the mean over mTRE_l for all l is equal to the mean over mTRE_g for all g, but for the DCE-ABDOMEN data this is not the case due to the unequal number of landmarks per image.

5.4.1.2 *Overlap*

We extended the Dice coefficient to measure the overlap between more than two segmentations as:

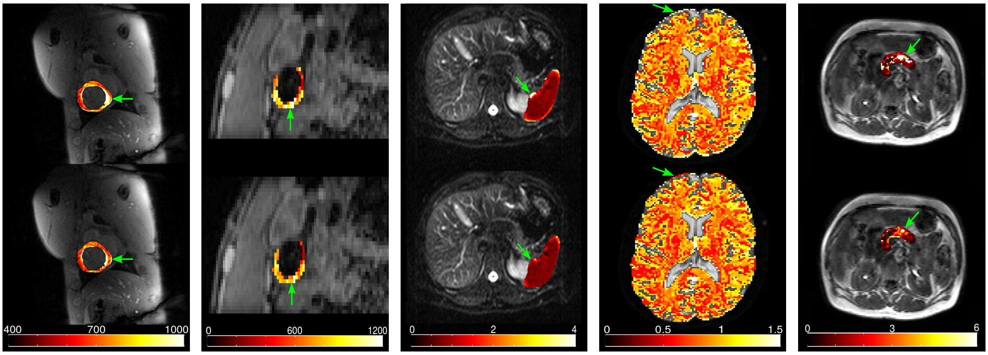
$$\text{Dice}_G = G\frac{|S_1 \cap S_2 \ldots \cap S_G|}{|S_1| + |S_2| \ldots + |S_G|}, \tag{5.33}$$

where S_g is the segmentation in the g^{th} image and G the number of segmentations. Note that Dice_G is sensitive to the misregistration of a single image, which is important in qMRI since a single misregistered image can severely reduce the quality of the qMRI model fit.

In the T1MOLLI-HEART experiment we measured the overlap of the manually outlined myocardium, outlined on six to nine images. Not all images were outlined because manual segmentation on some images proved to be difficult due to low contrast between the myocardium and surrounding tissues. For the ADC-ABDOMEN we measured the overlap of the manually outlined spleen in b=0 s/mm^2 image and in diffusion weighted images b=100 s/mm^2 in phase direction, b=100 s/mm^2 in slice direction, b=150 s/mm^2 in slice direction, b=200 s/mm^2 in read direction, b=300 s/mm^2 in read direction, b=500 s/mm^2 in phase direction and b=900 s/mm^2 in phase direction.

5.4.1.3 *Smoothness of the transformation*

Extreme and non-smooth deformations are unexpected for the experiments we conducted. Hence, the smoothness of the deformation field can be used to identify such undesirable trans-



(a) T_1-map T1MOLLI-HEART [ms]
(b) T_1-map T1VFA-CAROTID [ms]
(c) ADC-map ADC-ABDOMEN [$\mu m^2/ms$]
(d) MD-map DTI-BRAIN [$\mu m^2/ms$]
(e) K^{trans}-map DCE-ABDOMEN [min^{-1}]

Figure 5.6: Tissue maps, generated before (top) and after registration (bottom) with $\mathcal{D}_{\text{PCA2}}$. The green arrows indicate the differences in the tissue maps before and after registration.

formations. It is obtained by calculating the mean of the standard deviation of the determinant of $\partial \boldsymbol{T}_g/\partial \boldsymbol{x}$ over all $\boldsymbol{x}$ for all images:

$$\text{STD}_{|\partial \boldsymbol{T}_g/\partial \boldsymbol{x}|} = \frac{1}{G}\sum_{g=1}^{G} \text{STD}\left(|\partial \boldsymbol{T}_g/\partial \boldsymbol{x}|\right), \tag{5.34}$$

where $\text{STD}\left(|\partial \boldsymbol{T}_g/\partial \boldsymbol{x}|\right)$ is the standard deviation of $|\partial \boldsymbol{T}_g/\partial \boldsymbol{x}|$ over all $\boldsymbol{x} \in \text{ROI}$.

Table 5.3: Results of the SYNTH-MODEL experiment. Mean ± standard deviation of $\|\boldsymbol{d}^*_g(\boldsymbol{x})\|$ [mm].

	Affine	B-spline
(-)	6.2 ±3.7	3.0 ±1.6
$\mathcal{D}_{\text{MI}}$	7.7 ±8.9	9.3 ±7.1
$\mathcal{D}_{\text{VAR}}$	39.7 ±30.7	34.0 ±21.5
$\mathcal{D}_{\text{APE}}$	8.4 ±4.9	21.5 ±13.3
$\mathcal{D}_{\text{G-MI}}$	20.1 ±14.0	16.9 ±12.2
$\mathcal{D}_{\text{PCA}}$	0.1 ±0.1	0.6 ±0.6
$\mathcal{D}_{\text{PCA2}}$	0.1 ±0.3	1.5 ±1.9

Table 5.4: Results of the synthetic qMRI experiments. Mean ± standard deviation of $\|\boldsymbol{d}^*_g(\boldsymbol{x})\|$ [mm].

	s-T1MOLLI	s-T1VFA	s-ADC	s-DTI	s-DCE
(-)	1.55±1.10	1.02±0.79	0.98±0.63	1.32±0.91	1.18±0.64
$\mathcal{D}_{\text{MI}}$	1.83±1.94	0.43±0.38	1.80±2.03	1.46±1.21	3.03±5.34
$\mathcal{D}_{\text{VAR}}$	10.01±10.02	1.64±1.44	5.03±4.85	9.88±11.54	1.67±5.93
$\mathcal{D}_{\text{APE}}$	2.71±2.88	0.14±0.12	1.14±0.85	1.77±2.06	0.96±0.94
$\mathcal{D}_{\text{G-MI}}$	1.35±1.54	0.41±0.34	1.26±1.01	1.23±1.45	0.93±1.02
$\mathcal{D}_{T_1}$	0.23±0.21	-	-	-	-
$\mathcal{D}_{\text{PCA}}$	0.30±0.28	0.15±0.13	0.76±0.57	0.16±0.16	1.16±3.19
$\mathcal{D}_{\text{PCA2}}$	0.79±0.93	0.12±0.11	0.73±0.57	0.23±0.28	0.75±1.26

Table 5.5: 90%$\sqrt{\text{CRLB}}$ of T_1 [ms], ADC [μm^2/ms], MD [μm^2/ms] and K^{trans} [min^{-1}] in ROI of synthetic datasets.

	s-T1MOLLI	s-T1VFA	s-ADC	s-DTI	s-DCE
(-)	148	>1000	0.243	0.063	1.79
$\mathcal{D}_{\text{MI}}$	50	542	0.046	0.023	4.05
$\mathcal{D}_{\text{VAR}}$	121	>1000	0.043	0.021	2.76
$\mathcal{D}_{\text{APE}}$	75	206	0.054	0.018	0.92
$\mathcal{D}_{\text{G-MI}}$	47	390	0.093	0.007	0.39
$\mathcal{D}_{T_1}$	25	-	-	-	-
$\mathcal{D}_{\text{PCA}}$	27	203	0.039	0.008	1.31
$\mathcal{D}_{\text{PCA2}}$	34	222	0.028	0.008	0.32

5.4.1.4 *Uncertainty estimation of the qMRI fit*

The uncertainty of the estimated qMRI model parameters $\boldsymbol{\theta}$ was quantified by the Cramér–Rao lower bound (CRLB), which provides a lower bound for the variance of the ML estimated parameters [151, 166–168]. The CRLB of a function $f(\boldsymbol{\theta})$ is given by:

$$\text{CRLB}_f(\boldsymbol{\theta}, \varepsilon) = \left(\frac{\partial f}{\partial \boldsymbol{\theta}}\right)^{\text{T}} \mathbf{I}^{-1}(\boldsymbol{\theta}, \varepsilon) \left(\frac{\partial f}{\partial \boldsymbol{\theta}}\right) \tag{5.35}$$

where $\mathbf{I}(\boldsymbol{\theta}, \varepsilon)$ is the Fisher information matrix. For Rician distributed data, no closed-form expression for the Fisher information exists. To evaluate the CRLB, we used the method by [169], which approximates the integrals present in the evaluation of the Fisher information to double precision accuracy. To use the CRLB as indicator of misalignment we adopt the measure proposed by [6], which is the 90% percentile of the square root of the CRLB (90%$\sqrt{\text{CRLB}}$) over an ROI. This measure identifies misalignment, because misalignment may result in biologically implausible values of the estimated parameters, especially at tissue boundaries. Additionally,

Table 5.6: Results of the CT-LUNG experiment. Mean and standard deviation over all subjects for all evaluation measures and control point spacings 6 mm, 13 mm, and 20 mm.

mTRE [mm]	***6***	***13***	***20***
(-)	6.72±2.51	6.72±2.51	6.72±2.51
$\mathcal{D}_{\text{MI}}$	1.78±0.40	1.43±0.23	1.45±0.21
$\mathcal{D}_{\text{VAR}}$	1.59±0.59	1.42±0.39	1.47±0.37
$\mathcal{D}_{\text{APE}}$	1.47±0.60	1.40±0.37	1.45±0.34
$\mathcal{D}_{\text{G-MI}}$	1.37±0.38	1.28±0.21	1.31±0.21
$\mathcal{D}_{\text{PCA}}$	1.47±0.60	1.40±0.37	1.45±0.34
$\mathcal{D}_{\text{PCA2}}$	1.72±0.78	1.56±0.55	1.59±0.49
STD$_{\lvert\partial T_g/\partial x\rvert}$ [%]	***6***	***13***	***20***
(-)	0 ±0	0 ±0	0 ±0
$\mathcal{D}_{\text{MI}}$	28 ±8	15 ±4	11 ±3
$\mathcal{D}_{\text{VAR}}$	14 ±4	9 ±2	7 ±2
$\mathcal{D}_{\text{APE}}$	12 ±3	8 ±2	7 ±2
$\mathcal{D}_{\text{G-MI}}$	13 ±2	9 ±2	7 ±2
$\mathcal{D}_{\text{PCA}}$	12 ±3	8 ±2	7 ±2
$\mathcal{D}_{\text{PCA2}}$	10 ±2	7 ±2	6 ±1

Table 5.7: Results of the T1MOLLI-HEART experiment. Mean and standard deviation over all subjects for all evaluation measures and control point spacings 32 mm, 64 mm, and 128 mm and for reference images 1, 4, 7 and 11.

Dice_G [%]	**32**				**64**				**128**			
	1	***4***	***7***	***11***	***1***	***4***	***7***	***11***	***1***	***4***	***7***	***11***
(-)	48 ±8	48 ±8	48 ±8	48 ±8	48 ±8	48 ±8	48 ±8	48 ±8	48 ±8	48 ±8	48 ±8	48 ±8
$\mathcal{D}_{\text{MI}}$	26 ±11	22 ±10	22 ±15	39 ±14	37 ±11	34 ±16	19 ±15	52 ±7	44 ±14	40 ±13	24 ±17	51 ±7
$\mathcal{D}_{\text{VAR}}$	2 ±3	4 ±6	5 ±6	4 ±5	16 ±12	16 ±12	16 ±12	16 ±12	35 ±17	35 ±17	35 ±17	35 ±17
$\mathcal{D}_{\text{APE}}$	20 ±10	21 ±10	22 ±11	21 ±10	48 ±11	48 ±10	48 ±11	48 ±11	48 ±9	49 ±9	48 ±9	48 ±8
$\mathcal{D}_{\text{G-MI}}$	36 ±15	36 ±15	36 ±15	36 ±15	41 ±7	42 ±8	42 ±7	41 ±8	43 ±10	43 ±10	43 ±10	43 ±10
$\mathcal{D}_{\text{T1}}$	44 ±15	45 ±15	45 ±15	44 ±14	50 ±13	51 ±13	50 ±13	50 ±13	50 ±10	51 ±9	51 ±10	51 ±10
$\mathcal{D}_{\text{PCA}}$	54 ±9	54 ±9	54 ±9	54 ±9	53 ±7	53 ±7	53 ±7	53 ±8	51 ±9	50 ±9	51 ±9	51 ±9
$\mathcal{D}_{\text{PCA2}}$	52 ±13	52 ±13	52 ±13	52 ±13	52 ±11	52 ±10	52 ±11	53 ±10	52 ±7	52 ±8	52 ±8	52 ±8
$\text{STD}_{\lvert\partial T_g/\partial x\rvert}$ [%]	**32**				**64**				**128**			
	1	***4***	***7***	***11***	***1***	***4***	***7***	***11***	***1***	***4***	***7***	***11***
(-)	0 ±0	0 ±0	0 ±0	0 ±0	0 ±0	0 ±0	0 ±0	0 ±0	0 ±0	0 ±0	0 ±0	0 ±0
$\mathcal{D}_{\text{MI}}$	24 ±8	28 ±8	29 ±9	10 ±4	7 ±2	7 ±2	8 ±3	3 ±1	1 ±0	1 ±0	1 ±0	1 ±0
$\mathcal{D}_{\text{VAR}}$	82 ±31	60 ±20	57 ±21	78 ±27	16 ±14	9 ±5	10 ±6	11 ±6	1 ±1	1 ±0	1 ±0	1 ±0
$\mathcal{D}_{\text{APE}}$	17 ±7	17 ±7	18 ±8	13 ±6	3 ±2	3 ±3	4 ±4	2 ±2	0 ±0	0 ±0	0 ±0	0 ±0
$\mathcal{D}_{\text{G-MI}}$	10 ±3	12 ±11	15 ±13	7 ±4	2 ±1	2 ±2	2 ±2	1 ±1	0 ±0	0 ±0	0 ±0	0 ±0
$\mathcal{D}_{\text{T1}}$	15 ±9	11 ±5	14 ±8	8 ±3	3 ±2	3 ±2	3 ±2	2 ±1	0 ±0	0 ±0	0 ±0	0 ±0
$\mathcal{D}_{\text{PCA}}$	6 ±4	4 ±2	4 ±2	4 ±1	2 ±1	1 ±1	2 ±1	1 ±1	0 ±0	0 ±0	0 ±0	0 ±0
$\mathcal{D}_{\text{PCA2}}$	6 ±2	5 ±2	6 ±2	4 ±1	1 ±1	1 ±1	2 ±1	1 ±1	0 ±0	0 ±0	0 ±0	0 ±0
$90\%\sqrt{\text{CRLB}}\ T_1$ [ms]	**32**				**64**				**128**			
	1	***4***	***7***	***11***	***1***	***4***	***7***	***11***	***1***	***4***	***7***	***11***
(-)	92 ±19	92 ±19	92 ±19	92 ±19	92 ±19	92 ±19	92 ±19	92 ±19	92 ±19	92 ±19	92 ±19	92 ±19
$\mathcal{D}_{\text{MI}}$	119 ±26	157 ±61	467 ±610	93 ±19	97 ±16	103 ±18	146 ±52	83 ±13	91 ±14	95 ±16	121 ±29	81 ±12
$\mathcal{D}_{\text{VAR}}$	523 ±1044	279 ±220	>1000	164 ±62	179 ±202	118 ±28	173 ±111	111 ±25	124 ±70	98 ±20	102 ±25	97 ±19
$\mathcal{D}_{\text{APE}}$	94 ±20	93 ±25	92 ±22	89 ±17	85 ±12	82 ±14	84 ±17	82 ±10	86 ±12	81 ±12	84 ±15	82 ±10
$\mathcal{D}_{\text{G-MI}}$	90 ±14	88 ±15	88 ±24	85 ±15	86 ±14	81 ±12	85 ±21	79 ±11	89 ±7	84 ±8	83 ±11	80 ±5
$\mathcal{D}_{\text{T1}}$	83 ±18	76 ±17	76 ±19	80 ±19	82 ±19	77 ±18	76 ±19	75 ±13	80 ±15	77 ±14	76 ±15	75 ±13
$\mathcal{D}_{\text{PCA}}$	83 ±17	79 ±17	78 ±11	85 ±16	87 ±16	84 ±20	83 ±13	85 ±13	88 ±14	89 ±21	103 ±50	85 ±10
$\mathcal{D}_{\text{PCA2}}$	81 ±15	79 ±16	79 ±12	79 ±12	83 ±12	79 ±14	80 ±10	81 ±10	84 ±12	82 ±15	82 ±10	81 ±11

Table 5.8: Results of the T1VFA-CAROTID experiment. Mean and standard deviation over all subjects for all evaluation measures and control point spacings 8 mm, 16 mm, and 32 mm.

mTRE [mm]	***8***	***16***	***32***
(-)	1.47 ± 0.54	1.47 ± 0.54	1.47 ± 0.54
$\mathcal{D}_{\text{MI}}$	1.26 ± 0.44	1.22 ± 0.43	1.23 ± 0.45
$\mathcal{D}_{\text{VAR}}$	2.65 ± 0.79	1.84 ± 0.36	1.16 ± 0.37
$\mathcal{D}_{\text{APE}}$	1.24 ± 0.55	1.11 ± 0.41	1.10 ± 0.43
$\mathcal{D}_{\text{G-MI}}$	1.16 ± 0.48	1.08 ± 0.42	1.10 ± 0.43
$\mathcal{D}_{\text{PCA}}$	1.25 ± 0.56	1.11 ± 0.42	1.10 ± 0.43
$\mathcal{D}_{\text{PCA2}}$	1.13 ± 0.46	1.08 ± 0.39	1.10 ± 0.43
$\text{STD}_{\lvert \partial T_g / \partial x \rvert}$ [%]	***8***	***16***	***32***
(-)	0 ± 0	0 ± 0	0 ± 0
$\mathcal{D}_{\text{MI}}$	7 ± 1	2 ± 0	0 ± 0
$\mathcal{D}_{\text{VAR}}$	49 ± 25	6 ± 3	1 ± 0
$\mathcal{D}_{\text{APE}}$	6 ± 1	2 ± 1	0 ± 0
$\mathcal{D}_{\text{G-MI}}$	5 ± 1	1 ± 0	0 ± 0
$\mathcal{D}_{\text{PCA}}$	6 ± 1	2 ± 1	0 ± 0
$\mathcal{D}_{\text{PCA2}}$	5 ± 1	1 ± 0	0 ± 0
$90\%\sqrt{\text{CRLB}}\ T_1$ [ms]	***8***	***16***	***32***
(-)	>1000	>1000	>1000
$\mathcal{D}_{\text{MI}}$	530 ± 136	501 ± 83	523 ± 93
$\mathcal{D}_{\text{VAR}}$	>1000	>1000	848 ± 605
$\mathcal{D}_{\text{APE}}$	539 ± 154	498 ± 94	530 ± 95
$\mathcal{D}_{\text{G-MI}}$	514 ± 134	491 ± 83	515 ± 82
$\mathcal{D}_{\text{PCA}}$	540 ± 154	498 ± 93	530 ± 94
$\mathcal{D}_{\text{PCA2}}$	532 ± 154	510 ± 110	523 ± 87

the model will fit less accurately to the data, resulting in a higher estimated noise level and thus higher CRLB [6]. A 90% percentile was chosen because registration errors are mainly visible at the edges of a region and many voxels in the ROI will not make a difference, however we do not want to be sensitive to a very low number of extremely high outliers. Please note that when the $90\%\sqrt{\text{CRLB}}$ is high, the data is likely to be misaligned. However, when the $90\%\sqrt{\text{CRLB}}$ is low, it is not certain if the data is well aligned, which is why this measure should preferably be used in combination with other criteria.

5.5 RESULTS

The results in Tables 5.3 - 5.11 are color-coded. A red-to-white color scale is used to visually depict the (rank of the) results, where red indicates a worse result and white indicates a better result.

5.5.1 *Results on synthetic data*

The mean and standard deviation of $\|\boldsymbol{d}_g^*(\boldsymbol{x})\|$ of the SYNTH-MODEL images are shown in Table 5.3. The left column shows the mean and standard deviation of $\|\boldsymbol{d}_g^*(\boldsymbol{x})\|$ in the case of

Table 5.9: Results of the ADC-ABDOMEN experiment. Mean and standard deviation over all subjects for all evaluation measures and control point spacings 32 mm, 64 mm, and 128 mm and an affine transformation.

Dice_G [%]	*32*	*64*	*128*	Affine
(-)	70 ±4	70 ±4	70 ±4	70 ±4
$\mathcal{D}_{\text{MI}}$	61 ±18	64 ±16	73 ±8	72 ±7
$\mathcal{D}_{\text{VAR}}$	48 ±25	50 ±27	58 ±17	52 ±14
$\mathcal{D}_{\text{APE}}$	69 ±5	72 ±4	73 ±5	73 ±5
$\mathcal{D}_{\text{G-MI}}$	58 ±10 *	69 ±5 *	72 ±5 *	71 ±5
$\mathcal{D}_{\text{PCA}}$	65 ±13	71 ±5	71 ±4	72 ±4
$\mathcal{D}_{\text{PCA2}}$	75 ±7	75 ±5	73 ±5	71 ±5

$\text{STD}_{\partial T_g/\partial x}$ [%]	*32*	*64*	*128*	Affine
(-)	0 ±0	0 ±0	0 ±0	-
$\mathcal{D}_{\text{MI}}$	25 ±11	8 ±3	2 ±1	-
$\mathcal{D}_{\text{VAR}}$	33 ±13	8 ±5	2 ±1	-
$\mathcal{D}_{\text{APE}}$	32 ±15	7 ±4	1 ±1	-
$\mathcal{D}_{\text{G-MI}}$	27 ±11 *	7 ±3 *	1 ±0 *	-
$\mathcal{D}_{\text{PCA}}$	11 ±3	3 ±1	1 ±0	-
$\mathcal{D}_{\text{PCA2}}$	8 ±2	3 ±1	0 ±0	-

$90\%\sqrt{\text{CRLB}}$ ADC [$\mu\text{m}^2/\text{ms}$]	*32*	*64*	*128*	Affine
(-)	1.37±0.83	1.37±0.83	1.37±0.83	1.37±0.83
$\mathcal{D}_{\text{MI}}$	0.23±0.05	0.25±0.05	0.29±0.05	0.52±0.35
$\mathcal{D}_{\text{VAR}}$	0.20±0.08	0.23±0.10	0.35±0.16	0.43±0.21
$\mathcal{D}_{\text{APE}}$	0.42±0.06	0.42±0.13	0.40±0.14	0.50±0.29
$\mathcal{D}_{\text{G-MI}}$	0.44±0.10*	0.42±0.14*	0.37±0.05*	0.46±0.12
$\mathcal{D}_{\text{PCA}}$	0.15±0.04	0.23±0.06	0.46±0.27	0.55±0.30
$\mathcal{D}_{\text{PCA2}}$	0.23±0.03	0.27±0.05	0.41±0.18	0.50±0.26

*The registration of one subject failed and was not included in the results

Table 5.10: Results of the DTI-BRAIN experiment. Mean and standard deviation of the $90\%\sqrt{\text{CRLB}}$ MD over all subjects.

$90\%\sqrt{\text{CRLB}}$ MD [$\mu\text{m}^2/\text{ms}$]	Affine
(-)	0.096 ±0.029
$\mathcal{D}_{\text{MI}}$	0.084 ±0.028
$\mathcal{D}_{\text{VAR}}$	1.930 ±4.000
$\mathcal{D}_{\text{APE}}$	0.085 ±0.029
$\mathcal{D}_{\text{G-MI}}$	0.120 ±0.039
$\mathcal{D}_{\text{PCA}}$	0.085 ±0.029
$\mathcal{D}_{\text{PCA2}}$	0.084 ±0.028

$\boldsymbol{T}_g(\boldsymbol{x};\boldsymbol{\mu}_{\text{init},g})$ being an affine transformation, and in the right column the case of $\boldsymbol{T}_g(\boldsymbol{x};\boldsymbol{\mu}_{\text{init},g})$ being a B-spline transformation. Only the proposed methods $\mathcal{D}_{\text{PCA}}$ and $\mathcal{D}_{\text{PCA2}}$ were succesful in aligning the images.

Table 5.4 shows the mean ± standard deviation of $\|\boldsymbol{d}^*_g(\boldsymbol{x})\|$ of the synthetic qMRI data. This evaluation measure is calculated within the registration mask when that is present or else on the entire image. For all experiments except DCE-ABDOMEN, $\mathcal{D}_{\text{VAR}}$ showed the worst registration accuracy compared to all other methods. For s-T1MOLLI, $\mathcal{D}_{T_1}$ performed best, which is to be expected, since, in this simulation example, the intensities perfectly adhere to

Table 5.11: Results of the DCE-ABDOMEN experiment. Mean and standard deviation over all subjects for all evaluation measures and control point spacings 32 mm, 64 mm, and 128 mm and an affine transformation.

mTRE [mm]	*32*	*64*	*128*	Affine
(-)	8.49±4.54	8.49±4.54	8.49±4.54	8.49±4.54
$\mathcal{D}_{\text{MI}}$	6.73±2.02	6.46±2.32	6.47±2.37	7.35±3.55
$\mathcal{D}_{\text{VAR}}$	14.92±5.65	6.98±1.45	6.29±1.98	7.01±2.10
$\mathcal{D}_{\text{APE}}$	13.46±5.48	6.86±2.33	6.36±2.37	6.91±3.05
$\mathcal{D}_{\text{G-MI}}$	7.51±1.47	6.41±2.37	6.41±2.38	7.31±3.79
$\mathcal{D}_{\text{PCA}}$	6.21±2.25	6.11±2.32	6.24±2.37	6.54±2.69
$\mathcal{D}_{\text{PCA2}}$	5.89±2.23	5.99±2.17	6.18±2.27	6.62±2.75
$\text{STD}_{\lvert\partial T_g/\partial x\rvert}$ [%]	*32*	*64*	*128*	Affine
(-)	0 ±0	0 ±0	0 ±0	-
$\mathcal{D}_{\text{MI}}$	20 ±9	4 ±2	1 ±1	-
$\mathcal{D}_{\text{VAR}}$	22 ±14	4 ±1	1 ±0	-
$\mathcal{D}_{\text{APE}}$	21 ±12	4 ±1	1 ±0	-
$\mathcal{D}_{\text{G-MI}}$	17 ±6	3 ±1	1 ±0	-
$\mathcal{D}_{\text{PCA}}$	11 ±4	4 ±2	1 ±0	-
$\mathcal{D}_{\text{PCA2}}$	6 ±3	2 ±1	0 ±0	-
$90\%\sqrt{\text{CRLB}}\ K^{\text{trans}}$ [min^{-1}]	*32*	*64*	*128*	Affine
(-)	2.84±2.30	2.84±2.30	2.84±2.30	2.84±2.30
$\mathcal{D}_{\text{MI}}$	3.85±2.41	3.64±4.13	2.54±2.58	3.79±4.11
$\mathcal{D}_{\text{VAR}}$	3.08±2.68	1.38±0.74	1.28±0.86	1.54±0.95
$\mathcal{D}_{\text{APE}}$	2.18±1.56	1.49±1.30	1.59±1.68	1.63±1.24
$\mathcal{D}_{\text{G-MI}}$	1.61±1.05	1.42±1.15	1.37±1.23	2.08±2.05
$\mathcal{D}_{\text{PCA}}$	1.69±1.48	1.52±1.18	1.46±1.09	1.52±1.17
$\mathcal{D}_{\text{PCA2}}$	1.17±0.87	1.27±0.92	1.38±1.16	1.81±1.70

the T1-MOLLI model. For the remaining applications either $\mathcal{D}_{\text{PCA}}$ or $\mathcal{D}_{\text{PCA2}}$ outperformed the other methods.

Table 5.5 shows the $90\%\sqrt{\text{CRLB}}$ of the tissue property of interest of the synthetic datasets, before and after registration, evaluated in a specified ROI. For s-T1MOLLI the model-based metric $\mathcal{D}_{T_1}$ performed best. For all other applications both $\mathcal{D}_{\text{PCA}}$ or $\mathcal{D}_{\text{PCA2}}$ achieved the lowest $90\%\sqrt{\text{CRLB}}$ of the tissue property.

