MICCAI 2014 Workshop Proceedings Challenge on Computer-Aided Diagnosis of Dementia based on Structural MRI Data

Esther E. Bron, MSc Marion Smits, MD, PhD Prof. John C. van Swieten, MD, PhD Prof. Wiro J. Niessen, PhD Stefan Klein, PhD

Erasmus MC, Rotterdam, the Netherlands http://caddementia.grand-challenge.org





Preface

Promising results of computer-aided diagnosis (CAD) methods for dementia have been presented in the literature, but it is unclear how the different algorithms would perform on previously unseen data, and thus, how they would perform in clinical practice. To address this, we initiated the CADDementia challenge on computer-aided diagnosis of dementia based on structural MRI data.

This challenge aims to objectively compare algorithms for dementia classification based on a clinically representative multi-center data set. Research groups were invited to run their classification methods on the data and make a multiclass classification of patients with Alzheimers disease (AD), patients with mild cognitive impairment (MCI) and healthy controls (CN). The clinical diagnosis assessed with the most current and widely used criteria is used as a reference standard. For evaluation of the algorithms, a standardized methodology has been developed based on three key points: comparability, generalizability, and clinical applicability. First, comparability was enhanced by the use of the same data set and evaluation methods. Second, by providing a previously unseen multicenter data with blinded ground truth diagnoses overtraining is avoided and generalizability of the methods is promoted. Third, according to the current clinical standards, a multi-class diagnosis of AD, MCI and controls is evaluated.

The workshop of the CADDementia challenge is held on September 18th, in conjunction with the 17th International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI 2014) in Boston, USA. At this workshop, the first results and participating methods of the challenge are presented. Fifteen workshop papers of challenge participants were accepted for presentation at this workshop. Based on technical quality and results, five of the papers were selected to be presented in an oral presentation. For the other papers, the authors were given the opportunity to present their algorithms with a poster presentation.

We would like to thank all the participants for their hard work in processing the 384 images and developing a classification method over a relatively short time period. We also would like to thank the MICCAI 2014 organizers, the guest lecturer Prof. Giovanni Frisoni, the data-contributing centers and the workshop sponsors Quantib BV and Biomediq.

These proceedings consist of 1) the program of the workshop, 2) more information on the challenge framework, and 3) the 15 accepted workshop papers. More details can be found on http://caddementia.grand-challenge.org.

Esther E. Bron Marion Smits Prof. John C. van Swieten Prof. Wiro J. Niessen Stefan Klein

Workshop program

- 8.00-8.15 Opening by Prof. Wiro Niessen
- 8.15-9.00 Guest lecture by Prof. Giovanni Frisoni: Structural features for diagnosis and tracking of Alzheimer's disease
 Laboratory of Epidemiology, Neuroimaging & Telemedicine, IRCCS San Giovanni di Dio Fatebenefratelli, Italy; Memory Clinic and Laboratory of Neuroimaging of Aging, University Hospitals and University of Geneva, Switzerland
- 9.00-9.15 Dementia classification based on brain age estimation K Franke, C Gaser Structural Brain Mapping Group, Departments of Neurology & Psychiatry, Jena University Hospital, Germany
- 9.15-9.30 Alzheimer's disease state classification using structural volumetry, cortical thickness and intensity features C Ledig, R Guerrero, T Tong, K Gray, A Schmidt-Richberg, A Makropoulos, RA Heckemann, D Rueckert Department of Computing, Imperial College London, UK
- 9.30-9.35 Preview of challenge results
- 9.35-10.00 Poster Session

1. Voxel-based multi-class classification of AD, MCI, and elderly controls. Blind evaluation on an independent test set

A Abdulkadir, J Peter, T Brox, O Ronneberger, S Klöppel

Department of Psychiatry and Psychotherapy, University Medical Centre Freiburg, Germany

2. PRISMA-CAD: Fully automated method for computeraided diagnosis of dementia based on structural MRI data N Amoroso, R Errico, R Bellotti, the ADNI

National Institute of Nuclear Physics, Branch of Bari, Italy

3. CADDementia based on structural MRI using supervised kernel-based representations

D Cárdenas-Peña, A Álvarez-Meza, G Castellanos-Dominguez Signal Processing and Recognition Group, Universidad Nacional de Colombia, Colombia

4. Classification of Alzheimer's disease using structural MRI CV Dolph, MD Samad, KM Iftekharuddin

Vision Lab, Old Dominion University, VA, USA

5. Detecting Alzheimer's disease by morphological MRI using hippocampal grading and cortical thickness SF Eskildsen, P Coupé, V Fonov, DL Collins, the ADNI

Center of Functionally Integrative Neuroscience, Aarhus University, Denmark

6. MRI based dementia classification using semi-supervised learning and domain adaptation

E Moradi, C Gaser, H Huttunen, J Tohka, the ADNI

Department of Signal Processing, Tampere University of Technology, Finland

7. Advanced feature selection in multinominal dementia classification from structural MRI data

A Sarica, G Di Fatta, G Smith, M Cannataro, D Saddy, the ADNI Department of Medical and Surgical Sciences, Magna Graecia University of Catanzaro, Italy

8. Global Disease Index, a novel tool for MTL atrophy assessment

F Sensi, L Rei, G Gemme, P Bosco, N Amoroso, A Chincarini, the ADNI

National Institute of Nuclear Physics, Branch of Genoa, Italy

9. Towards the computer-aided diagnosis of dementia based on the geometric and network connectivity of structural MRI data