5.5.2 *Results on real data*

The results of the CT-LUNG experiment are shown in Table 5.6. The best control point spacing for this data is 13 mm. Method $\mathcal{D}_{\text{G-MI}}$ had the best performance, and $\mathcal{D}_{\text{PCA}}$ performed similar to $\mathcal{D}_{\text{MI}}$, $\mathcal{D}_{\text{VAR}}$ and $\mathcal{D}_{\text{APE}}$. Additionally, to allow for a direct comparison, we computed the mean of mTRE_l for all l of $\mathcal{D}_{\text{VAR}}$ on the subset of CT-LUNG used by [120]. In this subset we observed 1.26(0.33) where 1.26(0.27) was reported by Metz *et al.* Slight differences may be due to different mask and sampling strategies.

The results of the experiments on the real T1MOLLI-HEART data are shown in Table 5.7. Even though the lowest $90\%\sqrt{\text{CRLB}}$ is obtained for $\mathcal{D}_{T_1}$, the methods $\mathcal{D}_{\text{PCA}}$ and $\mathcal{D}_{\text{PCA2}}$ have the highest Dice_G and the lowest $\text{STD}_{\lvert\partial T_g/\partial x\rvert}$. $\mathcal{D}_{\text{VAR}}$ performed worst on all evaluation measures. Also notice that the result of $\mathcal{D}_{\text{MI}}$ depends heavily on the choice of reference image. The best results are obtained for reference image 11, which is the image acquired with the longest inversion time and therefore having the highest SNR.

The results on the T1VFA-CAROTID data are shown in Table 5.8. Because the registration ran separately for the left and right carotid, the mean over the two carotids is taken per subject. For all methods the lowest $90\%\sqrt{\text{CRLB}}$ T_1 is obtained at $\nu = 16$ mm, of which $\mathcal{D}_{\text{G-MI}}$ performed best. At $\nu = 16$ mm a similar point correspondence and transformation smoothness are obtained by all methods except for $\mathcal{D}_{\text{VAR}}$.

The results of the real ADC-ABDOMEN data are shown in Table 5.9. The $90\%\sqrt{\text{CRLB}}$ ADC in the ROI is decreased after registration for all methods and all control point spacings. Extreme deformations, especially at higher b-values cause the lower overlap, which can be seen from the relatively high $\text{STD}_{|\partial \boldsymbol{T}_g/\partial \boldsymbol{x}|}$ for low control point spacings. $\mathcal{D}_{\text{PCA2}}$ performed best in terms of Dice_G overlap and has the lowest $\text{STD}_{|\partial \boldsymbol{T}_g/\partial \boldsymbol{x}|}$ and $90\%\sqrt{\text{CRLB}}$ ADC for $\nu = 32$ mm.

The results of the DTI-BRAIN data are shown in Table 5.10. Both $\mathcal{D}_{\text{VAR}}$ and $\mathcal{D}_{\text{G-MI}}$ increase the $90\%\sqrt{\text{CRLB}}$ MD. A similar $90\%\sqrt{\text{CRLB}}$ MD is obtained for the remaining methods.

The results of the DCE-ABDOMEN data are shown in Table 5.11. The best $90\%\sqrt{\text{CRLB}}$ K^{trans} was obtained for $\mathcal{D}_{\text{PCA2}}$ for $\nu = 32$ mm, whereas $\mathcal{D}_{\text{MI}}$ performed worst on that measure with $\nu = 32$ mm. The point correspondence is worst for $\mathcal{D}_{\text{VAR}}$ at $\nu = 32$ mm. Overall $\mathcal{D}_{\text{PCA2}}$ showed the best point correspondence and the lowest $\text{STD}_{|\partial \boldsymbol{T}_g/\partial \boldsymbol{x}|}$.

5.5.3 *Visual inspection of qMRI data*

Figure 5.6 shows tissue maps before registration and after registration with $\mathcal{D}_{\text{PCA2}}$ of specific datasets that are used in the experiments. The difference in the estimated tissue maps is clearly visible, especially at tissue boundaries, showing the importance of registration as a pre-processing step prior to fitting the qMRI data.

5.6 DISCUSSION

We proposed two general registration methods for qMRI and we compared our methods to five other state-of-the-art registration methods. The following sections discuss in more detail the results of the synthetic and real data, some limitations of the evaluation, and possibilities for future work.

5.6.1 *Synthetic data*

A purely synthetic dataset was created to clearly demonstrate the advantages that the proposed PCA-based methods have. In this example with low pairwise mutual information between the images, the experiments confirmed that the methods based on mutual information ($\mathcal{D}_{\text{MI}}$, $\mathcal{D}_{\text{G-MI}}$), or groupwise correlation ($\mathcal{D}_{\text{APE}}$) failed to register the data. The proposed PCA-based methods gave very good registration results in this challenging example. It should be emphasized that, as the images in this synthetic example are linear combinations of a few basis images, it is expected that methods based on PCA give good results. Nevertheless this example provides insight in the functionality of the registration methods.

On synthetic data created using existing qMRI models, the PCA-based methods perform well in all experiments for both evaluation measures, whereas the other methods either fail or have only good results in a selection of the experiments.

For the s-T1VFA-CAROTID and T1VFA-CAROTID we used $L = 1$ instead of $L = \Gamma = 3$ for metric $\mathcal{D}_{\text{PCA}}$. Visual inspection of initial results led to this choice. A possible explanation why $L = 1$ showed better results is that the relatively small registration mask and the small number of acquired images could lead to only one dominant eigenvector. The second proposed metric, $\mathcal{D}_{\text{PCA2}}$, eliminates the necessity to choose L and shows results comparable to $\mathcal{D}_{\text{PCA}}$.

5.6.2 *Real data*

In the experiments where a B-spline transformation was used, we investigated the influence of different values for the control point spacing. All experiments show that the proposed PCA-based metrics show less extreme deformations and have better results than the other methods, which is particularly the case for low control point spacings.

The CT-LUNG data was used to evaluate the proposed methods on data without intensity variation. The best result on the CT-LUNG data is obtained with $\mathcal{D}_{\text{G-MI}}$, possibly due to the slight intensity variations that are present in the images. However, $\mathcal{D}_{\text{MI}}$, $\mathcal{D}_{\text{VAR}}$, $\mathcal{D}_{\text{APE}}$ and $\mathcal{D}_{\text{PCA}}$ showed competitive results.

In the T1MOLLI-HEART experiment we used different images as a reference image. The results show that the pairwise approach using $\mathcal{D}_{\text{MI}}$ heavily depends on the choice of reference image, while the groupwise approaches do not use a reference image during registration (and hence have a consistent performance across reference images).

$\mathcal{D}_{T_1}$ performed well on the synthetic data, but on real data the Dice_G was lower and the deformations were less smooth than those of the proposed methods. Furthermore, this method needs customization for different qMRI models.

5.6.3 *Limitations of the evaluation*

Evaluation of image registration results is usually a hard task because no ground truth is available. By using both synthetic and real qMRI data we extensively evaluated the registration methods. Registration accuracy was measured in terms of a mean and standard deviation of a residual deformation field (synthetic data), point correspondence or a groupwise Dice overlap (real data), the smoothness of the obtained deformation (real data), and the uncertainty of the qMRI parameter of interest (real and synthetic data).

It was impossible to accurately outline the ROIs on all images of the T1MOLLI-HEART data, because in some images there was not enough contrast between the tissues. This may result in an overly optimistic overlap measure, since those images might also be harder to register. In the DCE-ABDOMEN data, it proved difficult to manually annotate landmarks in the images, due to sliding motion of the various organs combined with the breathing motion. These problems, associated with the lack of a ground truth, were alleviated by reporting multiple evaluation measures, and by visually inspecting the data. In this way, a relatively complete picture of registration performance was given.

Smoothness and periodicity over time, which were modelled explicitly in [120], were not assumed or enforced in any of the experiments in our work. It could, however, be incorporated into our PCA-based method in the same way as proposed in [120]. Both in DCE-ABDOMEN and ADC-ABDOMEN we observed sliding motion. We did not explicitly account for this in the transformation model we used. However, since our methods can be extended with any other transformation model, one could use e.g. the transformation models proposed by Delmon *et al.* [170] and Berendsen *et al.* [171], which are designed to account for sliding motion.

No explicit regularization of the deformation field was used in the registration experiments. Different amounts of implicit regularization were compared by using different values for the control point spacings of the B-spline transformation model. Adding a regularization term such as bending energy may even further improve the results.

The CRLB was used as a measure to detect voxels that are fitted with a high uncertainty due to e.g. misalignment. Taking the 90% percentile over a voxel-wise fitted ROI is a good measure for registration accuracy and the values we report should be seen as such. It should be stressed that, since the 90% percentile is a rather conservative statistic, the $90\%\sqrt{\text{CRLB}}$ values reported could be misleading when seen as an indicator of the precision typically attained. For that purpose, the entire distribution of CRLB values in the ROI would have to be used.

5.6.4 *Future work*

Since the qMRI models are typically non-linear, a linear dimension-reduction technique such as PCA can only approximate the dimension of the subspace. It would therefore be interesting to investigate replacing PCA by a non-linear dimension reduction technique such as kernel-PCA or Laplacian eigenmaps, similar as in e.g. [172]. However, this is computationally demanding and it may not result in better registrations if the minimum of the dissimilarity metric remains at the same location in the transform parameter space.

The proposed registration methods are potentially applicable to any set of images where the voxel intensities are defined by a low-dimensional model. Possible other applications in medical imaging are perfusion computed tomography (CT) and dynamic positron emission tomography (PET). Perfusion CT is used for e.g. calculation of the myocardial blood flow index to detect functionally significant coronary lesions [173]. In perfusion CT a contrast agent is injected in the coronary arteries and the heart is imaged at several time points, giving each image a different contrast. Dynamic PET is used for e.g. estimating the energy consumption in tumors. In dynamic PET a radio-active tracer is injected and multiple time frames are imaged to assess both the spatial and temporal pattern of the tracer uptake [174]. Both applications acquire multiple time points in which the contrast changes and a low-dimensional model is fitted to the data to obtain quantitative information about tissues. Therefore, the proposed methods could potentially be used to correct the motion in such data.

5.7 CONCLUSION

We proposed two novel generic groupwise registration methods for qMRI. We evaluated our methods on five different applications of qMRI and compared it to five other state-of-the-art registration methods. We showed the advantage of a groupwise approach versus a pairwise approach and showed that our proposed methods have better or equal registration accuracy as the other methods. Both methods that were proposed showed good results. When no or little intensity changes are expected we would recommend using $\mathcal{D}_{\text{PCA}}$ and choose $L = 1$ for minimal computational complexity and good registration results. For compensating misalignments in qMRI the preferred dissimilarity metric is $\mathcal{D}_{\text{PCA2}}$, because it is free of any user-defined parameters.

5.8 FUNDING AND ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007 – 2013) under grant agreement no. 601055, VPH-DARE@IT and from VIDI Grant 639.072.411 from the Netherlands Organisation for Scientific Research (NWO).

We would like to thank Dr. H.M.M. van Beusekom for helping with the acquisition of the T1MOLLI-HEART data. The acquisition of the T1MOLLI-HEART data was financially supported by Agentschap NL (SENTER-NOVEM): “A novel approach to myocardial regeneration” to Dr. H.M.M. van Beusekom *et al.*

We would like to thank N.M. deSouza, MD, PhD, L. Bernardin, MD, and N. Douglas, MSc, Institute of Cancer Research, London, UK, for providing the ADC-ABDOMEN data. The ADC-ABDOMEN data were acquired in the context of the QuicConcept project, which has received support from the Innovative Medicines Initiative Joint Undertaking (`http://www.imi.europa.eu`) under grant agreement number 115151, resources of which are composed of financial contribution from the European Union’s Seventh Framework Programme (FP7/2007-2013) and EFPIA companies’ in kind contribution.

6

FAST MULTI-DIMENSIONAL B-SPLINE ALGORITHMS USING TEMPLATE METAPROGRAMMING

Abstract. B-splines are piecewise polynomial splines that are non-zero only within a certain domain, also known as the support region. Their multi-dimensional variants are widely used in image intensity interpolation and in non-rigid transformation models. The computation of an interpolated intensity or a transformed point involves a nested sum over all control points within the support region, and the computational complexity of this summation rapidly increases with increasing image dimension and B-spline order. However, due to the separability of the multi-dimensional B-spline it is possible to reformulate this nested sum as a recursion. In this work, we exploit this recursion property to create fast multi-dimensional algorithms for B-spline interpolation and transformation. Besides the B-spline functions themselves, their derivatives with respect to spatial coordinates and control point coefficients are also addressed, as these are commonly needed. For the recursive computation of the spatial derivatives we exploit derivatives of lower order, thus even further reducing the number of floating point additions and multiplications compared to a straight-forward recursion. An advantage of the recursive algorithms is that they can be implemented by template metaprogramming (TMP). TMP allows the compiler to create fast assembly code e.g. by unrolling loops, while maintaining a single code for any image dimension and B-spline order, and thereby assuring code readability, maintainability and extensibility. The proposed algorithms are compared to the non-recursive algorithms that are currently implemented in the Insight ToolKit, a popular open source image processing library written in C++, and NiftyReg, an open source image registration package. We show that for increasing image dimension the proposed algorithms are substantially faster, up to a factor of 4 for intensity interpolation and a factor of 18 for coordinate transformation. Furthermore, the proposed methods are applied in medical image registration problems, where B-spline interpolation and transformation account for a significant part of the computational load. Using the proposed algorithms, a complete image registration was on average 2 times faster for images of dimension 2, and up to $\sim$3.5 times faster for images of dimension 4. The C++ source code is made publicly available.

This chapter contains the content of *Fast multi-dimensional B-spline algorithms using template metaprogramming*, W. Huizinga et al., *submitted*.

6.1 INTRODUCTION

Image interpolation is one of the most common and fundamental tasks in image processing. Many intensity interpolation algorithms exist, but one of the most widely used methods is the B-spline interpolation algorithm. The primary reason is that B-splines provide a high interpolation accuracy. Moreover, they are compactly supported, a property which makes the computation efficient, and they are smooth, meaning that splines of degree n are $n-1$ times continuously differentiable. A good introduction to B-spline interpolation can be found in [175].

B-spline interpolation is not only used on the intensity of images, but has a much more widespread application. One of the most prominent uses next to intensity interpolation is to interpolate deformations of structures and images. Image registration is a technique that often exploits both intensity as well as transformation interpolation [119]. In computer graphics, Hsue *et al.* and Lee *et al.* used the B-spline model for free-form deformations [176, 177]. Currently it is widely used in the field of image registration, and in the specialized field of medical image registration [115]. In this transformation model, a grid of control points is defined on the image. The transformation of a coordinate is determined by interpolating its neighboring control point coefficients. In intensity-based image registration a coordinate transformation is found by minimizing the dissimilarity between images based on their intensities, e.g. by minimizing the mean square difference between deformed images. During the optimization procedure the image needs to be transformed and interpolated in each iteration, and, when gradient-based optimization methods are used, the image gradients and the derivative of the transformation with respect to its parameters (the control point displacements) also need to be computed. These computations account for a substantial portion of the computation time in typical registration procedures. The registration procedure can be computationally expensive, especially for high-dimensional images ($>$ 2D), which is often the case in the medical field. Decreasing computation time of B-spline algorithms would therefore have major impact on speed of image registration methods, and image processing software in general. While various acceleration techniques for image registration have been proposed that focus on more efficient optimization algorithms, e.g. [178–186], interpolation and transformation procedures still form a large part of the computational cost.

Previously it has been proposed to accelerate B-spline interpolation by approximating the B-spline weights using a precomputed look-up table [187]. In that work a speed-up was achieved at the cost of approximation errors, whereas we propose a highly accelerated algorithm that produces results within machine precision.

The use of GPUs is another approach for accelerating computations that has become an important area of research. In various previously proposed algorithms, the B-spline model for free-form deformations was ported to the GPU [188]. While a substantial acceleration may be achieved when using the GPU, these previous works only implemented the model for specific image dimension and spline order.

We propose recursive algorithms for B-spline interpolation and transformation for any possible image dimension and spline order. We also created an efficient algorithm for their spatial derivatives up to the second order and their derivatives with respect to control point coefficients. Since the spatial derivatives of the B-spline functions have many elements in common with lower order spatial derivatives, the recursive formulations of the higher order derivatives utilize the recursive functions of lower order derivatives, thereby reducing the number of required floating point operations. A second advantage of the re-use of lower-order derivatives is that those are returned together with the higher order derivatives with very little additional computation cost. Moreover, the proposed algorithms can be efficiently implemented for arbitrary image dimension and B-spline order with template metaprogramming (TMP). In TMP, templates are used to let the compiler generate efficient machine code. Language features such as for-loops and if-statements can be replaced by template specialization and recursion, allowing elimination of run-time flow control instructions [189]. Manually unrolling the loops also removes those instructions, but this does not lead to maintainable and readable code. With TMP the code is readable, maintainable and easily extendable to general cases. We exploit TMP to devise very efficient recursive B-spline implementations for a wide range of B-spline operations that are useful in image processing. The proposed algorithms are distributed as part of an open source image registration software package, with minor dependencies on that software, meaning that the proposed algorithms can be straightforwardly used in any (C++) software package.

In summary, the contributions of this work are four-fold: 1) We propose highly accelerated algorithms for B-spline interpolation and transformation, including associated derivatives, which produce results within machine precision. 2) By exploiting recursive algorithms in combination with TMP, efficient code is generated for any dimension and spline order. 3) For the computation of derivatives, novel recursive formulations are derived that reduce the computational complexity by re-using lower-order derivatives. 4) Source code of the proposed algorithms is made publicly available.

6.2 METHODS

We start in Sec. 6.2.1 with a general formulation of B-spline interpolation, and provide its spatial derivatives up to the second order, and its derivatives with respect to control point coefficients. Section 6.2.2 explains how we exploit the separability property to derive recursive formulations for the equations shown in Sec. 6.2.1. Section 6.2.3 explains the concept of TMP and shows pseudo-code of the proposed algorithms. Section 6.2.4 describes the algorithms that are used as a reference in this work. In Sec. 6.2.6 equations for the number of arithmetic operations of the reference and proposed algorithms are derived. Section 6.2.5 gives some implementation details and in Sec. 6.2.7 the basics and terminology of intensity-based image registration are given, explaining how B-spline intensity interpolation and B-spline coordinate transformation are used in that context.

6.2.1 *B-splines*

B-splines are piecewise polynomial splines that are non-zero only within a domain known as the support region. B-spline interpolation is defined as a sum of weighted control point coefficients $\boldsymbol{c}(\boldsymbol{k}) \in \mathbb{R}^{\Upsilon}$. In the case of intensity interpolation, $\boldsymbol{c}(\boldsymbol{k})$ are the intensity values of a filtered version of the image that is to be interpolated (and would often be called "image intensity coefficients"). The coefficients $\boldsymbol{c}(\boldsymbol{k})$, however, are not necessarily scalar ($\Upsilon = 1$), but can be multi-dimensional as well ($\Upsilon > 1$), e.g. in the case of coordinate transformation where Υ equals the image dimension D. See Figure 6.1 for an example with $\Upsilon = 1$. The control point coefficients are specified on a grid, and accessed by a multi-dimensional index $\boldsymbol{k} \in \mathbb{Z}^D$.

B-spline interpolation at a spatial coordinate $\boldsymbol{x} \in \mathbb{R}^D$ is described by the function $\boldsymbol{u} : \mathbb{R}^D \to \mathbb{R}^{\Upsilon}$, as follows:

$$\boldsymbol{u}(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathcal{Z}^n(\boldsymbol{x})} B^n(\boldsymbol{x} - \boldsymbol{k})\, \boldsymbol{c}(\boldsymbol{k}), \tag{6.1}$$

where $B^n(\cdot)$ is the multi-dimensional B-spline polynomial of order n, and $\mathcal{Z}^n(\cdot)$ is its support region, both defined below in more detail. Without loss of generality, the notation of Eq. (6.1) assumes the argument of B^n is unit-less, i.e. $\boldsymbol{x} - \boldsymbol{k}$ is normalized by the spacing between the control points in the grid, commonly known as the control point spacing. The multi-dimensional B-spline polynomial $B^n : \mathbb{R}^D \to \mathbb{R}$ can be written as a product of its one-dimensional functions $\beta^n : \mathbb{R} \to \mathbb{R}$:

$$\begin{aligned} B^n(\boldsymbol{x}) &= \prod_{i=1}^{D} \beta^n(x_i) \\ &= \beta^n(x_1)\beta^n(x_2)\cdots\beta^n(x_D). \end{aligned} \tag{6.2}$$

The support region $\mathcal{Z}^n(\boldsymbol{x})$ is the set of $(n+1)^D$ control points $\boldsymbol{k}$ in the grid for which $B^n(\boldsymbol{x} - \boldsymbol{k})$ is non-zero:

$$\mathcal{Z}^n(\boldsymbol{x}) = \left\{ \boldsymbol{k} \in \mathbb{Z}^D | B^n(\boldsymbol{x} - \boldsymbol{k}) \neq 0 \right\}. \tag{6.3}$$

Let $\mathcal{Z}^n(x_d)$ be the support region in dimension d, which is given by [175]:

$$\mathcal{Z}^n(x_d) = \left\{ \left\lceil x_d - \frac{n+1}{2} \right\rceil + i \Bigg|_{i=0,\ldots,n} \right\}. \tag{6.4}$$

Defining k_d to be control point index element d, and using Eq. (6.2), then Eq. (6.1) can be rewritten into:

$$\boldsymbol{u}(\boldsymbol{x}) = \sum_{k_1 \in \mathcal{Z}^n(x_1)} \sum_{k_2 \in \mathcal{Z}^n(x_2)} \cdots \sum_{k_D \in \mathcal{Z}^n(x_D)} \prod_{i=1}^{D} \beta^n(x_i - k_i)\boldsymbol{c}(\boldsymbol{k}), \tag{6.5}$$

where we will refer to $\beta^n(x_i - k_i)$ as the B-spline weights.

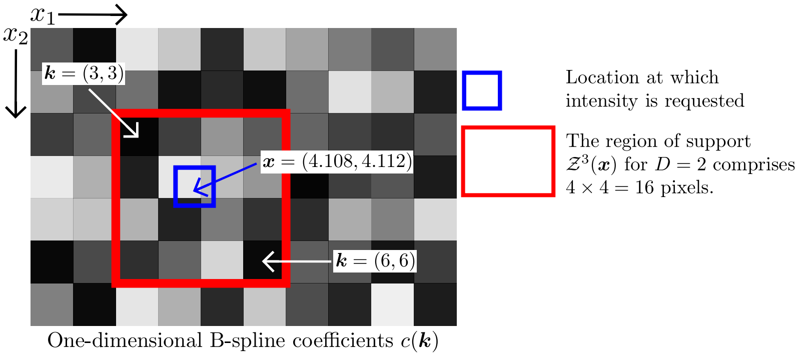


Figure 6.1: Graphical explanation of the variables used for the case $Y = 1$ and $D = 2$: $\boldsymbol{u}(\boldsymbol{x}) : \mathbb{R}^2 \to \mathbb{R}^1$. The intensity at $\boldsymbol{x} = (4.108, 4.112)$ is computed using B-spline interpolation.

6.2.1.1 *Spatial derivatives*

The first order spatial derivative of $\boldsymbol{u}(\boldsymbol{x})$ is given by:

$$\nabla \boldsymbol{u}(\boldsymbol{x}) = \begin{bmatrix} \partial \boldsymbol{u}/\partial x_1 \\ \partial \boldsymbol{u}/\partial x_2 \\ \vdots \\ \partial \boldsymbol{u}/\partial x_D \end{bmatrix} = \begin{bmatrix} \nabla_1 \boldsymbol{u} \\ \nabla_2 \boldsymbol{u} \\ \vdots \\ \nabla_D \boldsymbol{u} \end{bmatrix} \tag{6.6}$$

with

$$\nabla_i \boldsymbol{u} = \sum_{k_1 \in \mathcal{Z}^n(x_1)} \sum_{k_2 \in \mathcal{Z}^n(x_2)} \cdots \sum_{k_D \in \mathcal{Z}^n(x_D)} \frac{\partial \beta^n}{\partial x_i}(x_i - k_i) \left(\prod_{\substack{j=1 \\ j \neq i}}^{D} \beta^n(x_j - k_j) \right) \boldsymbol{c}(\boldsymbol{k}). \tag{6.7}$$

Similarly, the second order spatial derivative can be calculated as follows:

$$\nabla^2 \boldsymbol{u}(\boldsymbol{x}) = \begin{bmatrix} \nabla^2_{11}\boldsymbol{u} & \nabla^2_{12}\boldsymbol{u} & \dots & \nabla^2_{1D}\boldsymbol{u} \\ & \nabla^2_{22}\boldsymbol{u} & \dots & \nabla^2_{2D}\boldsymbol{u} \\ & & \ddots & \vdots \\ & & & \nabla^2_{DD}\boldsymbol{u} \end{bmatrix}, \tag{6.8}$$

with for $i \neq j$:

$$\nabla^2_{ij}\boldsymbol{u} = \nabla^2_{ji}\boldsymbol{u} =$$

$$\sum_{k_1 \in \mathcal{Z}^n(x_1)} \cdots \sum_{k_D \in \mathcal{Z}^n(x_D)} \frac{\partial \beta^n}{\partial x_i}(x_i - k_i) \frac{\partial \beta^n}{\partial x_j}(x_j - k_j) \times \left(\prod_{\substack{l=1 \\ l \neq i \\ l \neq j}}^{D} \beta^n(x_l - k_l) \right) \boldsymbol{c}(\boldsymbol{k}), \tag{6.9}$$

and for $i = j$:

$$\nabla^2_{ii}\boldsymbol{u} = \sum_{k_1 \in \mathcal{Z}^n(x_1)} \cdots \sum_{k_D \in \mathcal{Z}^n(x_D)} \frac{\partial^2 \beta^n}{\partial x_i^2}(x_i - k_i) \left(\prod_{\substack{l=1 \\ l \neq i}}^{D} \beta^n(x_l - k_l) \right) \boldsymbol{c}(\boldsymbol{k}). \tag{6.10}$$

6.2.1.2 *Derivatives with respect to* $\boldsymbol{c}$

We define $\boldsymbol{c}$ to be an array containing all Λ control points of dimension Υ, hence $\boldsymbol{c}$ is a $\Lambda \times \Upsilon$ tensor. An element of this tensor is accessed by $\boldsymbol{c}_j(\boldsymbol{k})$. An element i of the derivative of $\boldsymbol{u}(\boldsymbol{x})$ with respect to $\boldsymbol{c}_j(\boldsymbol{k})$ can be calculated as follows:

$$\frac{\partial u_i}{\partial c_j(\boldsymbol{k})}(\boldsymbol{x}) = \delta_{ij} \begin{cases} B^n(\boldsymbol{x} - \boldsymbol{k}) & \text{for } \boldsymbol{k} \in \mathcal{Z}^n(\boldsymbol{x}) \\ 0 & \boldsymbol{k} \notin \mathcal{Z}^n(\boldsymbol{x}) \end{cases} \tag{6.11}$$

where $i, j \in \{1, \ldots, \Upsilon\}$. Similarly we can compute the derivatives w.r.t. $\boldsymbol{c}$ of the spatial derivatives: $\partial \nabla \boldsymbol{u} / \partial \boldsymbol{c}$ and $\partial \nabla^2 \boldsymbol{u} / \partial \boldsymbol{c}$. Please note that $\partial \boldsymbol{u} / \partial \boldsymbol{c}$ is the derivative of a $\Upsilon \times 1$ tensor with respect to a $\Lambda \times \Upsilon$ tensor, resulting in a $\Lambda \times \Upsilon \times \Upsilon$ tensor. This tensor concept is used in the pseudocode of Algorithm 4.