GM Smith, ZV Stoyanov, DV Greetham, P Grindrod, JD Saddy, the ADNI

School of Systems Engineering, University of Reading, UK

10. MIND-BA: Fully automated method for computer-aided diagnosis of dementia based on structural MRI data S Tangaro, P Inglese, R Maglietta, A Tateo, the ADNI

National Institute of Nuclear Physics, Branch of Bari, Italy

10.00-10.30 Coffee break + Poster Session

10.30-10.45 Evaluation of morphometric descriptors of deep brain structures for the automatic classification of patients with Alzheimer's disease, mild cognitive impairment and elderly controls

A Routier, P Gori, AB Graciano Fouquier, S Lecomte, O Colliot, S Durrleman, the ADNI

Sorbonne Universités, UPMC Université Paris, France

10.45-11.00 Dementia diagnosis using MRI cortical thickness, shape, texture, and volumetry

L Sørensen, A Pai, C Anker, I Balas, M Lillholm, C Igel, M Nielsen Department of Computer Science, University of Copenhagen, Denmark

11.00-11.15 *BrainPrint* in the computer-aided diagnosis of Alzheimer's disease

C Wachinger, K Batmanghelich, P Golland, M Reuter Computer Science and Artificial Intelligence Lab, MIT, MA, USA

11.15-12.00 Presentation of the challenge results, including award ceremony

The CADDementia Evaluation framework

In this section, we describe the datasets, the reference standard and the evaluation measures used in the evaluation framework.

Data

As data for the evaluation framework, we composed a multi-center data set consisting of 384 scans. The participating centers are: Erasmus MC (EMC), Rotterdam, the Netherlands; VU University Medical Center (VUmc), Amsterdam, the Netherlands; University of Porto / Hospital de São João (UP), Porto, Portugal. This data set contains structural MRI (T1w) scans of subjects with the diagnosis of probable Alzheimer's disease (AD), mild cognitive impairment (MCI) and participants without a dementia syndrome (controls). In addition to the MR scans, demographic information (age, gender) and information on which data are from the same institute is included.

Most of the data is used for evaluation of the methods: the test set. A small part of the data is provided as training data set, which consists of 30 scans distributed over the diagnostic groups of which the diagnostic labels were made available. We decided to limit the training set to 30 subjects. Since we aim to evaluate the performance in a clinical situation, when not much data similar to the test data is available, we expect data from other sources to be used for training.

The data characterisics are listed in Table 1. The class sizes are not released before the workshop. The prior for each class is $\sim 1/3$.

Reference standard

In the data set AD patients, MCI patients, and controls are included. The data have been acquired either as part of clinical routine or as part of a research study at three centers. All patients underwent neurological and neuropsychological examination as part of their routine diagnostic work up. The clinical diagnosis is established by consensus of a multidisciplinary team. Patients with AD met the clinical criteria for probable AD [1,2]. MCI patients fulfilled the criteria specified by [3]. All subjects signed informed consent and the study was approved by the EMC medical ethical committee.

Data preprocessing

The T1-weighted MRI data were anonymized and faces were removed from the scan. Next to the original anonymized T1w scans, non-uniformity corrected scans were also made available. The correction was performed with N4ITK [4] with the following settings: shrink factor = 4, number of iterations = 150, convergence threshold = 0.00001, initial b-spline mesh resolution = 50.

Table 1: Data characteristics, FSPGR: fast spoiled gradient-recalled echo, IR: inversion recovery, MPRAGE: magnetization prepared rapid acquisition gradient echo,Prot.: Protocol

	EMC	VUmc	UP
Scanner	3T, GE Healthcare	3T, GE Healthcare	3T, Siemens
	Discovery MR750	Signa HDxt	Trio A Trim
Sequence	3D IR FSPGR	3D IR FSPGR	3D MPRAGE
Scan parameters	Prot. 1: 450ms/7.9ms/3.1ms		
(TI/TR/TE)	Prot. 2: 450ms/6.1ms/2.1ms	450ms/7.8ms/3.0ms	900ms/2300ms/3.0ms
	Prot. 3: 300ms/10.4ms/2.1ms		
Parallel imaging	Prot. 1: Yes (ASSET factor=2)		
	Prot. 2: Parallel imaging: No	Yes (ASSET factor=2)	No
	Prot. 3: Parallel imaging: No		
Resolution	Prot. 1: 0.94x0.94x1.0 mm (sagittal)		
	Prot. 2: 0.94x0.94x0.8 mm (axial)	0.9x0.9x1 mm	1x1x1.2 mm
	Prot. 3: 0.49x0.49x0.8 mm (axial)		
Number of scans	174	180	30
Age Mean (Std)	68.6 (7.8)	62.2(5.9)	67.8 (9.1)
Male (%)	63%	59%	50%

Evaluation

In this challenge, teams submitted outcomes on a large test set. Submitting the diagnostic label for each sample was obligatory. Additionally, submitting the output probabilities for each label was encouraged.

Accuracy and the true positive fraction for the three classes are calculated from the diagnostic labels. For every class, an ROC curve is calculated from the output probabilities, showing the ability of the classifier to separate that class from the other two classes. In addition, a multiclass AUC is calculated [5]. These measures are calculated separately on the test and training set. The Python scripts for evaluation can be downloaded from the web site.

Submitted algorithms are ranked based on accuracy on the test set. Algorithms for which output probabilities are available are also ranked on the AUC on the test set. The algorithm with the best accuracy (rank=1) on the test set, is the winning algorithm. In case two or more algorithms have equal accuracies, the average rank is assigned to these algorithms.

Web-based evaluation framework

The framework as proposed in this challenge is made publicly available through a web-based interface (*http://caddementia.grand-challenge.org*). From this protected web site, a part of the data was downloaded by the participants. The data available for download were, for the training set: 30 T1w scans from the probable AD, MCI and controls groups including diagnostic label, age, gender and scanner information; and for the test set: 254 T1w scans from the probable AD, MCI and control groups including age, gender and scanner information. The participants submitted their predictions and supporting workshop papers via the web site. We performed the validation for the test set with the software that is available for download as well.

Acknowledgements

We would like to thank several people for their contribution in setting up this challenge. We want to thank all data-providing centers: VU medical center (Frederik Barkhof, Wiesje van der Flier, Ronald van Schijndel), Erasmus MC (Janne Papma, Rebecca Steketee, Carolina Mendez Orellana, Rozanna Meijboom) and the University of Porto (Department of Medical Imaging) / Hospital de São João Departments of Psychiatry and Radiology). In addition, many thanks to Sjoerd Kerkstra (Diagnostic Image Analysis Group, Department of Radiology, Radboud University Medical Center Nijmegen, the Netherlands) for his assistance in building the web-based evaluation framework on *http://grand-challenge.org.*

References

- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M.: Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34(7) (1984) 939–44
- McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jr., C.R.J., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., Mohs, R.C., Morris, J.C., Rossor, M.N., Scheltens, P., Carillo, M.C., Thies, B., Weintraub, S., Phelps, C.H.: The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 7 (2011) 263–269
- Petersen, R.C.: Mild cognitive impairment as a diagnostic entity. J Intern Med 256(3) (2004) 183–94
- Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4ITK: improved N3 bias correction. IEEE Trans Med Imag 29(6) (2010) 1310–1320
- Hand, D.J., Till, R.J.: A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Mach Learn 45 (2001) 171–186

Workshop papers

	Page
A Abdulkadir et al Voxel-based multi-class classification of AD, MCI, and elderly controls. Blind evaluation on an independent test set	8
N Amoroso et al PRISMA-CAD: Fully automated method for Computer-Aided Diagnosis of Dementia based on structural MRI data	16
D Cárdenas-Peña et al $C\!ADDementia\ based\ on\ structural\ MRI\ using\ Supervised\ Kernel-based\ Representations$	24
CV Dolph et al Classification of Alzheimer's disease using structural $M\!RI$	30
$\label{eq:spectral} \begin{array}{l} {\rm SF\ Eskildsen\ et\ al.}\ -\ Detecting\ Alzheimer's\ disease\ by\ morphological\ MRI\\ using\ hippocampal\ grading\ and\ cortical\ thickness \end{array}$	38
K Franke et al Dementia classification based on brain age estimation	48
C Ledig et al Alzheimer's disease state classification using structural volumetry, cortical thickness and intensity features	55
${\rm E}$ Moradi et al MRI based dementia classification using semi-supervised learning and domain adaptation	65
A Routier et al Evaluation of morphometric descriptors of deep brain structures for the automatic classification of patients with Alzheimer's disease, mild cognitive impairment and elderly controls	74
A Sarica et al Advanced feature selection in multinominal dementia classification from structural MRI data	82
F Sensi et al Global Disease Index, a novel tool for MTL atrophy assessment	92
GM Smith et al Towards the Computer-Aided Diagnosis of Dementia based on the geometric and network connectivity of structural MRI data	101
L Sørensen et al Dementia diagnosis using MRI cortical thickness, shape, texture, and volumetry	111
S Tangaro et al MIND-BA: Fully automated method for Computer- Aided Diagnosis of Dementia based on structural MRI data	119
C Wachinger et al BrainPrint in the Computer-Aided Diagnosis of Alzheimer's disease	129

Voxel-Based Multi-Class Classification of AD, MCI, and Elderly Controls Blind Evaluation on an Independent Test Set

Ahmed Abdulkadir^{1,2,3}, Jessica Peter², Thomas Brox³, Olaf Ronneberger^{3,4}, and Stefan Klöppel^{1,2}

¹ Department of Psychiatry and Psychotherapy, University Medical Centre Freiburg, Freiburg, Germany

² Department of Neurology, University Medical Centre Freiburg, Freiburg, Germany ³ Department of Computer Science, University of Freiburg, Freiburg, Germany

⁴ BIOSS Centre for Biological Signalling Studies, University of Freiburg, Freiburg,

Germany

Abstract. We trained a multi-class support vector machine (SVM) with probabilistic outputs on a large, publicly available sample (n = 1429) of healthy controls, individuals with mild cognitive impairment (MCI), and patients with probable Alzheimer's disease (AD). The test performance on a small validation set (n = 30) was similar to the cross-validation performance of the training set. Average area under the curve was 0.84 for the validation and 0.79 for the training set. The model was then applied to the test set (n = 354) of which no labels were known and the predictions were submitted to the CADDementia Challange.

The method required one hour computation time on a single CPU per subject, and almost no manual intervention.