6.2.2 *Recursive B-splines*

Recursion in computer science is a method where the solution to a problem depends on solutions to smaller instances of the same problem [190]. The separability of B^n, as was shown by Thevenaz *et al.*, is used here to derive a recursive formulation for the B-spline functions [150]. Please note that all algorithms and equations in this paper are written using one-based counting, which means that the initial element of a sequence is assigned the index 1.

Using the separability of B^n, Eq. (6.1) can be rewritten into:

$$\boldsymbol{u}(\boldsymbol{x}) = \sum_{k_1 \in \mathcal{Z}^n(x_1)} \beta^n(x_1 - k_1) \sum_{k_2 \in \mathcal{Z}^n(x_2)} \beta^n(x_2 - k_2) \ldots \sum_{k_D \in \mathcal{Z}^n(x_D)} \beta^n(x_D - k_D) \boldsymbol{c}(\boldsymbol{k}) \tag{6.12}$$

To derive a recursive formulation of Eq. (6.12) we introduce a vectorized notation for $\boldsymbol{c}(\boldsymbol{k})$. The coefficients $\boldsymbol{c}(\boldsymbol{k})$, which are accessed by a D-dimensional index $\boldsymbol{k}$, are vectorized into $\widetilde{c}(\kappa)$ with $\kappa = 1 + \boldsymbol{\alpha} \cdot (\boldsymbol{k} - \mathbf{1})$, where α_d is defined as the step in the linear index κ required to obtain a unit step in dimension d in the coefficients $\boldsymbol{c}(\boldsymbol{k})$. Let λ_d be the size of the control point grid in dimension d. For a 3D grid of coefficients of size $\Lambda = \lambda_1 \times \lambda_2 \times \lambda_3$ we obtain for example $\boldsymbol{\alpha} = (1, \lambda_1, \lambda_1\lambda_2)$ (depending on the mapping of dimensions to memory). A graphical example in 2D is shown in Figure 6.2. Through this vectorization, the linear index can be calculated recursively. Using the vectorized coefficients $\widetilde{c}(\kappa)$ and Eq. (6.12) the recursive interpolation

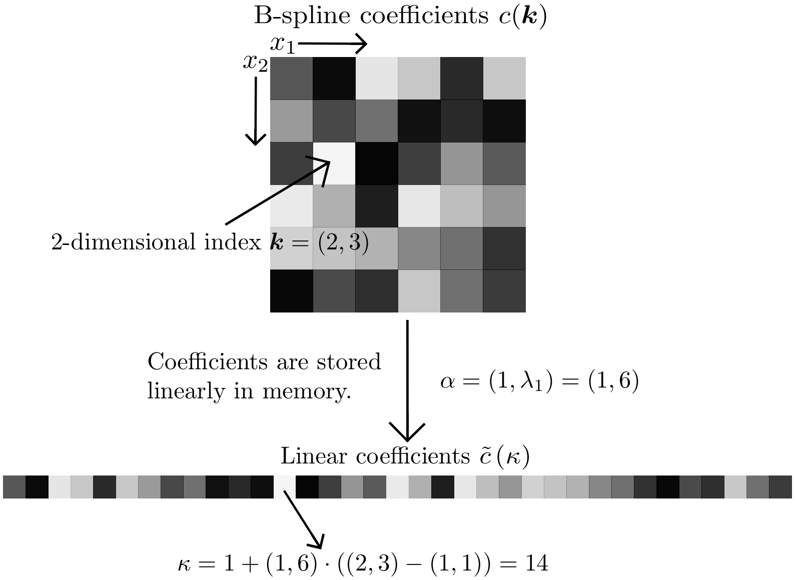


Figure 6.2: Graphical example of how to compute κ for $D = 2$. Please note that the top-left element of $\boldsymbol{c}(\boldsymbol{k})$ has index $\boldsymbol{k} = (1,1)$ and the first element of $\widetilde{c}(\kappa)$ has linear index $\kappa = 1$.

function $\boldsymbol{V}(d, \boldsymbol{x}, \kappa)$ (V for value) for dimension index d is given by:

$$\boldsymbol{V}(d, \boldsymbol{x}, \kappa) = \begin{cases} \sum\limits_{k_d \in \mathcal{Z}^n(x_d)} \beta^n(x_d - k_d)\boldsymbol{V}(d-1, \boldsymbol{x}, \kappa + \alpha_d(k_d - 1)) & d \geq 1 \\ \widetilde{c}(\kappa) & d = 0 \end{cases} . \tag{6.13}$$

The interpolated value at $\boldsymbol{x}$ equals:

$$\boldsymbol{u}(\boldsymbol{x}) = \boldsymbol{V}(D, \boldsymbol{x}, 1). \tag{6.14}$$

Pseudocode of the implementation of Eq. (6.13) is shown in Algorithm 1(a). Figure 6.3 shows the execution flow for the case $D = 3$ and $n = 3$.

6.2.2.1 *Recursion of spatial derivatives*

Often, both the value of $\boldsymbol{u}(\boldsymbol{x})$ as its derivatives are needed. The summations in Eq. (6.12) also appear in the computation of the spatial derivatives. A straight-forward recursive formulation for computing $\nabla \boldsymbol{u}$ would be to compute each $\nabla_i \boldsymbol{u}$ separately, using a similar function as in Eq. (6.13). However, due to the occurrence of repetitive elements, a recursive formula which utilizes $\boldsymbol{V}$ can be derived, leading to reduction of arithmetic operations compared to a straight-forward formulation. Moreover, when utilizing $\boldsymbol{V}$, both $\boldsymbol{u}$ and $\nabla \boldsymbol{u}$ are computed, which is an

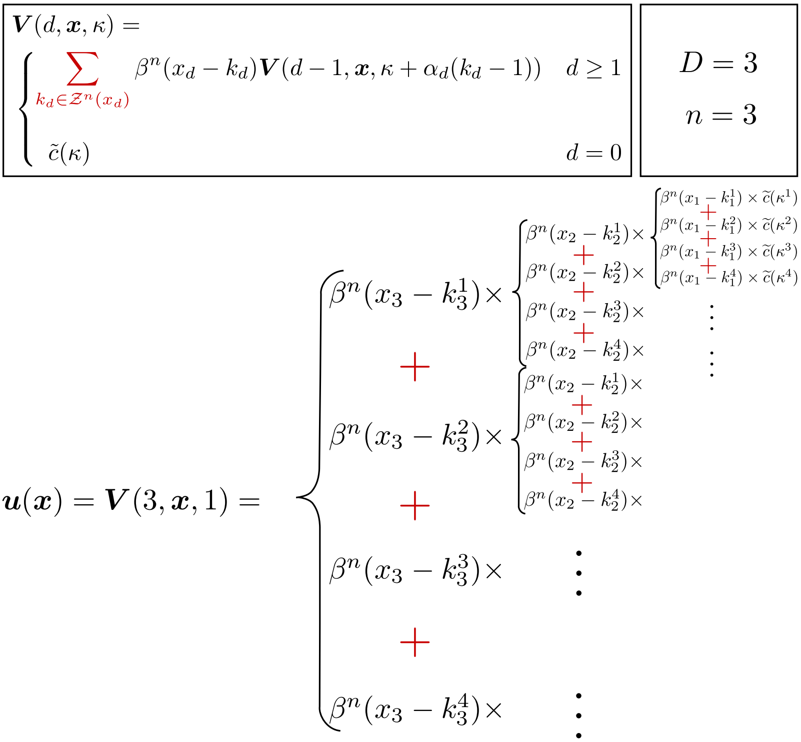


Figure 6.3: Execution flow for computing $V(3, x, 1)$ for $n = 3$. The number of control points in the region of support in each dimension is $n + 1 = 4$, meaning that the total number of control points equals $4^3 = 64$ and we get 64 different indices κ at the end of the flow, indicated with a superscript. In this scheme the summation is written explicitly over all $k_d \in \mathcal{Z}^n(x_d)$, also indicated with a superscript.

advantage when both are needed at the same time.

We can write the spatial derivative elements of $\boldsymbol{u}$ as:

$$\nabla_1 \boldsymbol{u} = \sum_{k_D \in \mathcal{Z}^n(x_D)} \beta^n(x_D - k_D) \cdots \sum_{k_2 \in \mathcal{Z}^n(x_2)} \beta^n(x_2 - k_2) \sum_{k_1 \in \mathcal{Z}^n(x_1)} \frac{\partial \beta^n}{\partial x_1}(x_1 - k_1) \boldsymbol{c}(\boldsymbol{k}), \tag{6.15}$$

$$\nabla_2 \boldsymbol{u} = \sum_{k_D \in \mathcal{Z}^n(x_D)} \beta^n(x_D - k_D) \cdots \sum_{k_2 \in \mathcal{Z}^n(x_2)} \frac{\partial \beta^n}{\partial x_2}(x_2 - k_2) \sum_{k_1 \in \mathcal{Z}^n(x_1)} \beta^n(x_1 - k_1) \boldsymbol{c}(\boldsymbol{k}), \tag{6.16}$$

etcetera, until the last element:

$$\nabla_D \boldsymbol{u} = \sum_{k_D \in \mathcal{Z}^n(x_D)} \frac{\partial \beta^n}{\partial x_D}(x_D - k_D) \cdots \sum_{k_2 \in \mathcal{Z}^n(x_2)} \beta^n(x_2 - k_2) \sum_{k_1 \in \mathcal{Z}^n(x_i)} \beta^n(x_1 - k_1) \boldsymbol{c}(\boldsymbol{k}). \tag{6.17}$$

We define the recursive function $\boldsymbol{G}(d, \boldsymbol{x}, \kappa)$ (G for gradient), which utilizes $\boldsymbol{V}(d, \boldsymbol{x}, \kappa)$. Let $\boldsymbol{V}_{d-1} = \boldsymbol{V}(d-1, \boldsymbol{x}, \kappa + \alpha_d(k_d - 1))$, and let $\boldsymbol{G}_{d-1} = \boldsymbol{G}(d-1, \boldsymbol{x}, \kappa + \alpha_d(k_d - 1))$, then:

$$\boldsymbol{G}(d, \boldsymbol{x}, \kappa) = \begin{cases} \displaystyle\sum_{k_d \in \mathcal{Z}^n(x_d)} \begin{bmatrix} \beta^n(x_d - k_d)\boldsymbol{G}_{d-1} \\ \frac{\partial \beta^n}{\partial x_d}(x_d - k_d)\boldsymbol{V}_{d-1} \end{bmatrix} & d \geq 1 \\ \\ [\,] & d = 0 \end{cases}, \tag{6.18}$$

with

$$\nabla \boldsymbol{u}(\boldsymbol{x}) = \boldsymbol{G}(D, \boldsymbol{x}, 1). \tag{6.19}$$

Here $[\,]$ is a $0 \times \Upsilon$ empty array, and $\boldsymbol{G}(d, \boldsymbol{x}, \kappa)$ returns a $d \times \Upsilon$ array, so its size reduces in each recursion.

The second order derivative $\nabla^2 \boldsymbol{u}(\boldsymbol{x})$ can be computed similarly with a recursive function $\boldsymbol{H}(d, \boldsymbol{x}, \kappa)$ (H for Hessian), which returns the upper-triangular elements of a symmetric $d \times d \times \Upsilon$ matrix. Here we can utilize both $\boldsymbol{G}(d, \boldsymbol{x}, \kappa)$ and $\boldsymbol{V}(d, \boldsymbol{x}, \kappa)$. Let $\boldsymbol{H}_{d-1} = \boldsymbol{H}(d-1, \boldsymbol{x}, \kappa + \alpha_d(k_d - 1))$, then:

$$\boldsymbol{H}(d, \boldsymbol{x}, \kappa) = \begin{cases} \displaystyle\sum_{k_d \in \mathcal{Z}^n(x_d)} \begin{bmatrix} \beta^n(x_d - k_d)\boldsymbol{H}_{d-1} & \frac{\partial \beta^n}{\partial x_d}(x_d - k_d)\boldsymbol{G}_{d-1} \\ \cdots & \frac{\partial^2 \beta^n}{\partial x_d^2}(x_d - k_d)\boldsymbol{V}_{d-1} \end{bmatrix} & d \geq 1 \\ \\ [\,]. & d = 0 \end{cases}, \tag{6.20}$$

where $[\,]$ is an empty array of size $0 \times 0 \times \Upsilon$. Please note that $\boldsymbol{G}_{d-1}$ and $\boldsymbol{V}_{d-1}$ are permuted to match size $d \times 1 \times \Upsilon$ and $1 \times 1 \times \Upsilon$, respectively. The second order derivative is given by

$$\nabla^2 \boldsymbol{u}(\boldsymbol{x}) = \boldsymbol{H}(D, \boldsymbol{x}, 1). \tag{6.21}$$

6.2.2.2 *Recursion of derivatives with respect to* $\boldsymbol{c}$

The derivative of $\boldsymbol{u}(\boldsymbol{x})$ with respect to the control point coefficients $\boldsymbol{c}(\boldsymbol{k})$ is a sparse array, since $\partial u_i / \partial c_j(\boldsymbol{k})$ is non-zero only when $i = j$ and when $\boldsymbol{k} \in \mathcal{Z}^n(\boldsymbol{x})$. From Eq. (6.11) we see that each element of this sparse array is a multiplication of the one-dimensional B-spline weights, i.e. the multi-dimensional B-spline weights $B^n(\boldsymbol{x} - \boldsymbol{k})$. This multiplication is obtained recursively for the entire sparse array $\partial \boldsymbol{u} / \partial \boldsymbol{c}$ using the following expression:

$$\boldsymbol{J}(d, \boldsymbol{x}, \kappa, v) = \begin{cases} \displaystyle\sum_{k_d \in \mathcal{Z}^n(x_d)} \boldsymbol{J}(d-1, \boldsymbol{x}, \kappa + \alpha_d(k_d - 1), v\beta^n(x_d - k_d)) & d \geq 1 \\ \\ v\mathbf{1}^{\kappa} & d = 0 \end{cases}. \tag{6.22}$$

$$\partial \boldsymbol{u}(\boldsymbol{x}) / \partial \boldsymbol{c} = \boldsymbol{J}(D, \boldsymbol{x}, 1, 1). \tag{6.23}$$

The variable $v \in \mathbb{R}$ was introduced to recursively build up the elements $B^n(\boldsymbol{x} - \boldsymbol{k})$, which are assigned to the correct elements of $\partial \boldsymbol{u}/\partial \boldsymbol{c}$ in the end case $d = 0$. The indicator matrix $\mathbf{1}^{\kappa}$ is of the same size as $\partial \boldsymbol{u}/\partial \boldsymbol{c}$, selecting the elements $\partial \boldsymbol{u}_j/\partial \boldsymbol{c}_j(\kappa)$ $\forall j$. As each value of κ in the end case ($d = 0$) occurs not more than once, the actual recursive implementation omits the summation over k_d; the tensor $\partial \boldsymbol{u}(\boldsymbol{x})/\partial \boldsymbol{c}$ is initialized by a $\Lambda \times \Upsilon \times \Upsilon$ array of zeros, and in the recursion all non-zero elements of $\partial \boldsymbol{u}(\boldsymbol{x})/\partial \boldsymbol{c}$ are stored at the correct locations in the array. The pseudo-code presented in the next section clarifies this execution flow.

In a similar manner, recursive functions for $\partial \nabla \boldsymbol{u}\left(\boldsymbol{x}\right)/\partial \boldsymbol{c}$ and $\partial \nabla^2 \boldsymbol{u}\left(\boldsymbol{x}\right)/\partial \boldsymbol{c}$ can be derived.

6.2.3 *Template metaprogramming*

Template metaprogramming (TMP) is a programming technique in which templates are used by a compiler to generate efficient assembly code [189]. Since D is relatively low and known at compile time it can be set as a template argument. Since n is also known at compile time, this allows the compiler to unroll the entire weighted sum of Eq. (6.1). For the loops in the algorithms of this paper, loop unrolling may lead to a significant reduction in run-time, since e.g. jumps and loop count decrements are eliminated from program execution. Due to the low number of arithmetic operations, the avoided overhead may be substantial.

Algorithm 1 shows the recursive and the TMP-recursive implementation to compute $\boldsymbol{u}(\boldsymbol{x})$ (see Eq. (6.13)). TMP pseudocode of the algorithms to compute $\nabla \boldsymbol{u}(\boldsymbol{x})$ and $\nabla^2 \boldsymbol{u}(\boldsymbol{x})$, and $\partial \boldsymbol{u}(\boldsymbol{x})/\partial \boldsymbol{c}$ are shown in Algorithm 2, Algorithm 3, and Algorithm 4, respectively. Template arguments are shown between $< \ldots >$ brackets, as in the C++ syntax. The outputs are initialized with zero-arrays, denoted by $\mathbf{0}_d$, where the zero-array is of size d. The zero-array $\mathbf{0}_{d\Upsilon}$ is of size $d \times \Upsilon$. Temporary outputs of a single recursion are denoted by a tilde, e.g. $\widetilde{\boldsymbol{u}}, \widetilde{\nabla \boldsymbol{u}}$. During each recursion $\nabla \boldsymbol{u}$ has d elements of dimension Υ. Element i of the spatial derivative array $\nabla \boldsymbol{u}$ is denoted by $\nabla_i \boldsymbol{u}$. The upper-triangular elements of $\nabla^2 \boldsymbol{u}(\boldsymbol{x})$ are stored in a one-dimensional array. Let $h = d(d+1)/2$, then the array indices are structured as:

$$\begin{bmatrix} 1 & 2 & 4 & \cdots & \\ & 3 & 5 & \cdots & \\ & & 6 & & \\ & & & \ddots & \\ & & & & h \end{bmatrix}.$$

Since $\boldsymbol{H}(d, \boldsymbol{x}, \kappa)$ returns the upper-triangular elements of a symmetric $d \times d \times \Upsilon$ matrix, during each recursion $\nabla^2 \boldsymbol{u}$ has $d(d+1)/2$ elements of dimension Υ. Element i of the second order spatial derivative array $\nabla^2 \boldsymbol{u}$ is denoted by $\nabla_i^2 \boldsymbol{u}$.

Since Algorithm 2 computes $\boldsymbol{u}\left(\boldsymbol{x}\right)$ as an almost free by-product, it returns both $\boldsymbol{u}\left(\boldsymbol{x}\right)$ and $\nabla \boldsymbol{u}\left(\boldsymbol{x}\right)$. Similarly, Algorithm 3 returns $\boldsymbol{u}\left(\boldsymbol{x}\right), \nabla \boldsymbol{u}\left(\boldsymbol{x}\right)$, and $\nabla^2 \boldsymbol{u}\left(\boldsymbol{x}\right)$. In Algorithm 4 $\partial \boldsymbol{u}/\partial \boldsymbol{c}$ is passed by reference (when an argument is passed by reference to a function the value of this argument can be modified by that function). This is necessary because this array is filled in the end-case with elements that are computed during the recursion. Since many elements of this array are zero, it is initialized as a zero-array and only the non-zero elements are written at the end of the recursion. The actual implementation uses a sparse array.

6.2.4 *Reference algorithms*

As a reference, we use the implementations of the B-spline functions in ITK and NiftyReg.

6.2.4.1 *ITK*

ITK is a widely used image processing library for C++ and many open source software packages for image processing use ITK, e.g. [119], [191]. The Elsevier journal Medical Image Analysis dedicated an entire issue to research that was enabled by the ITK library, showing the extent of its usage in multimodal registration, segmentation, algorithm validation, software integration, and medical image analysis education [192].

The B-spline functions in ITK are direct implementations of Eq. (6.5). This means that the implementations use nested for-loops resulting in a higher complexity than the separable and recursive formulations in this paper.

For a fair comparison of the reference and the proposed algorithms we made some improvements to the implementations in ITK. Firstly, derivatives that were not yet implemented were added in such a way that the implementation was similar to existing implementations. Secondly, since the proposed algorithms are templated over n and D, we also templated the reference algorithms over n and D if this was not yet the case. Finally, whenever possible, temporary memory was allocated on the stack.

6.2.4.2 *NiftyReg*

NiftyReg is an open-source software for efficient medical image registration [193]. We adopted the CPU implementation of the B-spline coordinate transformation function in NiftyReg and used it as a second reference. This algorithm also uses nested for-loops, but many loops have been manually unrolled, which may result in a more efficient implementation. NiftyReg has separate implementations for three cases of coordinate transformation, i.e. $\Upsilon = D$, in which D and n are hard-coded: $< D, n > \in \{< 2, 3 >, < 3, 1 >, < 3, 3 >\}$.

6.2.5 *Implementation details*

The methods were implemented in C++. We created separate base classes for the image interpolator and the coordinate transformation, which means that users can easily integrate these into their own applications. The algorithms are implemented in the open source image registration software Elastix [119], which is freely available under the liberal Apache 2.0 license. In

Algorithm 1 Pseudo code of recursive (a) and TMP-recursive (b) algorithms to compute $\boldsymbol{u}(\boldsymbol{x})$.

(a)

```
function $\boldsymbol{u} = \boldsymbol{V}(d, \boldsymbol{x}, \kappa)$
    if $d == 0$ then
        return $\widetilde{c}(\kappa)$
    else
        $\boldsymbol{u} = \boldsymbol{0}_Y$
        for $k_d \in \mathcal{Z}^n(x_d)$ do
            $\tilde{\boldsymbol{u}} = \boldsymbol{V}(d-1, \boldsymbol{x}, \kappa + \alpha_d(k_d - 1))$
            $\boldsymbol{u} = \boldsymbol{u} + \tilde{\boldsymbol{u}}\beta^n(x_d - k_d)$
        end for
        return $\boldsymbol{u}$
    end if
end function

$\boldsymbol{u}(\boldsymbol{x}) = \boldsymbol{V}(D, \boldsymbol{x}, 1)$
```

(b)

```
template< $d$ > class TMP
    function $\boldsymbol{u} = \boldsymbol{V}^{\mathrm{TMP}}(\boldsymbol{x}, \kappa)$
        $\boldsymbol{u} = \boldsymbol{0}_Y$
        for $k_d \in \mathcal{Z}^n(x_d)$ do
            $\tilde{\boldsymbol{u}}$ =TMP< $d-1$ >::$\boldsymbol{V}^{\mathrm{TMP}}(\boldsymbol{x}, \kappa + \alpha_d(k_d - 1))$
            $\boldsymbol{u} = \boldsymbol{u} + \tilde{\boldsymbol{u}}\beta^n(x_d - k_d)$
        end for
        return $\boldsymbol{u}$
    end function

// Template specialization for $d = 0$
template<> class TMP< 0 >
    function $\boldsymbol{u} = \boldsymbol{V}^{\mathrm{TMP}}(\boldsymbol{x}, \kappa)$
        return $\widetilde{c}(\kappa)$
    end function

$\boldsymbol{u}(\boldsymbol{x})$ = TMP< $D$ >::$\boldsymbol{V}^{\mathrm{TMP}}(\boldsymbol{x}, 1)$
```

the experiments we used ITK version 4.9.0 as reference and Elastix version 4.8.1.

The actual implementations differ slightly from the pseudocode shown in Algorithms 1 - 4:

- To be as flexible as possible, and to be able to compile our software for multiple settings of n and Y, n and Y are also template parameters in the actual implementations. For clarity, those parameters are not set as template parameters in Algorithms 1, 2, 3, and 4, but as constants, since they do not change in the recursion.
- The support region $\mathcal{Z}^n(\boldsymbol{x})$, the B-spline weights $\beta^n(x_d - k_d)$, the first order B-spline derivative weights $\partial\beta^n(x_d - k_d)/\partial x_d$ and the second order B-spline derivative weights $\partial^2\beta^n(x_d - k_d)/\partial x_d^2$ are precomputed and accessed in the functions $\boldsymbol{V}^{\mathrm{TMP}}$, $\boldsymbol{G}^{\mathrm{TMP}}$, $\boldsymbol{H}^{\mathrm{TMP}}$, and $\boldsymbol{J}^{\mathrm{TMP}}$. The computation of those weights is not included in the complexity analysis below, as those additional computations are identical in both the ITK and the TMP-recursive implementation. In both the non-recursive and the proposed implementa-

Algorithm 2 Pseudo code of TMP-recursive implementation to compute $\nabla \boldsymbol{u}(\boldsymbol{x})$, with $\boldsymbol{u}(\boldsymbol{x})$ as almost free by-product.

```
template< $d$ > class TMP
    function $\boldsymbol{u}, \nabla \boldsymbol{u} = \boldsymbol{G}^{\text{TMP}}(\boldsymbol{x}, \kappa)$
        $\boldsymbol{u} = \mathbf{0}_Y$
        $\nabla \boldsymbol{u} = \mathbf{0}_{dY}$
        for $k_d \in \mathcal{Z}^n(x_d)$ do
            $\widetilde{\boldsymbol{u}}, \widetilde{\nabla \boldsymbol{u}}$ = TMP< $d-1$ >::$\boldsymbol{G}^{\text{TMP}}\left(\boldsymbol{x}, \kappa + \alpha_d(k_d - 1)\right)$
            $\boldsymbol{u} = \boldsymbol{u} + \widetilde{\boldsymbol{u}} \beta^n(x_d - k_d)$
            for $i = 1$ to $d-1$ do
                $\nabla_i \boldsymbol{u} = \nabla_i \boldsymbol{u} + \widetilde{\nabla_i \boldsymbol{u}} \beta^n(x_d - k_d)$
            end for
            $\nabla_d \boldsymbol{u} = \nabla_d \boldsymbol{u} + \widetilde{\boldsymbol{u}} \frac{\partial \beta^n}{\partial x_d}(x_d - k_d)$
        end for
        return $\boldsymbol{u}, \nabla \boldsymbol{u}$
    end function

// Template specialization for $d = 0$
template<> class TMP< 0 >
    function $\boldsymbol{u}, \nabla \boldsymbol{u} = \boldsymbol{G}^{\text{TMP}}(\boldsymbol{x}, \kappa)$
        $\boldsymbol{u} = \widetilde{c}(\kappa)$
        $\nabla \boldsymbol{u} = []$
        return $\boldsymbol{u}, \nabla \boldsymbol{u}$
    end function

$\boldsymbol{u}(\boldsymbol{x}), \nabla \boldsymbol{u}(\boldsymbol{x})$ = TMP< $D$ >::$\boldsymbol{G}^{\text{TMP}}(\boldsymbol{x}, 1)$
```

tions temporary memory required during the evaluation of these functions was allocated on the stack.

- Note that most multiplications and additions in the algorithm require Y operations as they operate on Y long element arrays.
- To be consistent with the ITK algorithm, we apply mirror boundary conditions in the case of intensity interpolation ($Y = 1$). The boundary conditions are applied to the pre-computed support region $\mathcal{Z}^n(\boldsymbol{x})$ before calling the recursive functions.
- The input of $B^n(\cdot)$ is not unitless in the actual implementations, and is corrected by the control point spacing and the coordinate system directions of the image, as are the spatial derivatives.

6.2.6 *Complexity analysis*

A way to compare the computational complexity is by counting the number of arithmetic operations, which we define as the number of *floating point* additions and multiplications in the algorithms. Since the evaluations of $\beta^n(x_d - k_d)$, $\beta^n(x_d - k_d)/\partial x_d$ and $\partial^2 \beta^n(x_d - k_d)/\partial x_d^2$ are identical in the proposed and the ITK algorithm, the arithmetic operations to compute those are ignored. Let N be the number of arithmetic operations, depending on the image dimension D, the spline order n, and the control point coefficient dimension Y. The ITK algorithm for B-spline interpolation is a loop over $\mathcal{Z}^n$, consisting of $(n+1)^D$ control points, i.e.

Algorithm 3 Pseudo code of TMP implementation to compute $\nabla^2 \boldsymbol{u}(\boldsymbol{x})$, with $\nabla \boldsymbol{u}(\boldsymbol{x})$ and $\boldsymbol{u}(\boldsymbol{x})$ as almost free by-products.

```
template< $d$ > class TMP
    function $\boldsymbol{u}, \nabla\boldsymbol{u}, \nabla^2\boldsymbol{u} = \boldsymbol{H}^{\mathrm{TMP}}(\boldsymbol{x}, \kappa)$
        $\boldsymbol{u} = \boldsymbol{0}_{\Upsilon}$
        $\nabla\boldsymbol{u} = \boldsymbol{0}_{d\Upsilon}$
        $\nabla^2\boldsymbol{u} = \boldsymbol{0}_{h\Upsilon}$

        for $k_d \in \mathcal{Z}^n(x_d)$ do
            $\widetilde{\boldsymbol{u}}, \widetilde{\nabla\boldsymbol{u}}, \widetilde{\nabla^2\boldsymbol{u}}$ = TMP< $d-1$ >::$\boldsymbol{H}^{\mathrm{TMP}}\left(\boldsymbol{x}, \kappa + \alpha_d(k_d - 1)\right)$
            $\boldsymbol{u} = \boldsymbol{u} + \widetilde{\boldsymbol{u}}\beta^n(x_d - k_d)$
            for $i = 1$ to $d-1$ do
                $\nabla_i\boldsymbol{u} = \nabla_i\boldsymbol{u} + \widetilde{\nabla_i\boldsymbol{u}}\beta^n(x_d - k_d)$
            end for
            $\nabla_d\boldsymbol{u} = \nabla_d\boldsymbol{u} + \widetilde{\boldsymbol{u}}\frac{\partial\beta^n}{\partial x_d}(x_d - k_d)$
            for $i = 1$ to $d(d-1)/2$ do
                $\nabla^2_i\boldsymbol{u} = \nabla^2_i\boldsymbol{u} + \widetilde{\nabla^2_i\boldsymbol{u}}\beta^n(x_d - k_d)$
            end for
            for $i = d(d-1)/2 + 1$ to $h - 1$ do
                $j = i - d(d-1)/2$
                $\nabla^2_i\boldsymbol{u} = \nabla^2_i\boldsymbol{u} + \widetilde{\nabla_j\boldsymbol{u}}\frac{\partial\beta^n}{\partial x_d}(x_d - k_d)$
            end for
            $\nabla^2_h\boldsymbol{u} = \nabla^2_h\boldsymbol{u} + \widetilde{\boldsymbol{u}}\frac{\partial^2\beta^n}{\partial x_d^2}(x_d - k_d)$
        end for
        return $\boldsymbol{u}, \nabla\boldsymbol{u}, \nabla^2\boldsymbol{u}$
    end function

// Template specialization for $d = 0$
template<> class TMP< $0$ >
    function $\boldsymbol{u}, \nabla\boldsymbol{u}, \nabla^2\boldsymbol{u} = \boldsymbol{H}^{\mathrm{TMP}}(\boldsymbol{x}, \kappa)$
        $\boldsymbol{u} = \widetilde{c}(\kappa)$
        $\nabla\boldsymbol{u} = []$
        $\nabla^2\boldsymbol{u} = []$
        return $\boldsymbol{u}, \nabla\boldsymbol{u}, \nabla^2\boldsymbol{u}$
    end function

$\boldsymbol{u}(\boldsymbol{x}), \nabla\boldsymbol{u}(\boldsymbol{x}), \nabla^2\boldsymbol{u}(\boldsymbol{x})$ = TMP< $D$ >::$\boldsymbol{H}^{\mathrm{TMP}}(\boldsymbol{x}, 1)$
```

a direct implementation of Eq. (6.1). Each control point costs D multiplications of B-spline weights (see Eq. (6.2)), Υ additions and Υ multiplications with the control point coefficients $\boldsymbol{c}(\boldsymbol{k})$, hence $D + 2\Upsilon$ arithmetic operations for each control point in the support region. Thus, the total number of arithmetic operations per coordinate $\boldsymbol{x}$ is given by:

$$N(D, n, \Upsilon) = (n+1)^D (D + 2\Upsilon). \tag{6.24}$$

Each control point in the recursive algorithm is a loop over $n + 1$ elements in $\mathcal{Z}^n_d$. In this loop there are 2Υ operations (Υ additions and Υ multiplications) and a call to the function

Algorithm 4 Pseudo code of TMP implementation to compute $\partial \boldsymbol{u}(\boldsymbol{x}) / \partial \boldsymbol{c}$

Please note that $\partial \boldsymbol{u}/\partial \boldsymbol{c}$ is initialized before the call to the recursive function and is passed by reference. Its elements are recursively computed and the array is finally filled in the template specialization for $d = 0$.

```
template< $d$ > class TMP
    function $\boldsymbol{J}^{\mathrm{TMP}}(\boldsymbol{x}, \kappa, v, \partial \boldsymbol{u}/\partial \boldsymbol{c})$
        for $k_d \in \mathcal{Z}^n(x_d)$ do
            TMP< $d-1$ >::$\boldsymbol{J}^{\mathrm{TMP}}\left(\boldsymbol{x}, \kappa + \alpha_d(k_d - 1), v\beta^n_d, \partial \boldsymbol{u}/\partial \boldsymbol{c}\right)$
        end for
    end function

// Template specialization for $d = 0$
template<> class TMP< 0 >
    function $\boldsymbol{J}^{\mathrm{TMP}}(\boldsymbol{x}, \kappa, v, \partial \boldsymbol{u}/\partial \boldsymbol{c})$
        for $i = 1$ to $\Upsilon$ do
            $\partial u_i/\partial c_i(\kappa) = v$
        end for
    end function

$\partial \boldsymbol{u}/\partial \boldsymbol{c} = \mathbf{0}_{\Lambda\Upsilon\Upsilon}$
TMP< $D$ >::$\boldsymbol{J}^{\mathrm{TMP}}(\boldsymbol{x}, 1, 1, \partial \boldsymbol{u}/\partial \boldsymbol{c})$
```

itself where dimension index d is decreased by one, see (Eq. (6.13)). Therefore, the number of arithmetic operations for the recursive algorithm is given by a recursive function:

$$\begin{aligned} N^R(d, n, \Upsilon) &= \begin{cases} (n+1)\left(N^R(d-1, n, \Upsilon) + 2\Upsilon\right) & d \geq 1 \\ 0 & d = 0 \end{cases}, \\ &= \Upsilon(n+1)\frac{2(n+1)^D - 2}{n}. \end{aligned} \tag{6.25}$$

The number of arithmetic operations for the remaining functions are derived in a similar manner. Table 6.1 shows the complexities of the ITK and recursive algorithms. The derivatives of the B-spline functions with respect to the control point coefficients, see the last three rows of Table 6.1, do not depend on the output dimension Υ, as in the end-case the value is copied Υ times and as this copying is not a floating point operation, it is ignored in the current complexity analysis.

6.2.7 *Image registration*

Intensity-based image registration is the problem of finding a transformation $\boldsymbol{T}(\boldsymbol{x})$ that makes the deformed moving image $M(\boldsymbol{T}(\boldsymbol{x}))$ spatially aligned to the fixed image $F(\boldsymbol{x})$ by minimizing the dissimilarity between images based on their intensities. Since the moving image intensities are only defined at discrete locations (the voxel centers), interpolation is needed to obtain the intensity at the coordinate $\boldsymbol{T}(\boldsymbol{x})$. In parametric methods, the transformation is parameterised by a model with parameters $\boldsymbol{\mu}$: $\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{T}(\boldsymbol{x}; \boldsymbol{\mu})$, where we adopted the notation of [149].

Table 6.1: Functions to compute the number of arithmetic operations as a function of Υ, D, and n. Recursive explicit shows the non-recursive formula to compute the number of arithmetic operations for the recursive algorithms, derived from the recursive implicit formulas.

	ITK	Recursive implicit	Recursive explicit
$\boldsymbol{u}$	$(n+1)^D(D+2\Upsilon)$	$N^R(d,n,\Upsilon) = \begin{cases} (n+1)\left(N^R(d-1,n,\Upsilon)+2\Upsilon\right) & d \geq 1 \\ 0 & d = 0 \end{cases}$	$\Upsilon(n+1)\,\frac{2(n+1)^D-2}{n}$
$\nabla\boldsymbol{u}$	$D(n+1)^D(D+2\Upsilon)$	$N^R(d,n,\Upsilon) = \begin{cases} (n+1)\left(N^R(d-1,n,\Upsilon)+2\Upsilon(d+1)\right) & d \geq 1 \\ 0 & d = 0 \end{cases}$	$\Upsilon(n+1)\,\frac{2(n+1)^D(2n+1)-2n(D+2)-2}{n^2}$
$\nabla^2\boldsymbol{u}$	$D\,\frac{D+1}{2}\,(n+1)^D(D+2\Upsilon)$	$N^R(d,n,\Upsilon) = \begin{cases} (n+1)\left(N^R(d-1,n,\Upsilon)+\Upsilon(d+1)(d+2)\right) & d \geq 1 \\ 0 & d = 0 \end{cases}$	$\Upsilon(n+1)\,\frac{2(n+1)^D\left(3n^2+3n+1\right)-n^2\left(D^2+5D+6\right)-n(2D+6)-2}{n^3}$
$\partial\boldsymbol{u}/\partial\boldsymbol{c}$	$D(n+1)^D$	$N^R(d,n) = \begin{cases} (n+1)\left(N^R(d-1,n)+1\right) & d \geq 1 \\ 0 & d = 0 \end{cases}$	$(n+1)\,\frac{(n+1)^D-1}{n}$
$\partial\nabla\boldsymbol{u}/\partial\boldsymbol{c}$	$D^2(n+1)^D$	$N^R(d,n,D) = \begin{cases} (n+1)\left(N^R(d-1,n,D)+D-d+2\right) & d \geq 1 \\ 0 & d = 0 \end{cases}$	$(n+1)\,\frac{(n+1)^D(Dn+n-1)-n+1}{n^2}$
$\partial\nabla^2\boldsymbol{u}/\partial\boldsymbol{c}$	$D^2\,\frac{D+1}{2}\,(n+1)^D$	$N^R(d,n,D) =$ $\begin{cases} (n+1)\left(N^R(d-1,n,D)+D-d+2+\frac{(d-D+1)(d-D+2)}{2}\right) & d \geq 1 \\ 0 & d = 0 \end{cases}$	$(n+1)\,\frac{(n+1)^D\left(n^2\left(D^2+3D+2\right)-2(Dn+n-1)\right)-2\left(n^2-n+1\right)}{2n^3}$

Registration is in this case formulated as an optimization problem in which the cost function $\mathcal{C}$ is minimized w.r.t. to the parameters $\boldsymbol{\mu}$ of the transformation:

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \mathcal{C}(\boldsymbol{\mu}), \tag{6.26}$$

In image registration, the cost function usually takes the following form:

$$\mathcal{C}\left(\boldsymbol{\mu}\right) = \Psi\left(\frac{1}{|\Omega|}\sum_{\boldsymbol{x}_i\in\Omega}\mathcal{D}\left(F(\boldsymbol{x}_i), M\left(\boldsymbol{T}\left(\boldsymbol{x}_i;\boldsymbol{\mu}\right)\right)\right)\right) + \frac{\gamma}{|\Omega|}\sum_{\boldsymbol{x}_i\in\Omega}\mathcal{R}\left(\boldsymbol{x}_i;\boldsymbol{\mu}\right), \tag{6.27}$$

where $\mathcal{D}(F, M)$ and $\Psi(\cdot)$ are continuously differentiable functions that together define the dissimilarity between F and M, e.g. the sum of squared differences or mutual information [194], [195], and Ω a discrete set of spatial coordinates in the fixed image domain [149]. $\mathcal{R}$ is a regularization term that constrains $\boldsymbol{T}\left(\boldsymbol{x};\boldsymbol{\mu}\right)$, and γ weighs similarity against regularity. Often, the regularization term is a function of the first or second order spatial derivatives of $\boldsymbol{T}$: $\nabla\boldsymbol{T}$ or $\nabla^2\boldsymbol{T}$, or a combination of the two [115, 196]. When a gradient-based optimizer is used to solve Eq. (6.26) the derivative of $\mathcal{C}$ with respect to $\boldsymbol{\mu}$ is needed:

$$\frac{\partial\mathcal{C}}{\partial\boldsymbol{\mu}} = \frac{1}{|\Omega|}\sum_{\boldsymbol{x}_i\in\Omega}\left(\frac{\partial\boldsymbol{T}}{\partial\boldsymbol{\mu}}^{\mathrm{T}}\frac{\partial M}{\partial \boldsymbol{x}}\frac{\partial\mathcal{D}}{\partial M}\frac{\partial\Psi}{\partial y} + \frac{\partial\mathcal{R}}{\partial\boldsymbol{\mu}}\right). \tag{6.28}$$

Therefore, interpolation of the moving image M and its gradient ∇M are needed. Moreover, when a regularization term is used, we also need $\partial\mathcal{R}/\partial\boldsymbol{\mu}$, which requires $\partial\nabla\boldsymbol{T}/\partial\boldsymbol{\mu}$ or $\partial\nabla^2\boldsymbol{T}/\partial\boldsymbol{\mu}$.

6.2.7.1 *Usage of B-splines in image registration*

In the case of intensity interpolation, $Y = 1$, hence $\boldsymbol{u} : \mathbb{R}^D \to \mathbb{R}$. In the experiment section we will denote intensity interpolation with the non-bold symbol u. The control point coefficients $\boldsymbol{c}(\boldsymbol{k})$ are pre-computed by filtering the image, such that when $\boldsymbol{x} = \boldsymbol{k}$ the interpolated value $u(\boldsymbol{x}) = M(\boldsymbol{x})$ [175]. The derivatives of M w.r.t. $\boldsymbol{c}$ are not necessary to solve Eq. (6.26), and are therefore not used in the field of image registration.

For non-rigid registration we use a B-spline model for $\boldsymbol{T}(\boldsymbol{x};\boldsymbol{\mu})$ as introduced in [115]. In this case $Y = D$ and therefore $\boldsymbol{u} : \mathbb{R}^D \to \mathbb{R}^D$. In the experiment section we will denote B-spline transformation with the bold symbol $\boldsymbol{u}$. The parameter vector $\boldsymbol{\mu}$ is formed by all control point coefficients $\boldsymbol{c}(\boldsymbol{k})$, which are optimized during image registration. The transformation $\boldsymbol{T}(\boldsymbol{x};\boldsymbol{\mu})$ is computed as:

$$\boldsymbol{T}(\boldsymbol{x};\boldsymbol{\mu}) = \boldsymbol{u}(\boldsymbol{x}) + \boldsymbol{x}. \tag{6.29}$$

To solve Eq. (6.26) we need $\partial\boldsymbol{T}/\partial\boldsymbol{\mu}$ and therefore $\partial\boldsymbol{u}/\partial\boldsymbol{c}$ is needed. If a regularization term is used $\partial\nabla\boldsymbol{u}/\partial\boldsymbol{c}$ and/or $\partial\nabla^2\boldsymbol{u}/\partial\boldsymbol{c}$ are also needed.

First we evaluate the accuracy of the proposed algorithms by comparing the output to those of the reference algorithms. Second, we evaluate the computation time of the intensity interpolation and coordinate transformation algorithms in separate experiments. Finally we evaluate the reduction in computation time of typical image registration experiments. We will evaluate two cases: $\Upsilon = 1$, which corresponds to image intensity interpolation, and $\Upsilon = D$ which corresponds to non-rigid coordinate transformation. For the case $\Upsilon = 1$, the derivatives w.r.t. $\boldsymbol{c}$ are not used in image registration problems (see Sec. 6.2.7). Hence, for $\Upsilon = 1$ we only evaluate the spatial derivatives.

All experiments are performed on a Dell laptop, using a single Intel® i7-3740QM core, operating Ubuntu 14.04 64-bit. The clock-rate of the CPU is 3.7 GHz and the cache memory 6 MB. All software was compiled with GCC version 4.8.4 using the -O3 compiler optimization level.

6.3.1 *Accuracy*

Due to the fact that floating point additions and multiplications are non-associative, the reordering of the summations might introduce small round-off differences between the reference and TMP-recursive algorithms. The accuracy of the proposed TMP-recursive algorithms was evaluated by computing the sum of squared differences between elements of the proposed and reference algorithms output. The control point coefficients $\boldsymbol{c}(\boldsymbol{k})$ were drawn from a $\mathcal{N}(0,1)$ distribution. The output for 10^3 random spatial coordinates was computed and the average accuracy is reported.

6.3.2 *Complexity analysis*

Using the equations in Table 6.1, we evaluate the complexities of the B-spline functions for $D \in \{2,3,4\}$ and $n \in \{1,2,3\}$ for two cases of Υ. Complexities for $\Upsilon = 1$, i.e. intensity interpolation, and for $\Upsilon = D$, i.e. coordinate transformation, are reported for the ITK and the proposed algorithms.

6.3.3 *Computation time*

The computation time of the proposed algorithms was evaluated against the ITK algorithm in three experiments:

- Intensity interpolation, i.e. $\Upsilon = 1$.
- Coordinate transformation, i.e. $\Upsilon = D$.
- Image registration in which the above mentioned cases are combined.

6.3.3.1 *Intensity interpolation*

For the evaluation of the computation time for intensity interpolation we interpolate 10^7 points. The control point coefficients $\boldsymbol{c}(\boldsymbol{k})$ were drawn from a $\mathcal{N}(0,1)$ distribution and $\Lambda = 15^D$. The following experimental parameters were varied:

- Algorithm: ITK and TMP-recursive
- Dimension: $D \in \{2,3,4\}$
- Interpolation spline order: $n \in \{1,2,3\}$
- Algorithm output: $u(\boldsymbol{x})$, $\nabla u(\boldsymbol{x})$ and $\nabla^2 u(\boldsymbol{x})$

6.3.3.2 *Coordinate transformation*

In experiment 6.3.3.2 we transform the same 10^7 points as in experiment 6.3.3.1. Please note that the actual methods for coordinate transformation in the context of image registration include the addition of $\boldsymbol{x}$ to $\boldsymbol{u}(\boldsymbol{x})$, as in Eq. (6.29), which is not taken into account in the complexity analysis, but identical for both implementations. The control point coefficients $\boldsymbol{c}$ are again drawn from a $\mathcal{N}(0,1)$ distribution and $\Lambda = 15^D$. The following experimental parameters were varied:

- Algorithm: ITK and TMP-recursive
- Dimension: $D \in \{2,3,4\}$
- Transform spline order: $n \in \{1,2,3\}$
- Algorithm output: $\boldsymbol{u}(\boldsymbol{x})$, $\nabla \boldsymbol{u}(\boldsymbol{x})$, $\nabla^2 \boldsymbol{u}(\boldsymbol{x})$, $\partial \boldsymbol{u}(\boldsymbol{x}) / \partial \boldsymbol{c}$, $\partial \nabla \boldsymbol{u}(\boldsymbol{x}) / \partial \boldsymbol{c}$, and $\partial \nabla^2 \boldsymbol{u}(\boldsymbol{x}) / \partial \boldsymbol{c}$

To measure the computation time without the use of TMP, the recursive algorithm of Algorithm 1(a) was implemented. Algorithm 1(a) uses the recursive formulation of Eq. (6.13), *but it does not use TMP*. In this experiment the same 10^7 spatial coordinates as in the previous experiments were transformed and the following parameters were varied:

- Algorithm: Recursive
- Dimension: $D \in \{2,3,4\}$
- Transform spline order: $n \in \{1,2,3\}$
- Algorithm output: $\boldsymbol{u}(\boldsymbol{x})$

Finally we measure the computation time of the image registration package NiftyReg. Please note that not all combinations of D and n are implemented in this package. The following experimental parameters were varied:

- Algorithm: NiftyReg
- Dimension, transform spline order:
  $< D, n > \in \{< 2,3 >, < 3,1 >, < 3,3 >\}$
- Algorithm output: $\boldsymbol{u}(\boldsymbol{x})$ and $\nabla \boldsymbol{u}(\boldsymbol{x})$

6.3.3.3 *Image registration*

In these experiments we analyze the computation time of typical image registration tasks in the open source image registration software Elastix, in which the usage of ITK algorithms for intensity interpolation and coordinate transformation were compared to the usage of the proposed algorithms. We analyze the computation times for images of dimensions $D \in \{2, 3, 4\}$, all of which have applications in the field of (medical) image processing. An application of 2D image registration is registration-based slice interpolation, as proposed by [197]. Image registration with 3D images is often used in the medical field. A typical 4D registration problem is when multiple 3D images are acquired to analyze the motion of an organ, e.g. the lungs [120].

For the experiments images were extracted from a publicly available dataset: the 4D CT lung dataset from the DIR-LAB database [152]. In this dataset lungs from multiple subjects were imaged as a function of time. For our experiments we selected one of those subjects (4DCT1). For the 2D registration we extracted two slices from a single time-frame, for the 3D registration we extracted two time-frames and for the 4D registration the entire dataset was used.

The goal of image registration is to solve Eq. (6.26), and in Elastix we use the adaptive stochastic gradient descent (ASGD) optimization method [149], which uses a different set of randomly selected spatial coordinates in each iteration to estimate $\mathcal{C}$ and $\partial\mathcal{C}/\partial\boldsymbol{\mu}$. In our experiments we chose to use 2000 random spatial coordinates for this estimation in the 2D and 3D registration. In the 4D registration, we use 2000 random spatial coordinates per time-frame, and since we have 10 time-frames, the total number of spatial coordinates is 20000 in this experiment. The number of resolution steps was equal to 4 and in each resolution 2000 iterations were used. In summary, to solve Eq. (6.26), there were approximately 16 million computations of $u(\boldsymbol{x})$, $\nabla u(\boldsymbol{x})$, $\boldsymbol{u}(\boldsymbol{x})$, and $\partial\boldsymbol{u}(\boldsymbol{x})/\partial\boldsymbol{c}$ per 2D and 3D registration experiment, and 160 million for the 4D registration experiment. In the experiment where a regularization term was used, there were also 16 million computations of $\nabla^2\boldsymbol{u}(\boldsymbol{x})$ and $\partial\nabla^2\boldsymbol{u}(\boldsymbol{x})/\partial\boldsymbol{c}$.

The ASGD optimizer requires parameters for estimating a stepsize to ensure a proper convergence in each resolution [149]. Since these parameters differ per application, we first estimated what the values of those parameters were in our registration experiments using the methods described in [149]. We estimated the computation time of each image registration using the predetermined estimated ASGD parameters.

We performed four image registrations. In each registration we used a different dissimilarity metric. For transforming and interpolating the moving image, the spline order for intensity interpolation, n_i, and the spline order for coordinate transformation, n_t, were set to $n_i \in \{1, 2, 3\}$ and $n_t \in \{1, 2, 3\}$. The number of control points for the intensity interpolation, Λ_i, is constant throughout the image registration and equal to the image size. The number of control points for the coordinate transformation, Λ_t, is doubled in each resolution and depends on n_t. The number of control points in the final resolution Λ_t, the image sizes Λ_i and the dissimilarity metrics $\mathcal{C}$ used in each image registration experiment are reported in Table 6.2.

Table 6.2: Settings for the image registration experiments 6.3.3.3(a) - (d).

Experiment	D	Λ_i	Λ_t		
			$n_t = 1$	$n_t = 2$	$n_t = 3$
6.3.3.3(a)	2	256×256	21×21	22×22	23×23
6.3.3.3(b)	3	256×256×94	21×21×20	22×22×21	23×23×22
6.3.3.3(c)	4	256×256×94×10	21×21×20×11	22×22×21×12	23×23×22×13
6.3.3.3(d)	3	256×256×94	21×21×20	22×22×21	23×23×22

Table 6.2: Settings for the image registration experiments 6.3.3.3(a) - (d) (continued).

Experiment	$\mathcal{C}$
6.3.3.3(a)	Mutual information [194, 195]
6.3.3.3(b)	Normalized cross-correlation
6.3.3.3(c)	Variance-based [120]
6.3.3.3(d)	Normalized cross-correlation with $\mathcal{R} = 1/D \sum_{i=1}^{D} \left\| \nabla^2 T_i \right\|_{\mathrm{fro}}^2$ [115]

6.4 RESULTS

6.4.1 *Accuracy*

The mean squared differences between the proposed and the ITK algorithms for $\boldsymbol{u}(\boldsymbol{x})$, its spatial derivatives, and its derivatives w.r.t. $\boldsymbol{c}$ for $\Upsilon = 1$ and $\Upsilon = D$ over 10^3 random spatial coordinates were $< 10^{-28}$ and therefore negligable. The mean squared difference between the proposed and the NiftyReg algorithm for $\boldsymbol{u}(\boldsymbol{x})$ and $\nabla \boldsymbol{u}(\boldsymbol{x})$ was $< 10^{-12}$ and therefore also negligable.

6.4.2 *Complexity analysis*

Tables 6.3 and 6.4 show the number of arithmetic operations for the ITK and recursive algorithms for $\Upsilon = 1$ and $\Upsilon = D$, respectively. The $\partial^2 \beta^n (x_d - k_d)/\partial x_d^2$ are zero and discontinuous for $n = 1$, hence the number of arithmetic operations for $\nabla^2 u(\boldsymbol{x})$ and $\nabla^2 \boldsymbol{u}(\boldsymbol{x})$ are not available for $n = 1$. For the case $\Upsilon = 1$, the number of operations increases for both algorithms with increasing D and n, however, the recursive algorithm has a lower number for every setting. For $D = 2$ and $n = 1$, the ratio of complexity estimates is $16/12 = 1.3$, whereas for $D = 4$ and $n = 3$ this ratio equals 2.3, indicating that especially for high dimension and high spline order the recursive algorithms are more efficient. Note that for the spatial derivatives the ratio is even higher because of the smart use of the lower order derivatives by the recursive algorithms. For example, for $D = 4$ and $n = 3$ the complexity ratio of $\nabla^2 u(\boldsymbol{x})$ is 5.6. For the case $\Upsilon = D$, see Table 6.4, the recursive algorithms show less decrease in the number of arithmetic

Table 6.3: Complexity analysis of the ITK and the recursive B-spline algorithms for $Y = 1$.

		$D = 2$			$D = 3$			$D = 4$		
		$n = 1$	$n = 2$	$n = 3$	$n = 1$	$n = 2$	$n = 3$	$n = 1$	$n = 2$	$n = 3$
$u(\boldsymbol{x})$	ITK	16	36	64	40	135	320	96	486	1536
	Recursive	12	24	40	28	78	168	60	240	680
$\nabla u(\boldsymbol{x})$	ITK	32	72	128	120	405	960	384	1944	6144
	Recursive	28	54	88	72	186	384	164	588	1576
$\nabla^2 u(\boldsymbol{x})$	ITK	-	108	192	-	810	1920	-	4860	15360
	Recursive	-	90	144	-	330	656	-	1080	2744

Table 6.4: Complexity analysis of the ITK and the recursive B-spline algorithms for $Y = D$.