1 Introduction

Diagnosis of dementia is an important task in clinical routine. In vivo brain imaging supplements clinical assessments by providing information about structure and function and can be used for assisting the diagnosis using automated machine learning methods [11, 15]. In a direct comparison, an automated method for diagnosing Alzheimer's disease (AD) performed as well as or better than clinicians [12]. In a previous study, that we conducted with four different data sets using functional and structural MRI markers, the structural markers were as sensitive as functional imaging markers in diagnosing pre-symptomatic Huntington's disease [2], despite the fact that functional dysfunction precedes structural degeneration in the central nervous system [10]. A direct comparison of different automated methods for diagnosing AD based on structural MRI was conducted by Cuingnet et al. [6]. The data used for the study was acquired on multiple centers for the Alzheimer's Disease Neuroimaging Initiative (ADNI). One of the best performing methods [13] used features similar to those in voxel-based morphometry (VBM) [4]. We showed in multiple studies that the typical pre-processing for VBM studies leads to systematic differences between scanners, but at the same time, the extracted data was sufficiently robust to classify the presence of a disease across sites, and acquisition protocols [13, 1, 14, 20, 3]. Due to the high performance in previous study and robustness in different scenarios, VBM features were selected as the means of classification. In order to increase sensitivity and specificity, we applied data driven feature selection. Further, we aimed to reduce confounding effects of age, head size, and sex.

2 Materials

2.1 Test Data

The test data was provided through the web site on the challenge on Computer-Aided Diagnosis of Dementia (CADDementia) based on structural MRI data¹. CADDementia provided T1 weighted MRI along with age and sex as basic demographic covariates. Data was acquired on three different scanners with five different scanning sequences. The 354 subjects included in the study were classified in three groups; Healthy controls (HC), individuals with mild cognitive impairment (MCI), and patients with AD. Patients labeled AD met the clinical criteria for probable AD according to [16, 17]. Patients labeled MCI met the criteria stipulated by [19]. One site used three different protocols; the other two sites acquired the images using a single protocol. No diagnostic labels were provided for the test data. Demographic data are shown on Table 1.

Table 1. Demographic data of the training, validation, and test set. HC: healthycontrols, MCI: mild cognitive impairment, AD: Alzheimer's disease, F: female, M: male,N.A.: information was not available

HC/MCI/AD	F/M	age [years]
371/631/287	582/707	73.7 ± 7.3
79/31/30	80/60	74.5 ± 7.4
12/9/9	13/17	65.2 ± 7.0
N.A.	141/213	65.1 ± 7.8
	HC/MCI/AD 371/631/287 79/31/30 12/9/9 N.A.	HC/MCI/ADF/M371/631/287582/70779/31/3080/6012/9/913/17N.A.141/213

2.2 Validation Data

The validation dataset - also provided by the CADDementia Challenge - consisted of thirty examples that also included class labels. Acquisition sites, parameters, and inclusion criteria were identical to the test data set.

¹ http://caddementia.grand-challenge.org

2.3 Training Data

Structural MRI from baseline scans of the Alzheimer's Disease Neuroimaging Initiative (ADNI) database² [18] and from the Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL) database³ [8] were used. A goal of ADNI has been to test whether serial magnetic resonance imaging, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. The AIBL database consists of several hundred structural scans that were acquired on a single scanner. The study methodology has been reported previously [8]. We used baseline images from about 1429 subjects from AIBL and ADNI. Images from the ADNI were removed if either the subject converted or reverted during the course of the study. A demographic summary of the training data can be found on Table 1.

Four our study, individuals in the training data set were classified in AD, mild cognitive impairment (MCI), and healthy controls (HC).

3 Methods

3.1 Image pre-processing

The goal of image pre-processing was to obtain the input data for the automated classification process. We extracted very high-dimensional GM intensity maps for voxel-wise classification. Pre-processing of images was identical for training, validation, and test sets. Initially, the raw images were coregistered to the canonical T1 template in SPM8 using a rigid registration implemented in the SPM8 toolbox⁴. Then, using VBM8 toolbox⁵, we computed voxel-wise densities of gray matter (GM) that were normalized to a reference space. Maps were subsequently modulated by the determinant of the Jacobian of the local deformation field. The modulation thus accounted for non-linear volume changes, but ignored global (affine) volume changes. If multiple baseline images were available for a subject, the mean of all available GM maps was taken. The initial registration failed in some cases, which required manually registering the images to the template. Thus, the employed method was semi automatic, although the manual intervention was minor and did not require expert knowledge. The image preprocessing per image took about one hour on a single core. Computation of one column of the kernel matrix, which included computing pair-wise dot-products between GM maps, took a couple of seconds. Manual registration (required in approximately 10% of the test cases) took a few minutes per subject.

² http://adni.loni.usc.edu

³ http://www.aibl.csiro.au

⁴ http://www.fil.ion.ucl.ac.uk/spm/software/spm8/

⁵ http://dbm.neuro.uni-jena.de/vbm

3.2 Feature Selection

In previous studies [12, 3, 14, 2, 1] we used a linear SVM in combination with high-dimensional GM density maps including either all voxels or selecting voxels *a priori*. Here, we used the linear binary classifier with a hard margin in order to make the feature selection. Using 20% of the training data, one model for every binary classification was trained. For each classifier, using the method proposed by Gaonkar and Davatzikos [9] the *p*-values of the weights were computed and features were included only if the *p*-value was lower than a certain threshold. The threshold was 0.0001, 0.001, and 0.01 for ADvsHC, MCIvsNC, and ADvsMCI, respectively. The kernel computed from the subset of significant features was then used for the multi-class classification as explained in the next subsection. Computation for training a model and performing feature selection required about one minute of computation time, provided that the kernel matrix was computed and all required data was in the memory.