		$D = 2$			$D = 3$			$D = 4$		
		$n = 1$	$n = 2$	$n = 3$	$n = 1$	$n = 2$	$n = 3$	$n = 1$	$n = 2$	$n = 3$
$\boldsymbol{u}(\boldsymbol{x})$	ITK	24	54	96	72	243	576	192	972	3072
	Recursive	24	48	80	84	234	504	240	960	2720
$\nabla \boldsymbol{u}(\boldsymbol{x})$	ITK	48	108	192	216	729	1728	768	3888	12288
	Recursive	56	108	176	216	558	1152	656	2352	6304
$\nabla^2 \boldsymbol{u}(\boldsymbol{x})$	ITK	-	162	288	-	1458	3456	-	9720	30720
	Recursive	-	180	288	-	990	1968	-	4320	10976
$\partial \boldsymbol{u}(\boldsymbol{x}) / \partial \boldsymbol{c}$	ITK	8	18	32	24	81	192	64	324	1024
	Recursive	6	12	20	14	39	84	30	120	340
$\partial \nabla \boldsymbol{u}(\boldsymbol{x}) / \partial \boldsymbol{c}$	ITK	16	36	64	72	243	576	256	1296	4096
	Recursive	16	33	56	48	141	312	128	546	1592
$\partial \nabla^2 \boldsymbol{u}(\boldsymbol{x}) / \partial \boldsymbol{c}$	ITK	-	54	96	-	486	1152	-	3240	10240
	Recursive	-	63	108	-	333	748	-	1548	4588

operations and in a few cases the recursive algorithm has even slightly more arithmetic operations, e.g. for computing $\boldsymbol{u}(\boldsymbol{x})$ for the cases $D = 3$ and $n = 1$, and $D = 4$ and $n = 1$, and for computing $\nabla\boldsymbol{u}(\boldsymbol{x})$ for the case $D = 2$ and $n = 1$.

6.4.3 *Computation time*

6.4.3.1 *Intensity interpolation*

Table 6.5 shows the computation time in seconds needed to interpolate 10^7 points with the ITK, NiftyReg, (non-TMP) recursive and the TMP-recursive algorithms. The $\beta^n(x_d - k_d)/\partial x_d^2$ are zero and discontinuous for $n = 1$, hence the computation times for $\nabla^2 u$ are not available for $n = 1$. The experiment was repeated ten times and the coefficient of variation of the reported computation times was at most 8%. For $D = 4$ and $n = 3$, the speed-up with the proposed algorithm to obtain $u(\boldsymbol{x})$ equals $11.5/2.6 = 4.4$, which is higher than the factor by which the number of arithmetic operations is decreased. This shows the effect of TMP. When obtaining $\nabla^2 u(\boldsymbol{x})$ for $D = 4$ and $n = 3$ the proposed algorithm is 15.7 times faster.

6.4.3.2 *Coordinate transformation*

Table 6.6 shows the computation time needed to transform 10^7 points with the ITK, the NiftyReg (for the available cases), and the TMP-recursive algorithms. The $\partial^2\beta^n(x_d - k_d)/\partial x_d^2$ are zero and discontinuous for $n = 1$, hence the computation times for $\nabla^2\boldsymbol{u}(\boldsymbol{x})$ and $\partial\nabla^2\boldsymbol{u}(\boldsymbol{x})/\partial\boldsymbol{c}$ are not available for $n = 1$. For most algorithm outputs the proposed TMP-recursive algorithm is the fastest implementation, where highest speed-up compared to the reference algorithm ITK is obtained for $n = 2$ and $D = 4$: a factor of 17.9 to compute $\boldsymbol{u}(\boldsymbol{x})$. The NiftyReg algorithm was faster than ITK for computing $\boldsymbol{u}(\boldsymbol{x})$. For the case $D = 3$, $n = 1$, NiftyReg was also faster than TMP-recursive. For all other cases TMP-recursive was fastest.

The second row of Table 6.6 shows the effect of the recursive implementation without TMP for computing $\boldsymbol{u}(\boldsymbol{x})$. Despite the number of arithmetic operations of the recursive algorithm being only slightly less than the ITK algorithm, the recursive implementation was still faster than the ITK. This may be explained by the presence of non-floating point operations in the ITK algorithm, e.g. iterator updating and loop control instructions, which are not present in the recursive algorithm.

6.4.3.3 *Image registration*

Table 6.7 shows the computation time in seconds for image registration tasks for images of dimensions $D \in \{2, 3, 4\}$, transform spline orders $n_t \in \{1, 2, 3\}$ and interpolation spline orders $n_i \in \{1, 2, 3\}$. Note that in experiment 6.2.7(c) we have 10 times as many computations of $u(\boldsymbol{x})$, $\nabla u(\boldsymbol{x})$, $\boldsymbol{u}(\boldsymbol{x})$, and $\partial\boldsymbol{u}(\boldsymbol{x})/\partial\boldsymbol{c}$ than in the other experiments, explaining the much higher computation time of the image registration for $D = 4$. For $D = 2$, the registration was on average 1.5 times faster when the proposed algorithms are used, $\sim$2 times faster for $D = 3$,

Table 6.5: Results of experiment 6.3.3.1: computation time (s) for $u(\boldsymbol{x})$ and its derivatives of 10^7 random points.

		$D=2$			$D=3$			$D=4$		
		$n=1$	$n=2$	$n=3$	$n=1$	$n=2$	$n=3$	$n=1$	$n=2$	$n=3$
$u(\boldsymbol{x})$	ITK	0.3	0.4	0.6	0.5	1.1	2.4	1.0	4.0	11.5
	TMP-recursive	0.2	0.3	0.4	0.3	0.5	0.9	0.5	1.2	2.6
$\nabla u(\boldsymbol{x})$	ITK	0.5	0.7	1.1	1.2	3.4	7.3	3.9	16.4	50.6
	TMP-recursive	0.3	0.4	0.7	0.5	0.8	1.6	0.9	2.0	5.0
$\nabla^2 u(\boldsymbol{x})$	ITK	-	1.4	2.2	-	8.0	18.1	-	48.1	149.3
	TMP-recursive	-	0.7	0.9	-	1.7	2.9	-	3.6	9.5

Table 6.6: Results of experiment 6.3.3.2: computation time (s) of $\boldsymbol{u}(\boldsymbol{x})$ and its derivatives of 10^7 random points.

		$D=2$			$D=3$			$D=4$		
		$n=1$	$n=2$	$n=3$	$n=1$	$n=2$	$n=3$	$n=1$	$n=2$	$n=3$
$\boldsymbol{u}(\boldsymbol{x})$	ITK	1.5	2.0	2.6	5.6	11.2	18.9	17.7	55.4	132.4
	Recursive	0.5	0.8	1.1	1.3	3.1	6.2	3.1	10.8	29.2
	TMP-recursive	0.2	0.3	0.4	0.5	0.8	1.4	1.2	3.1	7.5
	NiftyReg	-	-	1.0	0.3	-	3.7	-	-	-
$\nabla \boldsymbol{u}(\boldsymbol{x})$	ITK	1.8	2.6	3.6	6.4	13.9	26.8	20.5	71.2	178.2
	TMP-recursive	0.3	0.4	0.6	0.9	1.5	2.7	2.8	6.2	14.0
	NiftyReg	-	-	2.5	-	-	8.3	-	-	-
$\nabla^2 \boldsymbol{u}(\boldsymbol{x})$	ITK	-	3.0	4.2	-	19.0	37.4	-	99.6	264.4
	TMP-recursive	-	0.8	1.1	-	3.4	5.5	-	12.7	29.1
$\partial \boldsymbol{u}(\boldsymbol{x})/\partial \boldsymbol{c}$	ITK	0.5	0.6	0.8	0.7	1.4	2.6	7.1	34.4	106.9
	TMP-recursive	0.3	0.4	0.5	0.5	1.0	1.8	1.0	2.7	7.0
$\partial \nabla \boldsymbol{u}(\boldsymbol{x})/\partial \boldsymbol{c}$	ITK	0.9	1.5	2.0	2.0	4.1	8.6	10.2	54.9	194.0
	TMP-recursive	0.4	0.5	0.8	0.8	1.6	3.2	1.6	6.4	23.9
$\partial \nabla^2 \boldsymbol{u}(\boldsymbol{x})/\partial \boldsymbol{c}$	ITK	-	3.2	4.5	-	12.9	25.0	-	86.2	280.4
	TMP-recursive	-	0.7	1.0	-	3.0	10.2	-	30.2	91.7

and ∼3.5 times faster for $D = 4$. This shows that the proposed algorithms become more efficient with increasing D. The $\partial^2\beta^n(x_d - k_d)/\partial x_d^2$ are zero and discontinuous for $n = 1$, hence the computation times for the registration with regularization are not available for $n_t = 1$.

Table 6.7: Results of experiments 6.3.3.3(a) - 6.3.3.3(d): computation time (s) of the image registrations.

Experiment	D		$n_t = 1$		
			$n_i = 1$	$n_i = 2$	$n_i = 3$
6.3.3.3(a)	2	ITK	14	15	17
		TMP-recursive	10	10	11
6.3.3.3(b)	3	ITK	26	32	43
		TMP-recursive	16	19	23
6.3.3.3(c)	4	ITK	700	994	1769
		TMP-recursive	258	316	442
6.3.3.3(d) regularized	3	ITK	-	-	-
		TMP-recursive	-	-	-
Experiment	**D**		**$n_t = 2$**		
			$n_i = 1$	$n_i = 2$	$n_i = 3$
6.3.3.3(a)	2	ITK	17	18	19
		TMP-recursive	10	12	12
6.3.3.3(b)	3	ITK	41	46	56
		TMP-recursive	20	24	27
6.3.3.3(c)	4	ITK	2015	2300	3081
		TMP-recursive	604	670	789
6.3.3.3(d) regularized	3	ITK	114	120	131
		TMP-recursive	53	57	61
Experiment	**D**		**$n_t = 3$**		
			$n_i = 1$	$n_i = 2$	$n_i = 3$
6.3.3.3(a)	2	ITK	21	22	22
		TMP-recursive	12	12	13
6.3.3.3(b)	3	ITK	60	66	77
		TMP-recursive	29	32	36
6.3.3.3(c)	4	ITK	5113	5445	6159
		TMP-recursive	1468	1541	1664
6.3.3.3(d) regularized	3	ITK	203	210	222
		TMP-recursive	96	100	103

6.5 DISCUSSION

Accelerating (non-rigid) image registration algorithms is necessary in a variety of fields. There is a need for fast real-time registration in e.g. image-guided interventions [188], radiotherapy [198, 199], multi-view range image registration [200], etc. When the data is very large, efficient algorithms are necessary to complete the task in a reasonable time, for example in reconstruction of high resolution microscopy images [201], or the construction of large population template images [113]. We therefore expect the proposed algorithms to have a large impact in these fields.

We showed that the proposed algorithms have mostly fewer and sometimes slightly more arithmetic operations, but, thanks to the implementation with TMP, which was made possible by the recursive formulation, the proposed algorithms are faster than the reference algorithms. For increasing dimension and spline order the proposed algorithms are progressively more efficient.

When comparing the computation times of the proposed algorithms in Table 6.5 ($Y = 1$) with Table 6.6 ($Y = D$) a larger difference in computation time of the TMP-recursive algorithms might be expected since coordinate transformation has D times as much arithmetic operations as intensity interpolation. However, the expected difference is smaller, especially for low D and low n, possibly due to constant overhead.

The TMP-recursive algorithm was faster than the NiftyReg implementation for all cases but one: for $D = 3$ and $n = 1$. Part of the computation time differences is explained by differences in precomputation of e.g. the B-spline weights and the support region. A limitation of the NiftyReg implementations is that they only exist for specific combinations of D and n, as opposed to the proposed algorithms.

The work of [187], which proposes to precompute the B-spline weights with a look-up table, can easily be combined with the algorithms we propose, to even further speed-up the B-spline algorithms. However, due to rounding errors in that method, the accuracy will be somewhat reduced.

The image registration experiments with Elastix show a large speed-up with the proposed algorithms, from a factor 1.5 to a factor 3.5, depending on the image dimension. The intensity interpolation and coordinate transformation make up a large part of the computation time of an image registration, but not all. Another part that is computationally expensive is performing the gradient descent step of the optimizer, which is especially expensive when $\boldsymbol{\mu}$ is very high-dimensional.

The separability property was already shown in [150], however we showed that using this separability, the B-spline interpolation function can be reformulated into a recursive function. We extended this recursive formulation to the derivatives of the B-spline functions with respect to spatial coordinates and control point coefficients. Recursion in especially the spatial derivatives showed to be efficient, due to the re-use of the lower-order derivatives.

The exact mechanisms that explain the reduction in computation time achieved with TMP may depend on the compiler used, or even the used compilation settings. These mechanisms could e.g. be a reduction in the number of jumps in the code, more efficient access of the cache memory or a reduction in mean memory latency. Which of these mechanisms is actually being used by the compiler can potentially be learned from observing the intermediate representation or the assembly code. Investigation of the compiler dependency of the proposed methods was not part of this work, but may be done in the future.

In this paper we showed that the proposed algorithms substantially reduce the computation time for CPU implementations. In principle the proposed algorithms could be ported to GPUs. To benefit from their parallel processing architecture, individual points to be transformed or interpolated could be processed by separate threads. As each thread iterates over an identical support region size and as there are no data-dependent switches, thread divergence [202] is not expected to be a problem. However, it probably would lead to very scattered memory access and subsequently suboptimal performance [188]. This may be remedied by prefetching the support region of each point separately. An in-depth evaluation of portability to GPU would be an interesting topic for future research.

6.6 CONCLUSION

We proposed novel recursive algorithms for B-spline functions and provide efficient implementations that use template metaprogramming. We derived recursive formulations for the B-spline function and its derivatives with respect to spatial coordinates and control point coefficients. For increasing dimension the proposed algorithm showed to be substantially faster, up to a factor of 4 for intensity interpolation and a factor of 18 for coordinate transformation. The recursive formulation enabled reducing the algorithmic complexity by re-using intermediate results of lower-order derivatives. With TMP, the compiler exploits this reduced algorithmic complexity to create efficient code. Image processing software can use the recursive B-spline functions (with open source implementation) proposed in this paper to reduce computation times for image (re)sampling and transformation. We showed that image registration with the public Elastix software can be approximately two to four times faster when the proposed algorithms are used. Since B-spline functions are widely used in image processing and may comprise a large part of the computation time, we believe that this work is of importance for accelerating image processing algorithms in general.

6.7 FUNDING

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007 – 2013) under grant agreement no. 601055, VPH-DARE@IT and from the Netherlands Organization for Health Research and Development (ZonMW, IMDI program), grant no. 104003012.

7

GENERAL DISCUSSION

As neurodegeneration due to disease may be difficult to distinguish from that of normal aging, this thesis aimed to develop methods that can assess how normal aging affects the brain morphology, to obtain insight in the normal aging process and to support clinical decision making. The first part of this thesis aimed to evaluate and propose methods for clinical decision support using MR brain imaging features. In the second part, two novel techniques were presented that allow performing groupwise image registration on large imaging datasets, and that were used in the first part of this thesis. The following sections discuss the main findings, methodological contributions and future perspectives.

7.1 MAIN FINDINGS

Clinicians often face the challenge to interpret all the data obtained from a single patient when making diagnostic or prognostic decisions. This process would be supported if more objective quantitative information would be extracted from imaging data and its diagnostic and prognostic value would be established, along with other known risk factors for disease. Therefore, there is a large interest in computer-based methods that support clinical decision making. Specifically for the computer-aided diagnosis of dementia and its subtypes several methods have been previously proposed and evaluated [22, 23, 203]. In this thesis the previously proposed DSI [22], normative modeling, and a novel model of the aging brain were explored for the purpose of supporting clinical decision making for neurodegenerative diseases. In the next subsections the main findings of this thesis are discussed.

7.1.1 *Predicting cognitive decline*

The aim of Chapter 2 was to assess if the DSI [22] could be used to identify persons at risk for global cognitive decline, as this identification may aid in early detection of risk at dementia. Global cognition was assessed at two time-points, with an average follow-up time of 5.7 years, providing the possibility to define a measure for cognitive decline. In this study, a person was diagnosed as a significant global cognitive decliner when he or she belonged to the five percent of participants with the highest cognitive decline per year. With feature selection, the performance of the DSI was assessed on several feature sets. We however found that age only was the most important predictor. Other features, such as MR brain imaging features, cognitive test results at baseline, cardiovascular risk factors, genetics, gait, and education, showed having potential, but did not improve prediction performance when combined with age. A next step could be to use longitudinal features in DSI, as this might improve its prediction performance. To validate whether our findings are not due to limitations of DSI, also other methods need to be evaluated in this prediction challenge. Finally, to be able to detect younger people at risk of significant global cognitive decline in future studies, thresholds for cognitive decline should be carefully chosen depending on the population, for example be age-adjusted.

7.1.2 *Comparing automated brain region segmentation methods*

For the segmentation of MR brain images into anatomical regions, numerous fully automated methods have been proposed, e.g. [2, 3, 6–10, 13–15, 19, 67]. Since manual segmentation is currently the gold reference standard, segmentations resulting from some of these methods have been compared to manual segmentations. However, the impact of using different segmentation methods on the analyses of individual patients when using a normative modeling framework was unknown. To assess this impact, five state-of-the-art automated brain segmentation methods were applied to non-demented (ND) subjects and compared on the volumes of six brain regions. For most regions the correlation was good (> 0.75) indicating that the observed volume differences between methods in ND subjects are mainly due to systematic differences. The methods were also applied to brain images of 42 patients with Alzheimer's disease (AD), with the aim to answer two research questions: 1) to what extent are methods interchangeable, as long as the same method is being used for generating normative volume distributions and patient volumes? and 2) can different methods be used for generating normative volume distributions and calculating patient-specific volumes? Based on the results we concluded that it is essential that the normative data and the patient-specific data are segmented with the same method. When the same method is being used for generating normative and patient data, we found that the agreement on the AD patient's position relative to the normative distribution (ICC-z) was high for the regions thalamus and putamen. Our results are encouraging as they indicate that methods are to some extent interchangeable for selected regions. For the regions hippocampus, amygdala, caudate nucleus and accumbens, and globus pallidus, not all method combinations showed a high ICC-z. Whether two methods are indeed interchangeable should be confirmed for the specific application and dataset of interest.

7.1.3 *Aging brain morphology: ND versus AD*

In Chapter 4, a method was proposed to facilitate comparing the morphology of an entire individual brain to a normative morphology distribution. The distribution of the aging morphology of AD patients was compared with that of ND subjects. The morphology distribution of the ND subjects was on average lower than the individual morphology scores of AD patients, indicating that morphology differences between AD patients and ND subjects can be partly explained by accelerated aging. We showed that this method has potential for the detection of accelerated brain aging in a clinical setting. Furthermore, the proposed framework opens up new possibilities to study the effect of other determinants, such as lifestyle factors, on brain aging.

7.2 METHODOLOGICAL CONTRIBUTIONS

Three novel methods have been developed and evaluated in this thesis. The method proposed in Chapter 4 builds a reference model of the entire brain as a function of age, i.e. a spatio-temporal reference model, to which an individual brain morphology can be compared. The model was built using 988 brain images of non-demented subjects. With the groupwise image registration method presented in Chapter 5, deformation fields were computed, containing the displacement from a common template space to the space of each subject of every voxel for each of the 988 brain images. With partial least squares regression, a high-dimensional deformation space was reduced to obtain a single morphology score and a corresponding main mode of age induced deformation. A normative morphology distribution was obtained by fitting percentile curves as a function of age on these scores. The score can be interpreted using the main mode of deformation as shown on the developed web application `https://www.agingbrain.nl`. Here, the spatio-temporal atlas based on 988 brain images is made publicly available.

The method developed in Chapter 5 enables groupwise image registration of quantitative MR imaging datasets. This method simultaneously registers the images to a common space, to obtain the necessary anatomical correspondence between images. Due to the large contrast differences between the acquired images, the registration based on intensities is generally challenging. However, the proposed method performed as good as, or better than the current standard method for such datasets. As shown in Chapter 4, this method is also suitable for inter-subject groupwise image registrations of images with modest contrast differences.

The algorithms presented in Chapter 6 contributed to the computational efficiency of medical image registration. Novel recursive formulations of B-spline algorithms were derived, which are often used in image registration for interpolation and transformation of voxels in the images. The recursive algorithms on itself already improved efficency, but in combination with an implementation using template metaprogramming, they significantly reduced the computation time compared to reference algorithms. The methods presented in Chapter 5 and Chapter 6 have been made publicly available in the image registration software package elastix [119], available at `https://github.com/SuperElastix/elastix`.

7.3 FUTURE PERSPECTIVES

I expect improvements can be made in advanced image analysis for modeling the aging brain in three areas of model development, explained in more detail below: 1) the training data, 2) the methodology, and 3) the computation. Additionally, I propose future research on the heritability of brain aging.

7.3.1 *Training data*

The models for clinical decision support presented in this thesis all used cross-sectional data. However, features derived from longitudinal data might be better suited to separate controls and patients, which may improve the relevance of other features, such as imaging features, on the prediction of cognitive decline. In the context of Chapter 2, longitudinal data could improve the prediction of cognitive decline. In the context of Chapter 3, volume change between two time points instead of absolute volumes at baseline might be more (or less) reproducible across segmentation methods. In the context of Chapter 4, computing the brain morphology score at multiple time points could provide insight into the pace of brain aging in individual patients.

7.3.2 *Methodology*

Instead of looking at specific regions, the method presented in Chapter 4 of this thesis uses whole-brain regression. It was however shown that in neurodegenerative diseases regional atrophy may occur in a specific order, and this order may vary per disease type [204]. To estimate this order, and thereby estimating the disease progression, event-based models were previously proposed [204, 205]. To investigate if neurodegeneration due to aging also occurs in a specified order, event-based models could be an interesting direction for future research to improve modeling the aging brain.

The method developed in Chapter 4 could be applied to other voxelwise maps than deformation maps, such as diffusion or perfusion imaging derived maps, to investigate age-related changes of these tissue properties. This could possibly improve discriminative ability between people at risk of accelerated brain aging and people without risk. Besides applying this method to voxelwise maps, it can also be applied to selected points on the surface of a segmented anatomical structure of interest, for example the hippocampus. In that case, a spatio-temporal model and its variation in the population of that specific structure can be studied in more detail and individual anatomical structures can be compared to a reference shape distribution.

Deep learning is a specific category of machine learning in which the model is built using abstract features, learned by the algorithm itself, in multiple layers. The success of this type of machine learning is due to the fact that the abstract features work generally well for object detection in images. In the field of medical image analysis it is therefore widely used for object detection, but also in other areas, such as automated segmentation, image registration, and computer-aided diagnosis [206, 207]. Using deep learning for predicting cognitive decline, as an alternative approach for the DSI that was used in Chapter 2, could be an interesting approach. Fully connected networks, such as the multilayer perceptron are suitable when the input also contains non-imaging data. Extending Chapter 3 with segmentation methods that use deep learning would be interesting. The bias of the atlas is not eliminated with deep learning. However, previous methods showed promising results [13, 14] and may lead to higher segmentation accuracy, which could further increase the diagnostic value of brain region volumes in a normative modelling framework.

7.3.3 *Computation*

Although some effort was being made for computational efficiency, the computation time of the method in Chapter 4 is still too high for practical use. However, computational efficiency can be gained by re-designing the groupwise image registration algorithm. Especially on the graphical processing unit (GPU) a high efficiency can be obtained due to the large number of computations it can handle simultaneously. I therefore recommend extending the novel algorithms presented in Chapter 5 and 6 to implementations on the GPU. Also, it would be interesting to study the impact of using simplified and more efficient registration approaches for comparing patient data to the spatio-temporal model of Chapter 4. At model construction stage, computation time is not an issue so the full groupwise approach can be preserved, but for comparing new patient data to the model, the method should be accelerated.

Deep learning could also be used to increase computational efficiency of image registration, as computationally expensive iterative optimization algorithms are replaced by the application of previously learned deep networks [208–211].

7.3.4 *Imaging genetics*

Another interesting area of future research is imaging genetics in which imaging biomarkers are combined with genetic data to investigate genes that are expressed in the brain [212]. By combining genetic data with the morphology score as developed in this thesis, the heritability of this score may be estimated. This could provide insight in to what extent the pace of brain aging is controlled by genes.

7.4 CONCLUSION

This thesis aimed to develop and evaluate novel methods for clinical decision support, based on advanced, quantitative analysis of brain MR images. The research objectives stated in the introduction have been achieved.

1. The accuracy of predicting global cognitive decline in the general population using a multivariate classification framework was evaluated. In this study, age only was the most important predictor. Other features showed having potential, but did not improve prediction performance.
2. The impact of differences in automated brain region segmentation methods on single-subject analysis in a normative modelling framework was evaluated. The results indicated that methods are to some extent interchangeable for selected regions.

3. A novel spatio-temporal approach for extracting and modelling the brain morphology changes due to normal aging was developed and evaluated. With this method an individual brain morphology can be compared to a morphology distribution of a reference population.
4. A novel method for intra-subject non-rigid groupwise registration of multiple images with contrast differences was developed. The method performed as good as, or better than the current standard method for such datasets.
5. A highly efficient algorithm for B-spline interpolation and transformation was developed. This algorithm led to a substantial acceleration of non-rigid image registration methods.

The novel aging brain models and computer-aided diagnosis methods presented in this thesis support the differentiation between normal and abnormal neurodegeneration. This will help in establishing more accurate diagnoses of patients, and in identifying patients at risk of developing neurodegenerative disease before symptoms emerge. In the future, performance and efficacy of the developed methods should be evaluated in clinical practice.

SUMMARY

Both normal aging and neurodegenerative diseases such as Alzheimer's disease (AD) cause morphological changes of the brain due to neurodegeneration. As neurodegeneration due to disease may be difficult to distinguish from that of normal aging, interpretation of magnetic resonance (MR) brain images in the context of diagnosis of neurodegenerative diseases is challenging, especially in the early stages of the disease. This thesis presented comprehensive models of the aging brain and novel computer-aided diagnosis methods, based on advanced, quantitative analysis of brain MR images, facilitating the differentiation between normal and abnormal neurodegeneration.

CLINICAL DECISION SUPPORT USING MR BRAIN IMAGING

The first part of this thesis, comprising Chapters 2, 3 and 4, aimed to evaluate and develop methods for clinical decision support using features derived from MR brain images.

In **Chapter 2** the performance of a previously proposed classifier, the Disease State Index (DSI), was evaluated for the prediction global cognitive decline in the general population, because identification of persons at risk for global cognitive decline may aid in early detection of persons at risk of dementia. The best prediction performance was obtained using only age as input feature. Other features showed potential, but did not improve prediction when used in combination with age. Future studies should evaluate whether the performance could be improved when using new features, e.g., longitudinal features, or using other prediction methods.

In **Chapter 3** the impact on single-subject analysis when using different automated subcortical brain region segmentation methods for the generation of normative data was evaluated. We applied five methods on 988 non-demented (ND) subjects and computed the correlation (PCC-v) and absolute agreement (ICC-v) on the volumes of six brain regions. For most regions the PCC-v was good (> 0.75) indicating that volume differences between methods in ND subjects are mainly due to systematic differences. The ICC-v was generally lower, especially for the smaller regions, indicating that it is essential that the same method is used to generate normative and patient data. To evaluate the impact on single subject analysis we also applied the methods to 42 patients with Alzheimer's disease (AD). In the case where the normative distributions and the patient-specific volumes were calculated by the same method, the patient's distance to the normative distribution was assessed with the z-score. We determined the diagnostic value of this z-score, which showed to be consistent across the methods. We also determined the absolute agreement on the AD patient z-scores (ICC-z). We found that the ICC-z was high for the regions thalamus and putamen. Our results are encouraging as they indicate that methods are to some extent interchangeable for selected regions. For the

regions hippocampus, amygdala, caudate nucleus and accumbens, and globus pallidus, not all method combinations showed a high ICC-z. Whether two methods are indeed interchangeable should be confirmed for the specific application and dataset of interest.

In **Chapter 4** a method was proposed for developing a spatio-temporal model of morphological differences in the brain due to normal aging. This model facilitates making the distinction between neurodegeneration due to normal aging and neurodegenerative diseases. The proposed method utilized groupwise image registration to characterize morphological variation across brain scans of people with different ages. A normative distribution of brain morphologies was estimated, which can be used to compare an individual's brain morphology to a cognitively healthy reference population. To derive a normative distribution of brain morphologies, a data-driven approach was chosen which required a computationally efficient non-rigid groupwise image registration technique. The method was validated on two different datasets, using images from both cognitively normal subjects and patients with AD. The distribution of the aging morphology of AD subjects was compared with that of cognitive normals and these results indicated that AD at least partially manifests as accelerated aging.

EFFICIENT NON-RIGID GROUPWISE IMAGE REGISTRATION

In the second part two novel techniques were developed and evaluated that allow performing non-rigid groupwise image registration on large imaging datasets.

In **Chapter 5**, a novel technique for groupwise image registration was developed, initially for quantitative MR imaging datasets, in which the contrast between the images to register varies. The developed method showed to be as good as, or better than the current standard method for such datasets. The contrast variation between the brain images used to derive the model of Chapter 4 was small, but it was not absent, and therefore the method of Chapter 5 showed to be effective for a groupwise image registration between brain images of multiple subjects.

In **Chapter 6** recursive formulations of algorithms for B-spline interpolation and transformation were derived. In combination with template metaprogramming, the novel algorithms showed a decrease of the computation time of both interpolation and transformation significantly compared to reference algorithms.

CONCLUSION

The novel aging brain models and computer-aided diagnosis methods presented in this thesis facilitate the differentiation between normal and abnormal neurodegeneration. This will help in establishing more accurate diagnoses of patients, and in identifying patients at risk of developing neurodegenerative disease before symptoms emerge. In the future, the computational efficiency of the developed methods should be improved such that the performance and efficacy can be evaluated in clinical practice.

SAMENVATTING

Zowel normale veroudering als neurodegeneratieve ziektes zoals de ziekte van Alzheimer zorgen voor morfologische veranderingen in het brein door neurodegeneratie. Omdat neurodegeneratie door ziekte moeilijk te onderscheiden is van dat door normale veroudering, is interpretatie van magnetische resonantie (MR) beelden van het brein in de context van diagnose van neurodegeneratieve ziekten een uitdaging, vooral in vroege stadia van de ziekte. Dit proefschrift presenteert uitgebreide modellen van het verouderende brein en nieuwe computerondersteunende diagnostische methodes, gebaseerd op geavanceerde, kwantitatieve analyse van MR breinbeelden om het maken van onderscheid tussen normale en abnormale neurodegeneratie te vereenvoudigen.

ONDERSTEUNING VAN KLINISCHE BESLUITVORMING MET MR BREINBEELDVORMING

Het eerste deel van dit proefschrift, bestaande uit Hoofdstuk 2, 3 en 4, heeft als doel om methodes voor klinische besluitvorming te ontwikkelen en evalueren, gebruik makend van kenmerken die uit MR breinbeelden geëxtraheerd worden.

In **Hoofdstuk 2** is de nauwkeurigheid van een eerder voorgestelde klassificatiemethode, de Disease State Index (DSI), geëvalueerd op het voorspellen van globale cognitieve achteruitgang in de algemene populatie, omdat de identificatie van verhoogd risico op cognitieve achteruitgang de detectie van een verhoogd risico op dementie kan ondersteunen. De beste voorspelling werd behaald wanneer enkel de leeftijd van de persoon werd meegenomen als kenmerk. Andere kenmerken hadden voorspellende waarde, maar verhoogden niet de nauwkeurigheid van de voorspellingen wanneer deze werden gecombineerd met leeftijd. Toekomstige studies moeten evalueren of de nauwkeurigheid kan worden verbeterd door nieuwe kenmerken, bijvoorbeeld kenmerken op basis van longitudinale data, of andere predictiemethoden te gebruiken.

In **Hoofdstuk 3** worden verschillende methodes voor het automatisch segmenteren van sub-corticale breinstructuren vergeleken. Specifiek wordt onderzocht wat de impact is op de score van individuele patiënten als verschillende methodes worden gebruikt voor het genereren van normatieve verdelingen, waarbij de score een maat is voor hoeveel een individu afwijkt van de normatieve verdeling. Hiertoe werden vijf methodes toegepast op 988 niet-dementerende (ND) subjecten en de correlatie (PCC-v) en de absolute overeenstemming (ICC-v) bepaald op zes breinstructuren. Voor de meeste regio's was de PCC-v goed (> 0.75), wat betekent dat de geobserveerde volumeverschillen tussen de ND subjecten vooral ontstaan door systematische verschillen. De ICC-v was over het algemeen lager, vooral voor de kleinere regio's. Dit betekent dat het essentieel is dat dezelfde methode wordt gebruikt voor het genereren van de normatieve data en de patiëntdata. De methodes werden ook toegepast op 42 patiënten met de ziekte van Alzheimer (AD) om de impact op de score van individuele patiënten te onderzoeken. In het geval dat de normatieve verdelingen en de patiënt-specifieke volumes werden berekend door dezelfde methode, werd de afstand van de patiënt tot de normatieve verdeling bepaald met de z-score. De diagnostische waarde van de z-score was consistent over de methodes. De absolute overeenstemming van de z-scores (ICC-z) was hoog voor de regio's thalamus en putamen. Deze resultaten zijn bemoedigend, omdat ze aantonen dat de methodes tot op zekere hoogte uitwisselbaar zijn voor bepaalde regio's. Voor regio's hippocampus, amygdala, caudate nucleus en accumbens, en globus pallidus waren niet alle combinaties van de ICC-z hoog. Of twee methodes daadwerkelijk uitwisselbaar zijn moet worden bepaald per dataset en toepassing.

In **Hoofdstuk 4** is een methode voorgesteld waarmee de vormverandering van het brein ten gevolge van normale veroudering kan worden bepaald. Het spatio-temporeel model in deze methode helpt om onderscheid te maken tussen neurodegeneratie door normale veroudering en door neurodegeneratieve ziektes. De voorgestelde methode maakt gebruik van groepsgewijze beeldregistratie om de morfologische variatie tussen breinbeelden van mensen van verschillende leeftijden te karakteriseren. Hieruit werd een normatieve verdeling van breinmorfologieën afgeleid welke kan worden gebruikt om de morfologie van een individueel brein te vergelijken met die van een gezonde referentiepopulatie. Er is gekozen voor een datagedreven methode om de normatieve verdeling van breinmorfologieën te bepalen. Hiervoor was een efficiënte niet-rigide groepsgewijze beeldregistratie techniek nodig. De methode is gevalideerd op twee datasets, waarbij zowel beelden van cognitief gezonde personen als patiënten met de ziekte van Alzheimer zijn gebruikt. De distributie van de morfologie van de patiënten met de ziekte van Alzheimer is vergeleken met die van cognitief gezonde personen. De morphologische verschillen tussen patiënten en cognitief gezonde personen laten zien dat de ziekte van Alzheimer zich ten minste voor een deel manifesteert als versnelde veroudering.

EFFICIËNTE NIET-RIGIDE GROEPSGEWIJZE BEELDREGISTRATIE

In het tweede deel van dit proefschrift werden twee nieuwe technieken ontwikkeld en geëvalueerd waarmee niet-rigide groepsgewijze beeldregistratie kan worden toegepast op grote datasets van beelden.

In **Hoofdstuk 5**, is een nieuwe techniek voor groepsgewijze beeldregistratie ontwikkeld, in beginsel voor kwantitatieve MR beelden waar het contrast tussen de beelden die moeten worden geregistreerd variëert. Dit is een uitdaging, omdat methodes die dit soort beelden registreren vaak gebruik maken van de aanname dat de voxelintensiteit over de beelden sterk correleert. Dit is niet het geval voor kwantitatieve MRI beelden en daarom maakt de ontwikkelde methode gebruik van het feit dat de voxelintensiteit afhankelijk is van een kwantitatief MR model. De methode maakt gebruik van de (lage) dimensionaliteit van het model, maar is onafhankelijk van de exacte formulering en daarom breed toepasbaar voor allerlei kwantitatieve

MR beeldregistratieproblemen. De ontwikkelde methode was even goed als of beter dan de huidige standaard voor zulke datasets. De variatie in contrast tussen de breinbeelden die zijn gebruikt om het model in Hoofdstuk 4 te maken was klein, maar niet afwezig. Daarom is de methode in Hoofdstuk 5 ook effectief gebleken voor de groepsgewijze beeldregistratie van breinbeelden van verschillende personen.

In **Hoofdstuk 6** zijn recursieve algoritmes voor interpolatie en transformatie met B-splines afgeleid. De recursieve formulering vermindert de complexiteit van het algoritme door een verlaging van het aantal rekenkundige operaties. Daarnaast maakt de recursieve formulering het mogelijk om de algoritmes te implementeren met template metaprogramming, een techniek waarbij de overhead tijdens gebruik van het uitvoerbare bestand wordt verminderd. Hierdoor zijn de nieuwe algoritmes significant sneller dan de referentie algoritmes waarmee ze zijn vergeleken.

CONCLUSIE

Met de nieuwe modellen van het verouderende brein en de computerondersteunde methodes voor diagnose die zijn ontwikkeld en gepresenteerd in dit proefschrift wordt het maken van onderscheid tussen normale en abnormale neurodegeneratie vereenvoudigd. Dit zal helpen bij het stellen van een nauwkeurigere diagnose bij patiënten en het vroeg identificeren van patiënten die het risico lopen om neurodegeneratieve ziektes te ontwikkelen, voordat symptomen zichtbaar zijn. In de toekomst moet de efficiëntie van de ontwikkelde methodes worden verbeterd zodat de prestatie en effectiviteit geëvalueerd kan worden in de klinische praktijk.

BIBLIOGRAPHY

1. J.B. Brewer. Fully-automated volumetric MRI with normative ranges: Translation to clinical practice. *Behavioural Neurology* **21**, 21–28 (2009).
2. 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 & A.M. Dale. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355 (2002).
3. K.O. Babalola, T.F. Cootes, C.J. Twining, V. Petrovic & C.J. Taylor. *3D brain segmentation using active appearance models and local regressors* in *Medical Image Computing and Computer-Assisted Interventions - MICCAI 2008* (eds D. Metaxas, L. Axel, G. Fichtinger & G. Székely) **Lecture Notes in Computer Science, vol 5241** (Springer-Verlag, Berlin, Heidelberg, 2008), 401–408.
4. B. Patenaude, S.M. Smith, D.N. Kennedy & M. Jenkinson. A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage* **56**, 907–922 (2011).
5. F. Wenzel, C. Meyer, T. Stehle, J. Peters, S. Siemonsen, C. Thaler, L. Zagorchev & for the Alzheimer's Disease Neuroimaging Initiative. Rapid fully automatic segmentation of subcortical brain structures by shape-constrained surface adaptation. *Medical Image Analysis* **In Press, Accepted Manuscript** (2018).
6. E.E. Bron, R. Steketee, G.C. Houston, R. Oliver, H.C. Achterberg, M. Loog, J.C. van Swieten, A. Hammers, W.J. Niessen, M. Smits & S. Klein. Diagnostic classification of arterial spin labeling and structural MRI in presenile early stage dementia. *Human Brain Mapping* **35**, 4916–4931 (2014).
7. M.J. Cardoso, M. Modat, R. Wolz, A. Melbourne, D. Cash, D. Rueckert & S. Ourselin. Geodesic Information Flows: Spatially-Variant Graphs and Their Application to Segmentation and Fusion. *IEEE Transactions on Medical Imaging* **34**, 1976–1988 (2015).
8. C. Ledig, R.A. Heckemann, A. Hammers, J.C. Lopez, V.F.J. Newcombe, A. Makropoulos, J. Lötjönen, D. Menon & D. Rueckert. Robust whole-brain segmentation: Application to traumatic brain injury. *Medical Image Analysis* **21**, 40–58 (2015).
9. S. Murphy, B. Mohr, Y. Fushimi, H. Yamagata & I. Poole. *Fast, Simple, Accurate Multi-atlas Segmentation of the Brain* in *Biomedical Image Registration. WBIR 2014* (eds S. Ourselin & M. Modat) **Lecture Notes in Computer Science, vol. 8545** (Springer, Cham, 2014), 1–10.
10. Wang, J., C. Vachet, A. Rumple, S Gouttard, C. Ouziel, E. Perrot, G. Du, X. Huang, G. Gerig & M. Styner. Multi-atlas segmentation of subcortical brain structures via the AutoSeg software pipeline. *Frontiers in Neuroinformatics* **8**, 7 (2014).
11. R. Wolz, P. Aljabar, J.V. Hajnal, A. Hammers & D. Rueckert. LEAP: Learning embeddings for atlas propagation. *NeuroImage* **49**, 1316–1325 (2010).
12. F. van der Lijn, T. den Heijer, M.M. Breteler & W.J. Niessen. Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. *NeuroImage* **43**, 708–720 (2008).
13. S. Bao & A.C.S. Chung. Multi-scale structured CNN with label consistency for brain MR image segmentation. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* **6**, 113–117 (2018).

14. M. Shakeri, S. Tsogkas, E. Ferrante, S. Lippe, S. Kadoury, N. Paragios & Kokkinos, I. *Sub-cortical brain structure segmentation using F-CNN's* in *2016 IEEE 13th International Symposium onBiomedical Imaging (ISBI)* (IEEE, 2016), 269–272.
15. De Brébisson, A. & Montana, G. *Deep Neural Networks for Anatomical Brain Segmentation* in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (IEEE, 2015), 20–28.
16. A. Hammers, R.S.N. Liu, O. Colliot, J. Burdett, E. Bardinet, J.S. Duncan, L. Garnero & L. Lemieux. Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. *NeuroImage* **46**, 749–761 (2009).
17. J.J. Corso, Z. Tu, A. Yuille & A. Toga. *Segmentation of sub-cortical structures by the graph-shifts algorithm* in (eds N. Karssemeijer & B. Lelieveldt) **Lecture Notes in Computer Science, vol 4584** (Springer, Berlin, Heidelberg, 2007), 183–197.
18. J.H. Morra, Z. Tue, L.G. Apostolova, A.E. Green, C. Avedissian, S.K. Madsen, N.Parikshak, X. Hua, A.W. Toga, C.R. Jack, M.W. Weiner, P.M.Thompson & the Alzheimer's Disease Neuroimaging Initiative. Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. *NeuroImage* **43**, 59–68 (2008).
19. Z. Tue, K.L. Narr, P. Dollar, I. Dinov, P.M. Thompson & A.W. Toga. Brain anatomical structure segmentation by hybrid discriminative/generative models. *IEEE Transactions on Medical Imaging* **27**, 495–508 (2008).
20. F.P.M. Oliveira & J.M.R.S. Tavares. Medical image registration: a review. *Computer Methods in Biomechanics and Biomedical Engineering* **17**, 73–93 (2014).
21. A. Sotiras, C. Davatzikos & N. Paragios. Deformable Medical Image Registration: A Survey. *IEEE Transactions on Medical Imaging* **32**, 1153–1190 (2013).
22. J. Mattila, J. Koikkalainen, A. Virkki, A. Simonsen, M. van Gils, G. Waldemar, H. Soininen, J. Lötjönen & the Alzheimer's Disease Neuroimaging Initiative. A disease state fingerprint for evalulation of Alzheimer's disease. *Journal of Alzheimer's Disease* **27**, 162–176 (2011).
23. E.E. Bron, M. Smits, W.M. van der Flier, H. Vrenken, F. Barkhof, P. Scheltens, J.M. Papma, R.M.E. Steketee, C.P. Méndez Orellana, R. Meijboom, M. Pinto, J.R. Meireles, C. Garrett, A.J. Bastos-Leite, A. Abdukadir, O. Ronneberger, N. Amoroso, R. Bellotti, D. Cárdenas-Pe na, A.M. Álveraz-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. 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 & S. Klein. Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge. *NeuroImage* **111**, 562–579 (2015).
24. R. Cuingnet, Gerardin, E., J. Tessieras, G. Auzias, S. Leh'ericy, M.O. Habert, M. Chupin, H. Benali & O. Colliot. Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *NeuroImage* **56**, 766–781 (2011).
25. D.M.J. Tax. *One-class classification: concept-learning in the absence of counter-examples* phd (Delft University of Technology, 2001).
26. C.R. Jack, M.A. Bernstein, N.C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P.J. Britson, J.L. Whitwell, C. Ward, A.M. Dale, J.P. Felmlee, J.L. Gunter, D.L.G. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C.S. DeCarli, G. Krueger, H.A. Ward, G.J. Metzger, K.T. Scott, R. Mallozzi, D. Blezek, J. Levy, J.P. Debbins, A.S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler & M.W. Weiner. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging* **27**, 685–691 (2008).

27. S. Kloppel, 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 & R.S. Frackowiak. Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. *Brain* **131**, 2969–2974 (2008).

28. J. Mattila, J. Koikkalainen, A. Virkki, A. Simonsen, M. van Gils, G. Waldemar, H. Soininen, J. Lötjönen & the Alzheimer's Disease Neuroimaging Initiative. A disease state fingerprint for evaluation of Alzheimer's disease. *Journal of Alzheimer's Disease* **27**, 163–176 (2011).

29. J. Mattila, H. Soininen, J. Koikkalainen, D. Rueckert, R. Wolz, G. Waldemar & J. Lötjönen. Optimizing the diagnosis of early Alzheimer's disease in mild cognitive impairment subjects. *Journal of Alzheimer's Disease* **32**, 969–979 (2012).

30. A. Hall, J. Mattila, J. Koikkalainen, J. Lötjönen, R. Wolz, P.H. Scheltens, G.B. Frisoni, M. Tsolaki, F. Nobili, Y. Freund-Levi, L. Minthon, L. Frölich, H.J. Hampel, P.T. Visser & H. Soininen. Predicting progression from cognitive impairment to Alzheimer's disease with the Disease State Index. *Current Alzheimer Research* **12**, 69–79 (2015).

31. M.A. Muñoz-Ruiz, A. Hall, J. Mattila, J. Koikkalainen, S.K. Herukka, R. Vanninen, Y. Liu, J. Lötjönen, H, H. S. & the Alzheimer's Disease Neuroimaging Initiative. Comparing predictors of conversion to Alzheimer's disease using the disease state index. *Neurodegenerative Diseases* **13**, 200–202 (2014).

32. T. Pekkala, A. Hall, J. Lötjönen, J. Mattila, H. Soininen, T. Ngandu, T. Laatikainen, M. Kivipelto & A. Solomon. Development of a Late-Life Dementia Prediction Index with Supervised Machine Learning in the Population-Based CAIDE Study. *Journal of Alzheimer's Disease* **55**, 1055–1067 (2017).

33. J.A. Blumenthal, P.J. Smith, S. Mabe, A. Hinderliter, K. Welsh-Bohmer, J.N. Browndyke, P.H. Lin, W. Kraus, P.M. Doraiswamy, J. Burke & A. Sherwood. Lifestyle and Neurocognition in Older Adults with Cardiovascular Risk Factors and Cognitive Impairment. *Psychosomatic Medicine* **79**, 719–727 (2017).

34. E.P. Moll van Charante, E. Richard, L.S. Eurelings, J.W. van Dalen, S.A. Ligthart, E.F. van Bussel, M.P. Hoevenaar-Blom, M. Vermeulen & W.A. van Gool. Effectiveness of a 6-year multidomain vascular care intervention to prevent dementia (preDIVA): a cluster-randomised controlled trial. *Lancet* **388**, 797–805 (2016).

35. T. Ngandu, J. Lehtisalo, A. Solomon, E. Levälahti, S. Ahtiluoto, R. Antikainen, L. Bäckman, T. Hänninen, A. Jula, T. Laatikainen, J. Lindström, F. Mangialasche, T. Paajanen, S. Pajala, M. Peltonen, R. Rauramaa, A. Stigsdotter-Neely, T. Strandberg, J. Tuomilehto, H. Soininen & M. Kivipelto. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. *Lancet* **385**, 2255–2263 (2015).

36. J. Mattila, J. Koikkalainen, A. Virkki, M. van Gils, J. Lötjönen & for the Alzheimer's Disease Neuroimaging Initiative. Design and Application of a Generic Clinical Decision Support System for Multiscale Data. *IEEE Transactions on Biomedical Engineering* **59**, 234–240 (2012).

37. M.A. Ikram, A. van der Lugt, W.J. Niessen, P.J. Koudstaal, G.P. Krestin, A. Hofman, D. Bos & M.W. Vernooij. The Rotterdam Scan Study: design update 2016 and main findings. *European Journal of Epidemiology* **30**, 1299–1315 (2015).

38. M.A. Ikram, A. van der Lugt, W.J. Niessen, P.J. Koudstaal, G.P. Krestin, A. Hofman, D. Bos & M.W. Vernooij. The Rotterdam Scan Study: design update 2016 and main findings. *European Journal of Epidemiology* **30**, 1299–1315 (2015).

39. A. Hofman, G.G. Brusselle, S.D. Murad, C.M. van Duijn, O.H. Franco, A. Goedegebure, M.A. Ikram, T.E.C. Nijsten, C. K. T., R.P. Peeters, B.H. Stricker, H.W. Tiemeier, A.G. Uitterlinden & M.W. Vernooij. The Rotterdam Study: 2016 objectives and design update. *European Journal of Epidemiology* **30**, 661–708 (2015).

40. H.A. Vrooman, C.A. Cocosco, F. van der Lijn, R. Stokking, M.A. Ikram, M.W. Vernooij, M.M. Breteler & W.J. Niessen. Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification. *NeuroImage* **37**, 71–81 (2007).

41. R. de Boer, H.A. Vrooman, F. van der Lijn, M.W. Vernooij, M.A. Ikram, A. van der Lugt, M.M. Breteler & W.J. Niessen. White matter lesion extension to automatic brain tissue segmentation on MRI. *NeuroImage* **45**, 1151–1161 (2009).

42. A.M. Dale, B. Fischl & M.I. Sereno. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage* **9**, 179–194 (1999).

43. R.S. Desikan, F. Ségonne, B. Fischl, B.T. Quinn, B.C. Dickerson, D. Blacker, R.L. Buckner, A.M. Dale, R.P. Maguire, B.T. Hyman, M.S. Albert & R.J. Killiany. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006).

44. M.W. Vernooij, A. van der Lugt, M.A. Ikram, P.A. Wielopolski, H.A. Vrooman, A. Hofman, G.P. Krestin & M.M. Breteler. Total cerebral blood flow and total brain perfusion in the general population: the Rotterdam Scan Study. *Journal of Cerebral Blood Flow & Metabolism* **28**, 412–419 (2008).

45. G. Roob, R. Schmidt, P. Kapeller, A. Lechner, H.P. Hartung & F. Fazekas. MRI evidence of past cerebral microbleeds in a healthy elderly population. *Neurology* **52**, 991–994 (1999).

46. M.W. Vernooij, A. van der Lugt, M.A. Ikram, P.A. Wielopolski, W.J. Niessen, A. Hofman, G.P. Krestin & M.M. Breteler. Prevalence and risk factors of cerebral microbleeds: the Rotterdam Scan Study. *Neurology* **70**, 1208–1214 (2008).

47. M. de Groot, M.A. Ikram, S. Akoudad, G.P. Krestin, A. Hofman, A. van der Lugt, W.J. Niessen & M.W. Vernooij. Tract-specific white matter degeneration in aging: the Rotterdam Study. *Alzheimer's & Dementia* **11**, 321–330 (2015).

48. S. Akoudad, M.L. Portegies, P.J. Koudstaal, A. Hofman, A. van der Lugt, M.A. Ikram & M.W. Vernooij. Cerebral Microbleeds Are Associated With an Increased Risk of Stroke: The Rotterdam Study. *Circulation* **132**, 509–516 (2015).

49. P.R. Wenham, W.H. Price & G. Blandell. Apolipoprotein E genotyping by one-stage PCR. *Lancet* **337**, 1158–1159 (1991).

50. L. Lahousse, V.J. Verlinden, J.N. van der Geest, G.F. Joos, A. Hofman, B.H. Stricker, G.G. Brusselle & M.A. Ikram. Gait patterns in COPD: the Rotterdam Study. *European Respiratory Journal* **46**, 88–95 (2015).

51. V.J. Verlinden, J.N. van der Geest, Y.Y. Hoogendam, A. Hofman, M.M. Breteler & M.A. Ikram. Gait patterns in a community-dwelling population aged 50 years and older. *Gait Posture* **37**, 500–505 (2013).

52. M.L. Bleecker, K. Bolla-Wilson, J. Agnew & D.A. Meyers. Age-related sex differences in verbal memory. *Journal of Clinical Psychology* **44**, 403–411 (1988).

53. I. Goethals, K. Audenaert, F. Jacobs, E. Lannoo, C. Van de Wiele, H. Ham, A. Otte, K. Oostra & R. Dierckx. Cognitive neuroactivation using SPECT and the Stroop Colored Word Test in patients with diffuse brain injury. *Journal of Neurotrauma* **21**, 1059–1069 (2004).

54. C.J. Golden. Identification of brain disorders by the Stroop Color and Word Test. *Journal of Clinical Psychology* **32**, 654–658 (1976).

55. M.D. Lezak. Neuropsychological assessment in behavioral toxicology–developing techniques and interpretative issues. *Scandinavian Journal of Work, Environment & Health* **10**, 25–29 (1984).

56. K.A. Welsh, N. Butters, R.C. Mohs, D. Beekly, S. Edland, G. Fillenbaum & A. Heyman. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part V. A normative study of the neuropsychological battery. *Neurology* **44**, 609–614 (1994).

57. J. Desrosiers, R. Hébert, G. Bravo & E. Dutil. The Purdue Pegboard Test: normative data for people aged 60 and over. *Disability and Rehabilitation* **17**, 217–224 (1995).

58. Y.Y. Hoogendam, A. Hofman, J.N. van der Geest, A. van der Lugt & M.A. Ikram. Patterns of cognitive function in aging: the Rotterdam Study. *European Journal of Epidemiology* **29**, 133–140 (2014).

59. C. Nadeau & Y. Bengio. Inference for the Generalization Error. *Machine Learning* **52**, 239–281 (2003).

60. F. Falahati, D. Ferreira, H. Soininen, P. Mecocci, B. Vellas, M. Tsolaki, I. Kłoszewska, S. Lovestone, M. Eriksdotter, L.-O. Wahlund, A. Simmons, E. Westman, for the AddNeuroMed consortium & the Alzheimerś Disease Neuroimaging Initiative. The effect of age correction on multivariate classification in Alzheimerś disease, with a focus on the characteristics of incorrectly and correctly classified subjects. *Brain Topography* **29**, 296–307 (2016).

61. J. Koikkalainen, H. Pölönen, J. Mattila, M. van Gils, H. Soininen, J. Lötjönen & the Alzheimer's Disease Neuroimaging Initiative. Improved Classification of Alzheimer's Disease Data via Removal of Nuisance Variability. *PLoS One* **7**, e31112 (2012).

62. M. Rusanen, M. Kivipelto, E. Levälahti, T. Laatikainen, J. Tuomilehto, H. Soininen & T. Ngandu. Heart diseases and long-term risk of dementia and Alzheimer's disease: a population-based CAIDE study. *Journal of Alzheimer's disease* **42**, 183–191 (2014).

63. M. Kivipelto, T. Ngandu, T. Laatikainen, B. Winblad, H. Soininen & J. Tuomilehto. Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. *Lancet Neurology* **5**, 735–741 (2006).

64. B.C. Stephan, C. Tzourio, S. Auriacombe, H. Amieva, C. Dufouil, A. Alpérovitch & T. Kurth. Usefulness of data from magnetic resonance imaging to improve prediction of dementia: population based cohort study. *The British Medical Journal* **350**, h2863 (2015).