3.3 Nuisance Correction

We used kernel regression to correct for confounding effects such as age, sex and total intracranial volume to remove confounds from the linear dot-product matrix of the gray matter values. Computation time for this step was smaller than one second and thus was negligible compared to the time that was required for the pre-processing. Given the kernel matrix $\mathbf{K} \in \mathbb{R}^{N \times N}$, the detrended kernel $\tilde{\mathbf{K}}$ was computed as

$$\tilde{\mathbf{K}} = \mathbf{R}\mathbf{K}\mathbf{R}^{T}, \quad \mathbf{R} = \mathbf{I} - \mathbf{X}\left(\mathbf{X}^{T}\mathbf{X}\right)^{-1}\mathbf{X}^{T},$$
(1)

where **I** was the identity matrix and $\mathbf{X} \in \mathbb{R}^{N \times 3}$ the design matrix of N subjects coding sex, age, and total intra-cranial volume. This method was the same as previously proposed by Dukart et *al.* [7], but uses sex and TIV as additional covariate and performs the detrending in kernel space.

Unlike in previous work [14], we did not correct for scanner/sequence for this study, since not enough training data was available. Specifically, only about four images of healthy controls per scanner/sequence were available in the validation set. Of note, the age distribution in the training and test sets differed significantly (p < 0.05, Student's t-test). Since age, and AD both are associated with neuronal degeneration in partially overlapping regions, we expected a bias due to age. Specifically, we expected a lower sensitivity, because AD progression is positively correlated with age [7]. Thus younger subjects are less likely to be classified as AD.

3.4 Multi-class Classification

Multi-class linear classification in the one-versus-one setting is not well suited for classification of controls, MCI, and AD, because MCI is in between controls and MCI. We therefore employed a non-linear SVM with radial basis function $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma ||\mathbf{x}_i - \mathbf{x}_j||)$. For classification, we used one versus all support vector machine (SVM) as implemented in libsvm [5] with the option for probabilistic multi-class outputs [21]. The SVM parameters were set manually to C = 2 and $\gamma = 0.0002$. These combination achieved similar performance on the training and validation set (Figure 1). The performance on the validation set was obtained by using the predictions by the model trained using all training examples that were not used to estimate *p*-values for feature selection. The performance on the training set was computed in a ten fold cross-validation.

4 Results

Classification results were obtained for the training set (n = 1429) by crossvalidation, and on the validation set (n = 30) by applying the model trained on the entire training set. Test accuracy of binary classification of the validation (training) set of HC, MCI, and AD versus rest was 41.7% (62.1%), 66.7% (64.8%), and 77.8% (70.2%), respectively. Discriminability of the validation set in terms of AUC was highest for AD (96.8%), intermediate for HC (84.7%), and lowest for MCI (67.7%), as shown in Figure 1. The same order was observed in training set as well. Predictions on the test set were submitted to the



Fig. 1. Performance curves of training and validation set.

CADDementia committee.

5 Discussion

Training and test/validation sets differed in scanner hardware, acquisition parameters and inclusion criteria. Furthermore, the populations of controls and

patients were possibly more distinct in the training set, as conversion and reversion lead to exclusion. In addition, the test population was significantly younger than the test population. The correction for age effects and multi-centric studies conducted previously [3, 14], suggest that the most relevant factor that could lead to a discrepancy in cross-validated training performance versus test performance are the class-wise difference in population.

As expected, classification performance of the three classes HC, and AD was well above chance on the train and validation set. Discriminating MCI from the rest was more certain. Binary classification of HC and AD subjects reached up to 90 % accuracy. These results were obtained with an (almost fully) automated processing pipeline, which required no expert knowledge in the classification process.

One drawback of the presented methods is, that the classification process used the same features for all classification tasks. Although MCI can be seen as pre-state of AD, the optimally discriminative features between HC and MCI are not necessarily the same as the optimally discriminative features between MCI and AD or between HC and AD.

The SVM, as discriminative method, performed well in many similar classification tasks that were evaluated by cross-validation. In the present setting, the validation set remained entirely untouched. This reduced the risk of overfitting the model. However, since the parameters were hand-tuned and picked in such a way that the cross-validated performance on the training set was similar to the validation performance, there was a risk of overfitting to the validation set. We therefore expect a slightly lower performance on the test set.

6 Acknowledgments

This work was funded by a grant from the Deutsche Forschungsgemeinschaft (KL2415/2-1) to SK and OR and supported by the by the Excellence Initiative of the German Federal and State Governments (EXC 294) to OR. Data collection and sharing of the ADNI data set was funded by the Alzheimer's Disease Neuroimaging Initiative (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012).

References

- Abdulkadir, A., Mortamet, B., Vemuri, P., Jack Jr., C.R., Krüger, G., Klöppel, S.: Effects of hardware heterogeneity on the performance of SVM Alzheimer's disease classifier. NeuroImage 58(3), 785–792 (2011)
- Abdulkadir, A., Ronneberger, O., Christian Wolf, R., Pfleiderer, B., Saft, C., Klöppel, S.: Functional and Structural MRI Biomarkers to Detect Pre-Clinical Neurodegeneration. Current Alzheimer Research 10(2), 125–134 (2013)
- Abdulkadir, A., Ronneberger, O., Tabrizi, S.J., Klöppel, S.: Reduction of confounding effects with voxel-wise gaussian process regression in structural mri. In: Pattern Recognition in NeuroImaging (PRNI), 2014 International Workshop on (2014)