65. L.C. Löwe, C. Gaser, K. Franke & for the Alzheimer's Disease Neuroimaging Initiative. The Effect of the APOE Genotype on Individual BrainAGE in Normal Aging, Mild Cognitive Impairment, and Alzheimer's Disease. *PLoS One* **11**, e0157514 (2016).

66. L.W. Bonham, E.G. Geier, C.C. Fan, J.K. Leong, L. Besser, W.A. Kukull, J. Kornak, O.A. Andreassen, G.D. Schellenberg, H.J. Rosen, W.P. Dillon, C.P. Hess, B.L. Miller, A.M. Dale, R.S. Desikan & J.S. Yokoyama. Age-dependent effects of APOEϵ4 in preclinical Alzheimer's disease. *Annals of Clinical and Translational Neurology* **3**, 668–677 (2016).

67. L. Zagorchev, C. Meyer, T. Stehle, F. Wenzel, S. Young, J. Peters, J. Weese, K. Paulsen, M. Garlinghouse, J. Ford, R. Roth, L. Flashman & T. McAllister. Differences in Regional Brain Volumes Two Months and One Year after Mild Traumatic Brain Injury. *Journal of Neurotrauma* **33**, 29–34 (2015).

68. P. Scheltens, N. Fox, F. Barkhof & C. De Carli. Structural magnetic resonance imaging in the practical assessment of dementia: beyond exclusion. *Lancet Neurology* **1**, 13–21 (2002).

69. A. Convit, M.J. De Leon, C. Tarshish, S. De Shanti, W. Tsui, H. Rusinek & A. George. Specific hippocampal volume reductions in individuals at risk for Alzheimer's disease. *Neurobiology of Aging* **18**, 131–138 (1997).

70. T. den Heijer, M.I. Geerlings, F.E. Hoebeek, A. Hofman, P.J. Koudstaal & M.M. Breteler. Use of hippocampal and amygdalar volumes on magnetic resonance imaging to predict dementia in cognitively intact elderly people. *Archives of General Psychiatry* **63**, 57–62 (2006).

71. G. Ziegler, G.R. Ridgway, R. Dahnke, C. Gaser & for the Alzheimer's Disease Neuroimaging Initiative. Individualized Gaussian process-based prediction and detection of local and global gray matter abnormalities in elderly subjects. *NeuroImage* **97**, 333–348 (2014).

72. A.F. Marquand, I. Rezek, J. Buitelaar & C.F. Beckmann. Understanding heterogeneity in clinical cohort using normative models: beyond case-control studies. *Biological Psychiatry* **80**, 552–561 (2016).

73. M. de Onis, A. Onyango, E. Borghi, A. Siyam, A. Pinol & et al. *WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-forheight and body mass index-for-age: methods and development* tech. rep. (WHO Department of Health and Nutricion, 2006).

74. O. Grimm, S. Pohlack, R. Cacciaglia, T. Winkelmann, M.M. Plichta, T. Demirakca & H. Flor. Amygdalar and hippcampal volume: A comparison between manual segmentation, Freesurfer and VBM. *Journal of Neuroscience methods* **253**, 254–261 (2015).

75. R.A. Morey, C.M. Petty, Y. Xu, J.P. Hayes, H.R. Wagner, D.V. Lewis, K.S. LaBar, M. Styner & G. McCarthy. A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *NeuroImage* **45**, 855–866 (2009).

76. K.O. Babalola, B. Patenaude, P. Aljebar, J. Schnabel, D. Kennedy, W. Crun, S. Smith, T.F. Cootes, M. Jenkinson & D. Rueckert. *Comparison and Evaluation of Segmentation Techniques for Subcortical Structures in Brain MRI* in *Medical Image Computing and Computer-Assisted Interventions - MICCAI 2008* (eds D. Metaxas, L. Axel, G. Fichtinger & G. Székely) **Lecture Notes in Comupter Science, vol 5241** (Springer-Verlag, Berlin, Heidelberg, 2008), 409–416.

77. P. Aljabar, R. Heckemann, A. Hammers, J. Hajnal & D. Rueckert. *Classifier selection strategies for label fusion using large atlas databases* in *Medical Image Computing and Computer-Assisted Interventions - MICCAI 2007* (eds N. Ayache, S. Ourselin & A. Maeder) **Lecture Notes in Comupter Science, vol 4791** (Springer-Verlag, Berlin, Heidelberg, 2007), 523–531.

78. G. Perlaki, R. Horvath, S.A. Nagy, P. Bogner, T .Doczi, J. Janszky & G. Orsi. Comparison of accuracy between FSL's FIRST and Freesurfer for caudate nucleus and putamen segmentation. *Scientific Reports* **7**, 2418 (2017).

79. M.A. Ikram, der Lugt, A., W.J. Niessen, P.J. Koudstaal, G.P. Krestin, A. Hofman, D. Bos & M.W. Vernooij. The Rotterdam Scan Study: design update 2016 and main findings. *European Journal of Epidemiology* **30**, 1299–1315 (2015).

80. W. Huizinga, D.H.J. Poot, M.W. Vernooij, G.V. Roshchupkin, E.E. Bron, M.A. Ikram, D. Rueckert, W.J. Niessen, S. Klein & the Alzheimer's Disease Neuroimaging Initiative. A spatio-temporal reference model of the aging brain. *NeuroImage* **169**, 11–22 (2018).

81. J. Ashburner & K.J. Friston. Unified segmentation. *NeuroImage* **26**, 839–851 (2005).

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

83. A. Hammers, R. Allom, M.J. Koepp, S.L. Free, R. Myers, L. Lemieux, T.N. Mitchell, D.J. Brooks & J.S. Duncan. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Human Brain Mapping* **19**, 224–247 (2003).

84. N.J. Tustison, B.B. Avants, P.A. Cook, Y. Zheng, A. Egan, P.A. Yushkevich & J.C. Gee. N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging* **29**, 1310–1320 (2010).

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

86. D.S. Marcus, T.H. Wang, J. Parker, J.G. Csernansky, J.C. Morris & 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**, 1498–1507 (2007).

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

88. R.A. Heckemann, C. Ledig, K.R. Gray, P. Aljabar, D. Rueckert, J.V. Hajnal & A. Hammers. Brain Extraction Using Label Propagation and Group Agreement: Pincram. *PLoS One* **10**, e0129211 (2015).

89. K. Van Leemput, F. Maes, D. Vandermeulen & P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging* **18**, 897–908 (1999).

90. T.F. Cootes, C.J. Taylor, D.H. Cooper & J. Graham. *Training models of shape from sets of examples* in *BMVC92* (Springer, 1992), 9–18.

91. T. Cole & P. Green. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat. Med.* **11**, 1305–1319 (1991).

92. I. Yeo & R. Johnson. A new family of power transformations to improve normality or symmetry. *Biometrika* **87**, 954–959 (2000).

93. T.W. Yee. The VGAM Package for Categorical Data Analysis. *Journal of Statistical Software* **32**, 1–34 (2010).

94. K.O. McGraw & S.P. Wong. Forming inferences about some intraclass correlation coefficients. *Phsychological methods* **1**, 30–46 (1996).

95. M.M. Mukaka. A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal* **24**, 69–71 (2012).

96. S.M. Smith. Fast robust automated brain extraction. *Human Brain Mapping* **17**, 143–155 (2002).

97. M.W. Vernooij & M. Smits. Structural neuroimaging in aging and Alzheimer's disease. *Neuroimaging Clinics of North America* **22**, 33–55 (2012).

98. B.C. Davis, P.T. Fletcher, E. Bullitt & S. Joshi. Population shape regression from random design data. *International Journal of Computer Vision* **90**, 255–266 (2010).

99. H.C. Achterberg, F. van der Lijn, T. den Heijer, A. van der Lugt, M.B. Breteler, W.J. Niessen & M. de Bruijne. *Prediction of dementia by hippocampal shape analysis* in *MICCAI 2010: Medical Image Computing and Computer-Assisted Intervention, Machine Learning in Medical Imaging* **LNCS, vol. 6357** (Springer, 2010), 23–30.

100. S.G. Costafreda, I.D. Dinov, Z. Tu, Y. Shi, C.Y. Liu, I. Kloszewska, P. Mecocci, H. Soininen, M. Tsolaki, B. Vellas, L.O. Wahlund, C. Spenger, A.W. Toga, S. Lovestone & A. Simmons. Automated hippocampal shape analysis predicts the onset of dementia in mild cognitive impairment. *Neuroimage* **56**, 212–219 (2011).

101. A. Serag, P. Aljebar, G. Ball, S.J. Counsell, J.P. Boardman, M. Rutherford, D. Edwards, J.V. Hajnal & D. Rueckert. Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. *NeuroImage* **59**, 2255–2265 (2012).

102. E. Dittrich, T.R. Raviv, G. Kasprian, R. Donner, P.C. Brugger, D. Prayer & G. Langs. A spatio-temporal latent atlas for semi-supervised learning of fetal brain segmentations and morphological age estimation. *Medical Image Analysis* **18**, 9–21 (2014).

103. J. Fishbaugh, S. Durrleman, M. Prastawa & G. Gerig. Geodesic shape regression with multiple geometries and sparse parameters. *Medical Image Analysis* **39**, 1–17 (2017).

104. G. Ziegler, R. Dhanke & C. Gaser. Models of the aging brain structure and individual decline. *Frontiers in Neuroinformatics* **6**, 1–16 (2012).

105. A.F. Marquand, I. Rezek, J. Buitelaar & C.F. Beckmann. Understanding heterogeneity in clinical cohort using normative models: beyond case-control studies. *Biological Psychiatry* **80**, 552–561 (2016).

106. G. Ziegler, G.R. Ridgway, R. Dahnke, C. Gaser & for The Alzheimer's Disease Neuroimaging Initiative. Individualized Gaussian process-based prediction and detection of local and global gray matter abnormalities in elderly subjects. *NeuroImage* **97**, 333–348 (2014).

107. S. Baloch & C. Davatzikos. Morphological appearance manifolds in computational anatomy: Groupwise registration and morphological analysis. *NeuroImage* **45**, 573–585 (2009).

108. S. Wold, M. Sjöström & L. Eriksson. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **58**, 109–130 (2001).

109. Jong, S. SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems* **18**, 251–263 (1992).

110. A. Krishnan, L.J Williams, A.R. McIntosh & H. Abdi. Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review. *NeuroImage* **56**, 455–475 (2011).

111. G. Ziegler, R. Dahnke, A.D. Winkler & C. Gaser. Partial least squares correlation of multivariate cognitive abilities and local brain structure in children and adolescents. *NeuroImage* **82**, 284–294 (2013).

112. Singh, N., P.T. Fletcher, J.S. Preston, R.D. King, J.S. Marronb, M.W. Weinerc, S. Joshi & Alzheimer's Disease Neuroimaging Initiative (ADNI). Quantifying anatomical shape variations in neurological disorders. *Medical Image Analysis* **18**, 616–633 (2014).

113. W. Huizinga, D.H.J. Poot, G. Roschchupkin, E.E. Bron, M.A. Ikram, M.W. Vernooij, D. Rueckert, W.J. Niessen & S. Klein. *Modeling the brain morphology distribution in the general aging population* in *Proc. SPIE 9788, Medical Imaging 2016: Biomedical Applications in Molecular, Structural, and Functional Imaging* **97880I** (2016), 9788–7.

114. J. Mazziotta, A. Toga, A. Evans, Fox, P., J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike, C. Holmes, L. Collins, P. Thompson, D. MacDonald, M. Iacoboni, T. Schormann, K. Amunts, N. Palomero-Gallagher, S. Geyer, L. Parsons, K. Narr, N. Kabani, G. Le Goualher, D. Boomsma, T. Cannon, R. Kawashima. & B. Mazoyer. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosofical transactions of the Royal Society of London* **356**, 1293–1322 (2001).

115. D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach & D.J. Hawkes. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging* **18**, 712–721 (1999).

116. K.K. Bhatia, J.V. Hajnal, B.K. Puri, A.D. Edwards & D. Rueckert. *Consistent groupwise non-rigid registration for atlas construction* in *Proc. IEEE Int Symp on Biomed Imaging: Nano to Macro* (2004), 908–911.

117. S. Balci, P. Golland, M. Shenton & M.W. Wells. *Free-form B-Spline Deformation Model for Groupwise Registration* in *Proceedings of Medical Image Computing and Compututer-Assisted Intervention (MICCAI)* Statistical Registration Workshop: Pair-wise and Group-wise Alignment and Atlas Formation Part I (Springer, 2007), 23–30.

118. W. Huizinga, D.H.J. Poot, J.-M. Guyader, R. Klaassen, B.F. Coolen, M. van Kranenburg, R.J.M. van Geuns, A. Uitterdijk, M. Polfliet, J. Vandemeulebroucke, A. Leemans, W.J. Niessen & S. Klein. PCA-based groupwise image registration for quantitative MRI. *Medical Image Analysis* **29**, 65–78 (2016).

119. S. Klein, M. Staring, K. Murphy, M.A. Viergever & J.P.W. Pluim. elastix: a toolbox for intensity based medical image registration. *IEEE Transactions on Medical Imaging* **29**, 196–205 (2010).

120. Metz, C., Klein, S., Schaap, M., van Walsum, T. & Niessen, W. Nonrigid registration of dynamic medical imaging data using nD+t B-splines and a groupwise optimization approach. *Medical Image Analysis* **15**, 238–249 (2011).

121. A. Höskuldsson. PLS regression methods. *Journal of Chemometrics* **2**, 211–228 (1988).

122. S. Wiklund, D. Nilsson, L. Eroksson, M. Sjöström, S. Wold & K. Faber. A randomization test for PLS component selection. *Journal of Chemometrics* **21**, 427–439 (2007).

123. T. Cole & P. Green. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat. Med.* **11**, 1305–1319 (1991).

124. I. Yeo & R. Johnson. A new family of power transformations to improve normality or symmetry. *Biometrika* **87**, 954–959 (2000).

125. T.W. Yee. The VGAM Package for Categorical Data Analysis. *Journal of Statistical Software* **32**, 1–34 (2010).

126. J. Carpenter & J. Bithell. Bootstrap confidence intervals: when, which, what? *Statistics in medicine* **19**, 1141–1164 (2000).

127. E.M. Schrijvers, B.F. Verhaaren, P.J. Koudstaal, A. Hofman, M.A. Ikram & M.M. Breteler. Is dementia incidence declining?: Trends in dementia incidence since 1990 in the Rotterdam Study. *Neurology* **78**, 1456–1163 (2012).

128. L. Le Folgoc, H. Delingette, A. Criminisi & N. Ayache. Quantifying Registration Uncertainty with Sparse Bayesian Modelling. *IEEE Transactions on Medical Imaging* **36**, 607–617 (2016).

129. H. Sokooti, G. Saygili, B. Glocker, B.P.F. Lelieveldt & M. Staring. *Accuracy Estimation for Medical Image Registration Using Regression Forests* in *MICCAI 2016: Medical Image Computing and Computer-Assisted Intervention* **LNCS, vol. 9902** (Springer, 2016), 107–115.

130. J. Kybic. Bootstrap Resampling for Image Registration Uncertainty Estimation Without Ground Truth. *IEEE Transactions on Image Processing* **19**, 64–73 (2010).

131. P. Tofts. *Quantitative MRI of the Brain: Measuring Changes Caused by Disease* (John Wiley and Sons, 2003).

132. Mangin, J.-F., Poupona, C., Clark, C., Bihana, D. L. & Bloch, I. Distortion Correction and Robust Tensor Estimation for MR Diffusion Imaging. *Medical Image Analysis* **6**, 191–198 (2002).

133. Guyader, J.-M., Bernardin, L., Douglas, N., Poot, D., Niessen, W. & Klein, S. Influence of image registration on apparent diffusion coefficient images computed from free-breathing diffusion MR images of the abdomen. *Journal of Magnetic Resonance Imaging* **42**, 315–330 (2014).

134. Xue, H., Shah, S., Greiser, A., Guetter, C., Littmann, A., Jolly, M., Arai, A., Zuehlsdorff, S., Guehring, J. & Kellman, P. Motion Correction for Myocardial T1 Mapping Using Image Registration with Synthetic Image Estimation. *Magnetic Resonance in Medicine* **67**, 1644–1655 (2012).

135. G.A. Buonaccorsi, J.P.B. O'Connor, A. Caunce, C. Roberts, S. Cheung, Y. Watson, K. Davies, L. Hope, A. Jackson, G.C. Jayson & G.J.M. Parker. Tracer kinetic model-driven registration for dynamic contrast enhanced MRI time series data. *Magnetic Resonance in Medicine* **5**, 1010–1019 (2007).

136. Andersson, J. & Skare, S. A Model-Based Method for Retrospective Correction of Geometric Distortions in Diffusion-Weighted EPI. *NeuroImage* **16**, 177–199 (2002).

137. Hallack, A., Chappell, M., Gooding, M. & Schnabel, J. *A New Similarity Metric for Groupwise Registration of Variable Flip Angle Sequences for Improved* T_{10} *Estimation in DCE-MRI* in *Biomedical Image Registration: WBIR14, 6th International Workshop on Biomedical Image Registration* (eds S. Ourselin & M. Modat) **LNCS, vol. 8545** (Springer, Cham, 2014), 154–163.

138. M. Bhushan, J.A. Schnabel, L. Risser, M.P. Heinrich, J.M. Brady & M. Jenkinson. *Motion Correction and Parameter Estimation in dceMRI Sequences: Application to Colorectal Cancer* in *MICCAI 2011* **LNCS, vol. 6891** (2011), 476–483.

139. Hamy, V., Dikaios, N., Punwani, S., Melbourne, A., Latifoltojar, A., Makanyanga, J., Chouhan, M., Helbren, E., Menys, A., Taylor, S. & Atkinson, D. Respiratory motion correction in dynamic MRI using robust data decomposition registration – Application to DCE-MRI. *Medical Image Analysis* **18**, 301–313 (2014).

140. Melbourne, A., Atkinson, D., White, M., Collins, D., Leach, M. & Hawkes, D. Registration of dynamic contrast-enhanced MRI using a progressive principal component registration (PPCR). *Physics in Medicine and Biology* **52**, 5147–5156 (2007).

141. T.E.J. Behrens, H.J. Berg, S. Jbabdi, M.F.S. Rushworth & M.W. Woolrich. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage* **34**, 144–155 (2007).

142. B. Jeurissen, A. Leemans, J.D. Tournier, D.K. Jones & J. Sijbers. Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Human Brain Mapping* **34**, 2747–2766 (11 2013).

143. X. Liu, M. Niethammer, R. Kwitt, M. Mcormick & S. Aylward. *Low-Rank to the Rescue - Atlas-based Analysis in the Presence of Pathologies* in *MICCAI 2014* **LNCS, vol. 8673** (2014).

144. Miller, E., Matsakis, N. & Viola, P. *Learning from one example through shared densities on transforms* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* **1** (2000), 464–471.

145. C. Wachinger & N. Navab. Simultaneous registration of multiple images: similarity metrics and efficient optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**, 1221–1233 (2013).

146. D.R. Messroghli, A. Radjenovic, S. Kozerke, D.M. Higgins, M.U. Sivananthan & J.P. Ridgway. Modified Look-Locker Inversion Recovery (MOLLI) for High-Resolution T1 Mapping of the Heart. *Magnetic Resonance in Medicine* **52**, 141–146 (2004).

147. Y. Peng, A. Ganesh, J. Wright & Y. Ma. *RASL: Robust alignment by Sparse and Low-rank Decomposition for Linearly Correlated Images* in *IEEE conference on Computer Vision and Pattern Recognition* (2010).

148. N.P. van der Aa, H.G. ter Morsche & R.R.M. Mattheij. Computation of eigenvalue and eigenvector derivatives for a general complex-valued eigensystem. *ELA. The Electronic Journal of Linear Algebra [electronic only]* **16**, 300–314 (2007).

149. S. Klein, J.P.W. Pluim, M. Staring & M.A. Viergever. Adaptive Stochastic Gradient Descent Optimisation for Image Registration. *International Journal of Computer Vision* **81**, 227–239 (2009).

150. Thévenaz, P., Blu, T. & Unser, M. Interpolation revisited. *IEEE Transactions on Medical Imaging* **19**, 739–758 (2000).

151. D.H.J. Poot & S. Klein. Detecting statistically significant differences in quantitative MRI experiments, applied to Diffusion Tensor Imaging. *IEEE Transactions on Medical Imaging,* **34**, 1164–1176 (2015).

152. Castillo, R., Castillo, E., Guerra, R., Johnson, V., McPhail, T., Garg, A. & Guerrero, T. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine and Biology* **54**, 1849 (2009).

153. Coolen, B., Poot, D., Liem, M., Smits, L., Gao, S., Kotek, G., Klein, S. & Nederveen, A. 3D quantitative T1 and T2 mapping of the carotid artery: sequence design and in vivo feasibility. *Magnetic Resonance in Medicine* **75**, 1008–1017 (2016).

154. N. Takaya, C. Yuan, B. Chu, T. Saam, H. Underhill, J. Cai, N. Tran, N.L. Polissar, C. Isaac, M.S. Ferguson, G.A. Garden, S.C. Cramer, K.R. Maravilla, B. Hashimoto & T.S. Hatsukami. Association between carotid plaque characteristics and subsequent ischemic cerebrovascular events: a prospective assessment with MRI — initial results. *Stroke* **37**, 818–813 (2006).

155. P.J. Basser & D.K. Jones. Diffusion-tensor MRI: theory, experimental design and data analysis - a technical review. *Nuclear Magnetic Resonance in Biomedicine* **15**, 456–467 (2002).

156. D.K. Jones & A. Leemans. Diffusion tensor imaging. *Methods in Molecular Biology* **711**, 127–144 (2011).

157. J.D. Tournier, S. Mori & A. Leemans. Diffusion tensor imaging and beyond. *Magnetic Resonance in Medicine* **65**, 1532–1556 (2011).

158. A. Leemans, J. Sijbers, S. De Backer, E. Vandervliet & P. Parizel. Multiscale white matter fiber tract coregistration: a new feature-based approach to align diffusion tensor data. *Magnetic Resonance in Medicine* **55**, 1414–1423 (2006).

159. N. de Geeter, G. Crevecoeur, L. Dupré, W. Van Hecke & A. Leemans. A DTI-based model for TMS using the independent impedance method with frequency-dependent tissue parameters. *Physics in Medicine and Biology* **57**, 2169–2188 (2012).

160. H.C. Wang, J.L. Hsu & A. Leemans. Diffusion tensor imaging of vascular Parkonsonism: structural changes in cerebral white matter and the association with clinical severity. *Archives of Neurology* **69**, 1340–1348 (2012).

161. N.E. van der Aa, A. Leemans, F.J. Northington, H.L. van Straaten, I.C. van Haastert, F. Groenendaal, M.J. Benders & L.S. de Vries. Does diffusion tensor imaging-based tractography at 3 months of age contribute to the prediction of motor outcome after perinatal arterial ischemic stroke. *Stroke* **42**, 3410–3414 (2011).

162. Y.D. Reijmer, A. Leemans, S.M. Heringa, I. Wielaard, B. Jeurissen, H.L. Koek, G.J. Biessels & Z. Wang. Improved sensitivity to cerebral white matter abnormalities in Alzheimer's disease with spherical deconvolution based tractography. *PLoS one* **7**, 1371 (2012).

163. G.J.M. Parker, C. Roberts, A. Macdonald, G.A. Buonaccorsi, S. Cheung, D.L. Buckley, A. Jackson, Y. Watson, K. Davies & G.C. Jayson. Experimentally-derived functional form for a population-averaged high-temporal-resolution arterial input function for dynamic contrast-enhanced MRI. *Magnetic Resonance in Medicine* **56**, 993–1000 (2006).

164. Tofts, P., Brix, G., Buckley, D., Evelhoch, J., Henderson, E., Knopp, M., Larsson, H., Lee, T.-Y., Mayr, N., Parker, G., Port, R., Taylor, J. & Weisskoff, R. Estimating Kinetic Parameters From Dynamic Contrast-Enhanced T1-Weighted MRI of a Diffusable Tracer: Standardized Quantities and Symbols. *Journal of Magnetic Resonance Imaging* **10**, 223–232 (1999).

165. Klaassen, R., Gurney-Champion, O., ter Voert, E., Heerschap, A., Bijlsma, M., Besselink, M., van Tienhoven, G., C.Y. Nio, J., Punt, C., Wilmink, J., van Laarhoven, H. & Nederveen, A. *Motion correction of high temporal 3T Dynamic Contrast Enhanced MRI of pancreatic cancer - preliminary results* in *Proceedings of the 22st Annual Meeting International Society for Magnetic Resonance in Medicine* (2014), 3667.

166. J. Sijbers, A.J. den Dekker, E. Raman & D. Van Dyck. Parameter estimation from magnitude MR images. *International Journal of Imaging Systems and Technology* **10**, 109–114 (1999).

167. S. Cavassila, S. Deval, C. Huegen, D. van Ormondt & D. Graveron-Demilly. Cramér-Rao bounds: an evaluation tool for quantitation. *Nuclear magnetic resonance in biomedicine* **14**, 278–283 (2001).

168. C.R. Rao. Minimum variance and the estimation of several parameters. *Mathematical Proceedings of the Cambridge Philosophical Society* **43**, 280–283 (1946).

169. D.H.J. Poot, G. Kotek, W.J. Niessen & S. Klein. *Bias correction of maximum likelihood estimation in quantitative MRI* in *Proc. SPIE* **8669** (2013), 86691F–6.

170. Delmon, V., Rit, S., Pinho, R. & Sarrut, D. Registration of sliding objects using direction dependent B-splines decomposition. *Physics in Medicine and Biology* **58**, 1303–1314 (2013).

171. F.F. Berendsen, A.N.T.J. Kotte, M.A. Viergever & J.P.W. Pluim. *Registration of organs with sliding interfaces and changing topologies* in *Proceedings of SPIE Medical Imaging 9034: Medical Imaging 2014: Image Processing* **9034** (2014), 90340E.

172. C.F. Baumgartner, C. Kolbitsch, D.R. Balfour, P.K. Marsden, J.R. McClelland, D. Rueckert & A.P. King. High-resolution dynamic MR imaging of the thorax for respiratory motion correction of PET using groupwise manifold alignment. *Medical Image Analysis* **18**, 939–952 (2014).

173. A. Rossi, A. Dharampal, A. Wragg, L.C. Davies, R.J. van Geuns, C. Anagnostopoulos, E. Klotz, P. Kitslaar, A. Broersen, A. Mathur, K. Nieman, M.G.M. Hunink, P.J. de Feyter, S.E. Petersen & F. Pugliese. Diagnostic performance of hyperaemic myocardial blood flow index obtained by dynamic computed tomography: does it predict functionally significant coronary lesions? *European Heart Journal – Cardiovascular Imaging* **15**, 85–94 (2014).

174. M. Muzi, F. O'Sullivan, D. Mankoff, R. Doot, L. Pierce, B. Kurland, H. Linden & P. Kinahana. QIN: Quantitative Assessment of Dynamic PET Imaging Data in Cancer Imaging. *Magnetic Resonance Imaging* **30**, 1203–1215 (2012).

175. Unser, M. Splines: A Perfect Fit for Signal and Image Processing. *IEEE Signal Processing Magazine* **16**, 22–38 (1999).

176. Hsu, M., Hughes, J. & Kaufman, H. Direct Manipulaion of Free-Form Deformations. *Computer Graphics* **26**, 177–184 (1992).

177. S. Lee, G. Wolberg, K. Chwa & S.Y. Shin. Image metamorphosis with scattered feature constraints. *IEEE Transactions on Visualization and Computer Graphics* **2**, 337–354 (1996).

178. Y. Qiao, B. van Lew, B.P. Lelieveldt & M. Staring. Fast Automatic Step Size Estimation for Gradient Descent Optimization of Image Registration. *IEEE Transactions on Medical Imaging* **35**, 391–403 (2016).

179. M.P. Heinrich, W. Bartlomiej, J. Papiez, J.A. Schnabel & H. Handels. *Non-parametric Discrete Registration with Convex Optimisation* in *Biomedical Image Registration: WBIR14, 6th International Workshop on Biomedical Image Registration* (eds S. Ourselin & M. Modat) **LNCS, vol. 8545** (Springer, Cham), 51–61.

180. A. Dame & E. Marchand. Second-Order Optimization of Mutual Information for Real-Time Image Registration. *IEEE Transactions on Image Processing* **21**, 4190–4203 (2012).

181. J. Ashburner & K.J. Friston. Diffeomorphic registration using geodesic shooting and Gauss–Newton optimisation. *NeuroImage* **55**, 954–967 (2011).

182. T. Vercauteren, X. Pennec, A. Perchant & N. Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage* **45**, S61–S72 (2009).

183. M. Sdika. A Fast Nonrigid Image Registration With Constraints on the Jacobian Using Large Scale Constrained Optimization. *IEEE Transactions on Medical Imaging* **27**, 271–281 (2009).

184. R. Bhagalia & J.A. Fessler. Accelerated Nonrigid Intensity-Based Image Registration Using Importance Sampling. *IEEE Transactions on Medical Imaging* **28**, 1208–1216 (2009).

185. M.R. Sabuncu & P. Ramadge. Using spanning graphs for efficient image registration. *IEEE Transactions on Image Processing* **17**, 788–797 (2008).

186. S. Klein, M. Staring & J.P.W. Pluim. Evaluation of Optimization Methods for Nonrigid Medical Image Registration Using Mutual Information and B-Splines. *IEEE Transactions on Image Processing* **16**, 2879–2890 (2007).

187. Sarrut, D. & Vandemeulebroucke, J. B-LUT: Fast and low memory B-spline image interpolation. *Computer Methods and Programs in Biomedicine* **99**, 172–178 (2010).

188. O. Fluck, C. Vetter, W. Wein, A. Kamen, B. Preim & R. Westermann. A survey of medical image registration on graphics hardware. *Computer Methods and Programs in Biomedicine* **104**, e45–e57 (2011).

189. Veldhuizen, T. Using C++ template metaprograms. *C++ Report* **7**, 36–43 (1995).

190. Graham, R., Knuth, D. & Patashnik, O. in. 2nd (Addison-Wesley Publishing Company, 1990).

191. J.A. Schnabel, D. Rueckert, M. Quist, J.M. Blackall, A.D. Castellano Smith, T. Hartkens, G. P., W.A. Hall, H. Liu, C.L. Truwit, F.A. Gerritsen, D.L.G. Hill & D.J. Hawkes. *A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations.* in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2001 (4th International Conference, Utrecht, The Netherlands, October 14 - 17, 2001)* **2208** (Springer-Verlag), 573–581.

192. T.S. Yoo & D.N. Metaxas. ITK — Open science - combining open data and open source software: Medical image analysis with the Insight Toolkit. *Medical Image Analysis* **9**, 503–506 (2005).

193. M. Modat. NiftyReg. `http://cmictig.cs.ucl.ac.uk/wiki/index.php/NiftyReg`. Last modified on 11 March 2015 (2010).

194. Viola, P. & Wells, W. Alignment by Maximization of Mutual Information. *International Journal of Computer Vision* **24**, 137–154 (1997).

195. Thévenaz, P. & Unser, M. Optimization of mutual information for multiresolution image registration. *IEEE Transactions on Image Processing* **9**, 2083–2099 (2000).

196. Staring, M., Klein, S. & Pluim, J. A rigidity penalty term for nonrigid registration. *Medical Physics* **34**, 4098–4108 (2007).

197. Penney, G., Schnabel, J., Rueckert, D., Hawkes, D. & Niessen, W. *Registration-Based Interpolation Using a High-Resolution Image for Guidance* in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004 (7th International Conference, Saint-Malo, France, September 26-29, 2004. Proceedings, Part I)* **3216** (Springer-Verlag, 2004), 558–565.

198. S.S. Samant & J. Xia. High performance computing for deformable image registration: Towards a new paradigm in adaptive radiotherapy. *International Journal of Medical Physics Research and Practice* **35**, 3546–3553 (2008).

199. O.M. Dorgham, S.D. Laycock & M.H. Fisher. GPU accelerated generation of digitally reconstructed radiographs for 2-D/3-D image registration. *IEEE Transactions on biomedical engineering* **59**, 2594–2603 (2012).

200. S. Park, S. Choi, J. Kim & J.S. Chae. Real-time 3D registration using GPU. *Machine Vision and Applications* **22**, 837–850 (2011).

201. A.A. Ballin, D. Bock, R.C. Reid & S.K. Warfield. Accelerating image registration with the Johnson-Lindenstrauss lemma: Application to imaging 3-D neural ultrastructure with electron microscopy. *IEEE Transactions on Medical Imaging* **30**, 1427–1438 (2011).

202. E.Z. Zhang, Y. Jiang, Z. Guo & X. Shen. *Streamlining GPU applications on the fly: thread divergence elimination through runtime thread-data remapping* in *ICS '10 Proceedings of the 24th ACM International Conference on Supercomputing* Tsukuba, Ibaraki, Japan (271 – 281, 2010), 115–126.

203. E.E. Bron, M. Smits, J.M. Papma, R.M.E. Steketee, R. Meijboom, M. de Groot, J.C. van Swieten, W.J. Niessen & S. Klein. Multiparametric computer-aided differential diagnosis of Alzheimer's disease and frontotemporal dementia using structural and advanced MRI. *European Journal of Radiology* **27**, 3372–3382 (2017).

204. H.M. Fonteijn, M. Modat, M.J. Clarkson, J. Barnes, M. Lehmann, N.Z. Hobbs, R.I. Scahill, S.J. Tabrizi, S. Ourselin, N.C. Fox & D.C. Alexander. An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. *NeuroImage* **60**, 1880–1889 (2012).

205. V. Venkatraghavan, E.E. Bron, W.J. Niessen & S. Klein. *A Discriminative Event Based Model for Alzheimer's Disease Progression Modeling* in *Information Processing in Medical Imaging (IPMI)* (eds M. Niethammer & et al.) **LNCS, vol. 10265** (Springer, Cham, 2017), 121–133.

206. G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W.M. van der Laak, B. van Ginneken & C.I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis* **42**, 60–88 (2017).

207. W. Shen, G. Wu & H.-I. Suk. Deep learning in medical image analysis. *Annual review of Biomedical Engineering* **19**, 221–248 (2017).

208. S. Miao, Z.J. Wang & R. Liao. A CNN regression approach for real-time 2D/3D registration. *IEEE Transactions on Medical Imaging* **35**, 1352–1363 (2016).

209. H. Sokooti, B. de Vos, F. Berendsen, B.P.F. Lelieveldt, I. Isgum & M. Staring. *Nonrigid image registration using multi-scale 3D convolutional neural networks* in *MICCAI 2017: Medical Image Computing and Computer-Assisted Intervention* **LNCS, vol. 10433** (Springer, Cham, 2017), 232–239.

210. K.A.J. Eppenhof, M.W. Lafarge, P. Moeskops, M. Veta & J.P.W. Pluim. *Deformable image registration using convolutional neural networks* in *Proc. SPIE, Medical Imaging 2018: Image Processing* (2018).

211. X. Yang, R. Kwitt & M. Niethammer. *Fast predictive image registration* in *Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, LABELS 2016 and the Iternational Workshop on Deep Learning in Medical Image Analysis, DLMIA 2016. MICCAI 2016 workshop.* (eds G. Carneiro & et al.) **LNCS, vol. 10008** (Springer, Cham, 2018), 48–57.

212. G.V. Roshchupkin, B.A. Gutman, M.W. Vernooij, N. Jahanshad, N.G. Martin, K.L. McMahon, A. H. K., S.J. van der Lee, C.M. van Duijn, G.I. de Zubicaray, A.G. Uitterlinden, M.J. Wright, W.J. Niessen, P.M. Thompson, M.A. Ikram & H.H.H. Adams. Heritability of the shape of subcortical brain structures in the general population. *Nature Communications* **7**, 13738 (2016).

DANKWOORD

Ik wil graag iedereen bedanken die betrokken is geweest bij het tot stand komen van dit proefschrift. Een aantal mensen wil ik hieronder in het bijzonder bedanken.

Mijn eerste woord van dank gaat uit naar mijn promotor. Wiro, bedankt voor de kansen die je me bood om onderzoek te doen in jouw groep, zowel mijn Master als mijn PhD project. Jouw enthousiasme en positiviteit werken stimulerend in een onderzoeksomgeving.

Grote dank gaat uit naar mijn co-promotoren, Stefan en Dirk, die altijd voor me klaar stonden om vragen te beantwoorden, discussies te voeren en de zoveelste versie van een manuscript te lezen en te beoordelen. Stefan, jouw creativiteit lijkt geen grenzen te kennen. Excuses als: "dat kan (ik) niet", worden niet door jou getolereerd, want weinig is onmogelijk in jouw ogen. Ook bewonder ik jouw enorme integriteit, zowel op professioneel als op persoonlijk vlak. Jij bent een voorbeeld voor iedere wetenschapper. Dirk, ik ken niemand die zo snel kan redeneren en afleiden als jij. De discussies waren voor mij soms moeilijk te volgen, maar wel altijd heel interessant en leerzaam. Bedankt voor je waardevolle inzichten en je begeleiding, zowel tijdens mijn Master als mijn PhD traject.

I would like to thank the members of the reading committee who assessed my thesis. Special thanks go to Prof. Daniel Rueckert. Daniel, thank you for the opportunity to join your group for a few months. I learned a lot and had a great time. Thank you Prof. Stanley Durrleman and Prof. Aad van der Lugt for reading and approving my thesis.

In het bijzonder wil ik mijn paranimfen, Carolyn en Lotte, bedanken. Carolyn, we were office mates from the start and it immediately clicked. Thank you for the funny and the serious conversations, the chai lattes at Doppio, the scientific input and life lessons. The idea that we will be practically neighbors and that we can continue our friendship makes me very happy. Lotte, vanaf het begin van onze PhDs werden we bij elkaar gezet en het duo barbie en de nerd was geboren. Dat het zo'n geweldige samenwerking zou worden had ik niet van tevoren kunnen bedenken. Jij en ik weten hoe het is om tijdens een PhD traject moeder te worden en om te gaan met het feit dat een belangrijk familielid ernstig ziek is. Bedankt voor alles.

I would also like to thank all my colleagues at BIGR. It is very important to have such a great group of talented and friendly people to work with. Thank you Annegreet, Arna, Chao Ping, Erwin, Esther, Emilie, Gijs, Ghassan, Gennady, Gerardo, Hakim, Marcel, Marius, Pierre, Veronika, Vikram, Willem, Yao Yao, and Zahra. I would also like to thank the people from the RAMBO group for their valuable input, enthusiasm and cooperation. Thank you Diego, Floris, Gokhan, Jean-Marie, Marius, Martijn, Sebastian, Wei Sun and Yuchuan. Tenslotte wil ik Désirée en Petra bedanken voor alle administratieve hulp tijdens mijn PhD traject en de afronding ervan.

Ik wil ook mijn collega's bij radiologie & nucleaire geneeskunde bedanken voor de goede samenwerking. In het bijzonder wil ik Mark, Eline en Meike bedanken. Mark, ik zag je niet vaak, maar als je er was waren de gesprekken altijd gezellig, of ze nu werkgerelateerd of oudergeretaleerd waren. Eline, onze samenwerkingen was kort maar ik heb hem als zeer prettig ervaren. Meike, bedankt voor je waardevolle inzichten en de gezelligheid tijdens de VPH-DARE@IT meetings, zowel intern als extern.

In addition, I would like to thank my colleagues at Imperial College London and University College London for the nice time I had in London. Thank you Alexander, Andreas, Christian, Emma, Ira, Jose, Kerstin, Kevin, Lisa, Martin, Nico, Ricardo, Sarah and Stefan.

Ik wil ook al mijn nieuwe collega's bij TNO bedanken. Ondanks de relatief korte tijd die ik er zit voel ik me er thuis en ik ervaar de groep als hecht en betrokken. In het bijzonder wil ik Esther bedanken voor alle vrijheid die ik kreeg bij mijn start bij Intelligent Imaging, die lastig was door mijn thuissituatie. Iemand kan zich in zo'n situatie geen betere werkgever wensen.

Zingen is altijd een belangrijke hobby en uitlaatklep geweest in mijn vrije tijd. Ik wil iedereen bij Krashna Musika en Delft Blue bedanken voor de geweldige tijd die ik heb gehad tijdens het zingen bij deze koren.

Familie en vrienden hebben een onvergetelijke steun geleverd tijdens mijn promotie. Lieve Marije, bedankt voor dat ik alles wat ik met je kan delen, de geweldige tijd tijdens tournees en repetitieweekenden en de gezellige klaverjasavonden. Je bent een vriendin uit duizenden.

Lieve Folkert, Tessa, Gina, Michel, Jan-Jelle en Marloes, bedankt voor jullie interesse en betrokkenheid. Jullie zijn fantastische (schoon)familie.

Lieve Nico, Anneke en Renske, bedankt dat jullie altijd klaar staan om ons te helpen, van oppassen op Hidde tot klussen in en rond het huis.

Lieve mama, bedankt voor jouw liefde en zorgzaamheid, zowel voor mij als voor mijn gezin. Lieve papa, van jou heb ik geleerd om iedere letterlijke en figuurlijke berg te nemen met een portie doorzettingsvermogen en optimisme. Dit heeft mij tijdens mijn PhD traject enorm geholpen. Bedankt.

Lieve Hidde, hoe goed of slecht een dag ook is geweest, jij maakt hem altijd weer een stukje beter.

Lieve Hajo, jij spoort me aan en fluit me terug wanneer dat nodig is, al twaalf jaar lang. Bedankt dat je me altijd steunt en begrijpt. Ik kijk uit naar een lang (!) en gelukkig leven met ons gezin.

PUBLICATIONS

Publications in peer-reviewed journals:

M. Polfliet, S. Klein, **W. Huizinga**, M.M. Paulides, W.J. Niessen & J. Vandemeulebroucke. Intra-subject multimodal groupwise registration with the conditional template entropy. *Medical Image Analysis* **46**, 15–25 (2018).

J.M. Guyader, **W. Huizinga**, V. Fortunati, D.H.J. Poot, J.F. Veenland, M.M. Paulides, W.J. Niessen & S. Klein. Groupwise multi-channel image registration. *IEEE Journal of Biomedical and Health Informatics* (2018).

W. Huizinga, D.H.J. Poot, M.W. Vernooij, G. Roshchupkin, E.E. Bron, M.A. Ikram, D. Rueckert, W.J. Niessen, S. Klein & the Alzheimer's Disease Neuroimaging Initiative. A spatio-temporal reference model of the aging brain. *NeuroImage* **169**, 11–22 (2018).

C.D Langen, H.I. Zonneveld, T. White, **W. Huizinga**, L.G.M. Cremers, M. de Groot, M.A. Ikram, W.J. Niessen & M.W. Vernooij. White matter lesions relate to tract-specific reductions in functional connectivity. *Neurobiology of Aging* **51**, 97–103 (2017).

W. Huizinga, D.H.J. Poot, J.-M. Guyader, R. Klaassen, B.F. Coolen, M. van Kranenburg, R.J. van Geuns, A. Uitterdijk, M. Polfliet, J. Vandemeulebroucke, A. Leemans, W.J. Niessen & S. Klein. PCA-based groupwise image registration for quantitative MRI. *Medical Image Analysis* **29**, 65–78 (2016).

Publications in submission:

J.M. Guyader, **W. Huizinga**, D.H.J. Poot, M. van Kranenburg, A. Uitterdijk, W.J. Niessen & S. Klein. Groupwise image registrayion based on a total correlation dissimilarity measure for quantitative MRI and dynamic imaging data. submitted (2018).

L.G.M. Cremers*, **W. Huizinga***, D.H.J. Poot, J. Lötjönen, W.J. Niessen, M.A. Ikram, S. Klein & M.W. Vernooij. Predicting global cognitive decline in the general population using the Disease State Index. submitted, * *(shared first authorship)* (2018).

W. Huizinga, D.H.J. Poot, E.J. Vinke, F. Wenzel, N. Toussaint, C. Ledig, M.A. Ikram, M.W. Vernooij, W.J. Niessen & S. Klein. Differences between MR brain region segmentation methods: impact on single-subject analysis. submitted (2018).

W. Huizinga, M. Staring, S. Klein & D.H.J. Poot. Fast multi-dimensional B-spline algorithms using template metaprogramming. submitted (2018).

Publications in conference proceedings:

C.D. Langen, M.W. Vernooij, L.G.M. Cremers, **W. Huizinga**, M. de Groot, M.A. Ikram, T. White & W.J. Niessen. *The structural disconnectome: A pathology-sensitive extension of the structural connectome* in *Proceedings of the 14th IEEE International Symposium on Biomedical Imaging* Melbourne, Australia (IEEE, 2017).

M. Polfliet, S. Klein, **W. Huizinga**, J. de Mey & J. Vandemeulebroucke. *The Pythagorean averages as group images in efficient groupwise registration* in *Proceedings of the 13th IEEE International Symposium on Biomedical Imaging* Prague, Czech Republic (IEEE, 2016), 1261–1264.

J.-M. Guyader, **W. Huizinga**, V. Fortunati, D.H.J. Poot, M. van Kranenburg, J.F. Veenland, M.M. Paulides, W.J. Niessen & S. Klein. *Total correlation-based groupwise image registration for quantitative MRI* in *Workshop on Biomedical Image Registration - CVPRW 2016 Workshop* Las Vegas, United States of America (IEEE, 2016), 186–193.

W. Huizinga, D.H.J. Poot, G. Roshchupkin, E.E. Bron, M.A. Ikram, M.W. Vernooij, D. Rueckert, W.J. Niessen & S. Klein. *Modeling the brain morphology distribution in the general aging population* in *Proceedings of the SPIE Medical Imaging 2016: Biomedical Applications in Molecular, Structural, and Functional Imaging* San Diego, United States of America. **9788** (SPIE, 2016), 97880I 7.

J.-M. Guyader, **W. Huizinga**, V. Fortunati, Veenland, J., M.M. Paulides, W.J. Niessen & S. Klein. *Groupwise image registration of multimodal head-and-neck images* in *Proceedings of the 12th IEEE International Symposium on Biomedical Imaging* New York, United States of America (IEEE, 2015).

M. Polfliet, **W. Huizinga**, S. Klein, J. de Mey & J. Vandemeulebroucke. *Pythagorean Mean Images for Efficient Groupwise Registration* in *Imaging and Computer Assistance in Radiation Therapy - MICCAI 2015 Workshop* Munich, Germany (2015).

W. Huizinga, D.H.J. Poot, J.-M. Guyader, H. Smit, M. van Kranenburg, R.J. van Geuns, A. Uitterdijk, H. van Beusekom, B.F. Coolen, A. Leemans, W.J. Niessen & S. Klein. *Non-rigid groupwise image registration for motion compensation in quantitative MRI* in *Proceedings of the International Workshop on Biomedical Image Registration 2014* London, United Kingdom. **5845** (Springer, 2014), 184–193.

W. Huizinga, M. Staring, S. Klein & D.H.J. Poot. *Fast Multidimensional B-spline Interpolation Using Template Metaprogramming* in *Proceedings of the International Workshop on Biomedical Image Registration 2014* London, United Kingdom. **5845** (Springer, 2014), 11–20.

W. Huizinga, C.T. Metz, D.H.J. Poot, M. de Groot, A. Leemans, W.J. Niessen & S. Klein. *Groupwise registration for correcting subject motion and eddy current distortions in diffusion MRI using a PCA based dissimilarity metric* in *Computational Diffusion MRI and Brain Connectivity - MICCAI 2014 Workshop* Nagoya, Japan (Springer, 2014), 163–174.

Conference abstracts:

E.J. Vinke, **W. Huizinga**, M. Bergtholdt, M.A. Ikram, W.J. Niessen, F. Wenzel & M.W. Vernooij. *Comparing normative subcortical volume distributions across cohorts* in *European Conference of Radiology (ECR)* (2018).

Miscellaneous:

W. Huizinga & G. Roshchupkin. *Publicly available web-page: A spatio-temporal reference model of the aging brain* `https://www.agingbrain.nl`. 2017.

W. Huizinga, M.W. Vernooij, M.A. Ikram, S. Klein, D.H.J. Poot & W.J. Niessen. *Deliverable 6.8: Software tool for interaction with ageing brain models* Published as public deliverable in the VPH-DARE@IT project. 2017.

W. Huizinga, M.W. Vernooij, M.A. Ikram, S. Klein, D.H.J. Poot & W.J. Niessen. *Deliverable 6.6: Evaluation of method to contrast individual to ageing brain models* Published as public deliverable in the VPH-DARE@IT project. 2017.

F. Wenzel, M. Bergtholdt, Z. Taylor, **W. Huizinga** & N. Ravikumar. *Deliverable 6.2: Model-based segmentation of sub-cortical and cortical brain structures in T1-weighted MRI* Confidential deliverable in the VPH-DARE@IT project. 2016.

W. Huizinga, M.W. Vernooij, M.A. Ikram, S. Klein, L.G.M. Cremers, D.H.J. Poot & W.J. Niessen. *Deliverable 6.5: Ageing brain models in normal ageing and disease: virtual population for biophysical modelling* Published as public deliverable in the VPH-DARE@IT project. 2015.

W. Huizinga, M.W. Vernooij, M.A. Ikram, S. Klein, D.H.J. Poot & W.J. Niessen. *Deliverable 6.4: Software tool implementing generative 3D+t spatiotemporal models of the ageing brain* Published as public deliverable in the VPH-DARE@IT project. 2015.

N. Toussaint, D. Cash & **W. Huizinga**. *Deliverable 3.3: Workflow implementation and streamlining for high-throughput image analysis of large-scale studies* Published as public deliverable in the VPH-DARE@IT project. 2015.

W. Huizinga, W.J. Niessen, M.M.S. Jasperse, M. Koek, H.A. Vrooman, M.W. Vernooij, M.A. Ikram, S. Klein & D.H.J. Poot. *Deliverable 6.1: Normal ageing brain models – Initial phenomenological model and links to mechanical modelling framework* Published as public deliverable in the VPH-DARE@IT project. 2014.

PHD PORTFOLIO

PhD period	2013 - 2017
Departments	Medical Informatics and Radiology & Nuclear Medicine

IN-DEPTH COURSES

Summer school on advanced topics in medical image processing, Nijmegen, the Netherlands	2015
From human neuroanatomy to psychopathology, Maastricht, the Netherlands (MHeNs)	2015
Research integrity, Rotterdam, the Netherlands (Erasmus MC)	2015
Winter school on pattern analysis and classification, Amsterdam, the Netherlands (ABC)	2014
Biomedical writing and communication, Rotterdam, the Netherlands (Erasmus MC)	2014
Presentation course, Rotterdam, the Netherlands (Medical Informatics)	2014
Advanced MR physics, Utrecht, the Netherlands (ImagO)	2014
Front-end vision and multi-scale image analysis, Eindhoven, the Netherlands (ASCI)	2013
Summer school on imaging in neurology, Dubrovnik, Coratia (EIBIR)	2013

INTERNATIONAL CONFERENCES

SPIE Medical Imaging, San Diego, United States (oral presentation)	2016
Workshop on Biomedical Image Registration (WBIR ’14), Londen, United Kingdom (oral presentation)	2014
Workshop on Biomedical Image Registration (WBIR ’14), Londen, United Kingdom (poster)	2014

SEMINARS AND WORKSHOPS

Symposium Netherlands Forum for Biomedical Imaging (NFBIA), Nijmegen (oral presentation)	2015
Medical Imaging Symposium for PhD students (MISP), Eindhoven (oral presentation)	2016
Medical Imaging Symposium for PhD students (MISP), Amsterdam (attendance)	2015
Medical Image Computing and Computer-Assisted Intervention (MICCAI) workshop on Computational Diffusion MRI and Brain Connectivity, Nagoya, Japan (poster)	2013

RESEARCH SEMINAR SERIES

Biomedical Imaging Group Rotterdam Seminars, bi-weekly (2 oral presentations)	2014 - 2017
Medical Informatics Research Lunch, bi-weekly (3 oral presentations)	2014 - 2017
RAMBO Medical Image Registration meeting, bi-weekly (oral presentations)	2014 - 2017

AWARDS, NOMINATIONS AND GRANTS

Finalist in Robert F. Wagner all-conference best student paper award, *SPIE*	2016
Traveling grant for visiting Imperial College London, *Erasmus Trustfonds*	2014
Traveling grant for visiting Imperial College London, *Stichting de Drie Lichten*	2014

TEACHING EXPERIENCE

Teaching introduction to image processing to medical students	2015 - 2016
Supervision student project: Romain Niel. Project: Fast B-spline transformation using template metaprogramming	2015

OTHER

Open source software

Registration method for quantitative MRI data (`elastix.isi.uu.nl`)	2015
Fast algorithms for multidimensional B-spline interpolation and transformation (`elastix.isi.uu.nl`)	2017

General assembly meetings of the FP7 VPH-DARE@IT project

Venice (attendance)	2017
Hamburg (poster)	2016
Rotterdam (poster)	2015
Venice (poster)	2014
Manchester (attendance)	2014
Sheffield (attendance)	2013

European Commision review meetings of the FP7 VPH-DARE@IT project

Venice (oral presentation)	2017
Brussels (oral presentation)	2015
Brussels (oral presentation)	2014

ABOUT THE AUTHOR

Wyke Huizinga was born on 1989 in Alkmaar, the Netherlands. After finishing high school Wyke started the Bachelor's program in Applied Physics at the University of Technology in Delft. She did a minor in Biomedical Engineering and in that time she developed an interest for medical imaging. In 2013 Wyke received a Master of Science degree in Applied Physics with a specialization in image science and technology. Her graduation project was also performed in the Biomedical Imaging Group Rotterdam at the Erasmus MC, which resulted in a Master's thesis entitled "Groupwise Image Registration in Diffusion Weighted MRI using a PCA based Dissimilarity Metric".

As of August 2013 Wyke started her PhD project at BIGR, under the supervision of Prof. dr. Wiro Niessen, dr. Stefan Klein, and dr. Dirk Poot. Her work was part of a large European project, the FP7 VPH-DARE@IT project, which aimed to develop methods and clinical tools to aid the diagnosis of dementia. The main aim of Wyke's work was to obtain insight in the brain morphology change, to distinguish neurodegeneration due to normal aging from that of neurodegenerative diseases. During this period Wyke also extended her master's thesis projefct and contributed to the open source image registration software elastix.

From September 2014 till January 2015 Wyke visited the Biomedical Image Analysis group (BioMedIA, headed by Prof. dr. Rueckert) at Imperial College London. Here, Wyke collaborated with researchers from Imperial College London and University College London within the scope of the VPH-DARE@IT project.

As of March 2018 Wyke started as a researcher in the department of Intelligent Imaging at TNO.

ABOUT THE COVER

In several chapters of this thesis I propose the usage of normative data for brain MRI. During my PhD research my son Hidde was born which brought the usage of normative data to my personal life. The cover shows the percentile curves of the normative weight distribution of Dutch boys in the age range of 0 - 15 months. This normative data is acquired and distributed by TNO Healthy Living. The normative distributions are freely available and used by the Dutch consultation bureaus to monitor the growth of children under the age of four years. The individual data points in the curves on the cover show the weight of Hidde at multiple time points. This cover is therefore both contentwise as well as personally related to me during the time I did my PhD research.