- Ashburner, J., Friston, K.J.: Voxel-based morphometry The methods. NeuroImage 11(6), 805–821 (2000)
- 5. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2(3), 27–27 (2011)
- 6. Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O., Alzheimer's Disease Neuroimaging Initiative, Alzheimer's Disease Neuroimaging Initiative: Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. NeuroImage 56(2), 766–781 (2011)
- Dukart, J., Schroeter, M.L., Mueller, K., Alzheimer's Disease Neuroimaging Initiative: Age Correction in Dementia – Matching to a Healthy Brain. PloS one 6(7), e22193 (2011)
- Ellis, K.A., Bush, A.I., Darby, D., De Fazio, D., Foster, J., Hudson, P., Lautenschlager, N.T., Lenzo, N., Martins, R.N., Maruff, P., Masters, C., Milner, A., Pike, K., Rowe, C., Savage, G., Szoeke, C., Taddei, K., Villemagne, V., Woodward, M., Ames, D.: The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. International Psychogeriatrics 21(04), 672–687 (2009)
- Gaonkar, B., Davatzikos, C.: Analytic estimation of statistical significance maps for support vector machine based multi-variate image analysis and classification. NeuroImage 78, 270–283 (2013)
- Gili, T., Cercignani, M., Serra, L., Perri, R., Giove, F., Maraviglia, B., Caltagirone, C., Bozzali, M.: Regional brain atrophy and functional disconnection across Alzheimer's disease evolution. Journal of Neurology, Neurosurgery & Psychiatry 82(1), 58–66 (2011)
- Klöppel, S., Abdulkadir, A., Jack Jr, C.R., Koutsouleris, N.: Diagnostic neuroimaging across diseases. Neuroimage (2012)
- Klöppel, S., Stonnington, C.M., Barnes, J., Chen, F., Chu, C., Good, C.D., Mader, I., Mitchell, L.A., Patel, A.C., Roberts, C.C., Fox, N.C., Jack, C.R., Ashburner, J., Frackowiak, R.S.J.: Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. Brain 131(Pt 11), 2969–2974 (2008)
- Klöppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack, C.R., Ashburner, J., Frackowiak, R.S.J.: Automatic classification of MR scans in Alzheimer's disease. Brain 131(Pt 3), 681–689 (2008)
- Kostro, D., Abdulkadir, A., Durr, A., Roos, R., Leavitt, B.R.: Correction of interscanner and within-subject variance in structural MRI based automated diagnosing. NeuroImage (2014)
- Lemm, S., Lemm, S., Blankertz, B., Blankertz, B., Dickhaus, T., Dickhaus, T., Müller, K.R., Müller, K.R.: Introduction to machine learning for brain imaging. NeuroImage 56(2), 387–399 (2011)
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M.: Clinical-Diagnosis of Alzheimers-Disease - Report of the Nincds-Adrda Work Group Under the Auspices of Department-of-Health-and-Human-Services Task-Force on Alzheimers-Disease. Neurology 34(7), 939–944 (1984)
- McKhann, G.M., Knopman, D.S., Chertkow, H.: The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia (2011)

- Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L.: The Alzheimer's Disease Neuroimaging Initiative. Neuroimaging Clinics of North America 15(4), 869–877 (2005)
- Petersen, R.C.: Mild cognitive impairment as a diagnostic entity. Journal of Internal Medicine 256(3), 183–194 (2004)
- Stonnington, C.M., Tan, G., Klöppel, S., Chu, C., Draganski, B., Jack, C.R., Chen, K., Ashburner, J., Frackowiak, R.S.J.: Interpreting scan data acquired from multiple scanners: a study with Alzheimer's disease. NeuroImage 39(3), 1180–1185 (2008)
- 21. Wu, T.F., Lin, C.J., Weng, R.C.: Probability estimates for multi-class classification by pairwise coupling. Journal of Machine Learning Research (2004)

PRISMA-CAD: Fully automathed method for Computer-Aided Diagnosis of Dementia based on structural MRI data

N. Amoroso^{1,2}, R. Errico^{2,3}, and R. Bellotti^{1,2}

¹ Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Italy

 $^{2}\,$ Dipartimento Interateneo di Fisica, Università degli Studi di Bari, Italy

 $^{3}\,$ Istituto Nazionale di Fisica Nucleare, Sezione di Genova, Italy

for the Alzheimer's Disease Neuroimaging Initiative **

Abstract. Neurodegenerative association with structural changes of the brain has been widely investigated gaining profound knowledge on specific aspects of the healthy and diseased brain. Medial temporal lobe atrophy and, in particular, the hippocampal atrophy are important biomarkers for the Alzheimer's disease. In this paper we describe how MRI brain scans can be processed and analyzed, in a fully automated framework, to segment relevant anatomical structures, extract morphometric and statistical features and perform an accurate clinical classification on the basis of anatomical and statistical features. We trained an artificial neural on a population consisting of 288 subjects to discriminate normal control subjects (NC), from those affected by Alzheimer's disease (AD) and mild cognitive impairment (MCI) with a one versus one strategy. Performances were validated with k-fold procedure, NC-AD were discriminated with accuracy ACC (NC-AD) = 0.91 while the overall accuracy ACC (NC-MCI-AD) reached the 0.81 value.

1 Introduction

Neuroscience is generating exponentially growing volumes of data and knowledge on specific aspects of the healthy and diseased brain, in different species, at different ages. However, there is no effective strategy to experimentally map the brain across all its levels and functions, yet. A proof of interest in the field is the recent funding of worldwide initiatives, such as the Human Brain Project ⁴ and the Human Connectome Project ⁵1.

^{**} Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

 $^{^4}$ www.humanbrain project.eu

⁵ www.humanconnectomeproject.org

Medical image computing raises new challenges related to the scale and complexity of the required analyses. For example, magnetic resonance imaging (MRI) of the brain plays a fundamental role for detection of neurodegeneration. The manual segmentation on MRI has been so far considered the only available strategy to accurately access reliable structural biomarkers and therefore to achieve a sound quantitative clinical discrimination. Nevertheless, manual segmentation is a time-consuming task nor it can manage the intrinsic human intra-rater variability, this is why automated processing pipelines are needed as diagnosis support systems.

In the present paper a novel fully automated processing workflow is described. It consists of three main steps. Firstly the pre-processing, an automated rigid registration and histogram based equalization for spatial and intensity normalization. Then, a volume of interest (VOI) extraction is performed; this VOI individuates a gross region containing the left and right hippocampi, from this region important features as the hippocampal volume or its thickness are calculated. Finally, the classification (NC - MCI - AD) is obtained with an automated artificial neural network.

2 Materials

The goal of this work is to provide a fully automated and reliable diagnosis support system to discriminate NC - MCI - AD. The CADDementia challenge aims to compare several methods and protocols to unveil, on the basis of a common test set whether significant differences exist among the various algorithms. According to this a standardized evaluation framework is set up, consisting of 384 multi-center scans. The participating centers are: Erasmus MC (EMC), Rotterdam, the Netherlands; VU University Medical Center (VUmc), Amsterdam, the Netherlands; University of Porto / Hospital de São João (UP), Porto, Portugal This data set contains structural MRI (T1w) scans of subjects with the diagnosis of probable Alzheimer's disease (AD), mild cognitive impairment (MCI) and participants without a dementia syndrome (controls). In addition to the MR scans, demographic information (age, gender) and information on which data are from the same institute is included.

To reach this goal 30 MRI brain scan are provided by the MICCAI CADDementia challenge (http://caddementia.grand-challenge.org) for training. Nevertheless, an increased basis of knowledge should help classification to build more generalized model and this is why a second dataset consisting of 258 MRI brain scans shared by the Alzheimer's Disease NeuroImaging Initiative (ADNI) was used. The two training databases used are described with demographics given in table 1.

3 Methods

In this study a fully automated pattern recognition system for accurate and reproducible segmentation of the hippocampus and the peri-hippocampal region

Data	Size	Age	M/F	Subjects
ADNI	258	60 - 96	144/114	96 NC - 96 MCI - 66 AD
MICCAI	30	54 - 80	17/13	12 NC - 9 MCI - 9 AD

Table 1. Data demographics. Group size, range age (years) and resolution of the two clinical datasets, containing normal control (NC) subjects, Alzheimer's Disease (AD) and mild cognitive impairment (MCI) patients.

in structural Magnetic Resonance Imaging (MRI) was used. This procedure, described in detail in our previous works in [1,2,3,4], is schematically shown in figure 1.



Fig. 1. The figure represents the overall processing pipeline. The MRI scans are processed independently by FreeSurfer and an automated segmentation pipeline hippocampus-focused. The calculated features are used to train an artificial neural network in a 5-fold cross validation frameowrk, then finally classification is performed.

The system consisted of three processing levels: (a) MRI brain scans were linearly registered to the standard MNI152 template and an automated point distribution shape analysis method was used to define the peri-hippocampal region. (b) Feature extraction: the peri-hippocampal VOI was statistically analyzed; gray level distribution features such as means, standard deviations, kurtosis and skewness were calculated. Moreover, other morphometric hippocampal based features such as the whole volume, the thickness, or local geometric features were calculated. In addition other features were calculated with a publicly available brain segmentation package FreeSurfer v.5.1 ⁶ [5] for an overall amount of 248 features. (c) Subject classification: a back propagation neural network was used to classify examples as NC, MCI and AD. An unsupervised filter was used to explore the feature space and determine correlations and linear dependences, in this way a subsample of about 150 features was determined. Features were normalized with the intra-cranial volume and finally normalized then, a one versus all strategy was adopted for training the network.

The network architecture was kept as simple as possible to avoid over-training issues, just one hidden layer with 10 neurons was used and a regularized cost function was adopted. To improve the generalization of the trained model a 5-fold strategy was used, in addition a random sampling of 50 features for every cross-validation round was performed. We repeated this procedure for a hundred times, thus obtaining 1500 trained networks. For every cross-validation round a 288×3 score matrix is obtained. Training performances are obtained by averaging the score matrix for every cross-validation and then averaging the class probabilities obtained from the different classifiers, in fact according to the one versus we trained a NC vs AD classifier, a NC vs MCI classifier and an AD vs MCI classifier, thus for example for NC subjects two distinct probabilities were given for each cross-validation step.

3.1 Computational infrastructure

The analyses presented in this paper were developed in MATLAB framework and required substantial computational resources. The previously described automated processing pipeline required an overall processing time of about 13 hours for subject, this processing time was almost entirely due to FreeSurfer. In fact, the processing time required to extract the hippocampal and the statistical features did not exceed one hour per subject. Therefore the use of dedicated workflow manager such as the LONI pipeline processing environment [6][7] was used: a user-friendly and efficient software for complex data analyses, available at http://pipeline.loni.ucla.edu and an adequate distributed infrastructure was of fundamental importance.

The analyses were carried out using the local computer farm BC2S ⁷: a distributed computing infrastructure consisting of about 5000 CPU and allowing up to 1,8 PB storage. A further study for grid deployment was also performed, within the aim of creating a pipeline tool suitable for large clinical trials. It was carried out on the European Grid Infrastructure (EGI) which consists of about 300 geographically distributed sites around the world. The run-time reduction with the grid implementation allowed to produce results in a reasonable time with respect to the application execution as a sequential process on limited resources. The advantages of the grid execution were evident since we obtained the 90% of the analysis of 642 images after less than 16 hours.

 $^{^{6}}$ freesurfer.nmr.mgh.harvard.edu

⁷ http://www.recas-pon.ba.infn.it

4 Results

The overall training results showed a significant discrimination among the three populations. Performances were measured in terms of accuracy (ACC) and area under the receiver operating characteristic (AUC). Figure **??** shows the overall training results.



Fig. 2. The figure represents the receiver operating characteristics of the three classes: NC (blue), MCI (green) and AD (red) for both the overall training set (on the left) and the 30 MICCAI images (on the right). AUC is also reported for all of them, it can be seen how NC and AD are recognized slightly better than MCI.

Performances on the reduced training set (the MICCAI data) resulted significantly lower, with an average accuracy $A\bar{C}C = 0.67 \pm 0.3$. However the data size does not allow to draw statistically significant conclusions 3.



Fig. 3. The figure shows the confusion matrix relative to the overall training (on the left) and the 30 MICCAI training images (on the right). Classes 1-2-3 are respectively the NC, the MCI and the AD classes.

The networked already trained were then used to obtain the test predictions. To improve generalization, we randomly sampled 300 classifiers from the 1500 already trained (100 for each type: NC-AD, NC-MCI and AD-MCI) and obtained for test example a prediction to be NC, MCI and AD. As previously explained for training, the average class probability was reported as the final classification score.

5 Discussion and Conclusion

Accuracy and area under the curve results suggest the method is reliable, besides being fully automated it can be adopted for large studies without suffering of intra-rater variability nor requesting time-intensive manual work from experts.

Classification performances compare well state-of-the-art performances thus suggesting the overall analysis workflow is reliable. The reduced number of features used for classification also suggest the possibility to significantly improve performances with the individuation of new features. The most important features for classification resulted to be those correlated to temporal lobe and hippocampal atrophy, on one hand this demonstrates that the volumes obtained by the proposed workflow are reliable, on the other it should also suggest to explore and investigate new features to improve the clinical discrimination.

Unsupervised approaches, such as deep learning networks, could be naturally included in the proposed framework and will be investigated in future works. Besides, those issues deriving from the computational burden yielded by FreeSurfer should also be addressed, the most promising strategy, according to our results, could be the individuation of statistical features which could substitute those obtained by FreeSurfer.

Acknowledgments

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California - San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

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 W81XW H -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: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. 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 Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

This research was supported by Istituto Nazionale di Fisica Nucleare (INFN), Italy. This research was also supported by grants from Università degli Studi di Bari, Italy.

All authors disclose any actual or potential conflicts of interest, including any financial, personal, or other relationships with other people or organizations that could inappropriately influence their work. All experiments were performed with the informed consent of each participant or caregiver in line with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Local institutional ethics committees approved the study.

References

- Amoroso, N., Bellotti, R., Bruno, S., Chincarini, A., Logroscino, G., Tangaro, S., Tateo, A.: Automated Shape Analysis landmarks detection: Model and Developments. Proceedings of the International Symposium, CompIMAGE (2012) 139–142
- Tangaro, S., Amoroso, N., Bruno, S., Chincarini, A., Frisoni, G.B., Maglietta, R., Tateo, A., Bellotti, R.: Active Learning Machines for Automatic Segmentation of Hippocampus in MRI. Industrial Conference on Data Mining (ICDM 2013). LECTURE NOTES IN COMPUTER SCIENCE (2013)
- Maglietta, R., Amoroso, N., Bruno, S., Chincarini, A., Frisoni, G.B., Inglese, P., Tangaro, S., Tateo, A., Bellotti, R.: Random forest classification for hippocampal segmentation in 3D MR images. IMLA 2013 conference inproceedings (2013)
- Tangaro, S., Amoroso, N., Boccardi, P., Bruno, S., Chincarini, A., Ferraro, G., Frisoni, G.B., Maglietta, R., Redolfi, A., Rei, L., Tateo, A., Bellotti, R.: Automated voxel-by-voxel tissue classification for hippocampal segmentation: Methods and Validation. in press (2014)
- 5. Fischl, B.: FreeSurfer. Neuroimage 62 (2012) 774-781

- 6. Dinov, I.D., Lozev, K., Petrosyan, P., Liu, Z., Eggert, P., Pierce, J., Zamanyan, A., Chakrapani, S., Van Horn, J.D., Parker, D.S., Magsipoc, R., Leung, K., Gutman, B., Woods, R.P., Toga, A.W.: Neuroimaging Study Designs, Computational Analyses and Data Provenance Using the LONI Pipeline. PLoS ONE 5 9
- Rex, D.E., Ma, J.Q., Toga, A.W.: The LONI Pipeline Processing Environment. Neuroimage 19 (2003) 1033–1048

CADDementia based on structural MRI using Supervised Kernel-based Representations

D. Cárdenas-Peña, A. Álvarez-Meza, and G. Castellanos-Dominguez

Signal Processing and Recognition Group Universidad Nacional de Colombia Campus La Nubia, km 7 via al Magdalena, Manizales-Colombia {dcardenasp, amalvarezme, cgcastellanosd}@unal.edu.co

Abstract. Magnetic Resonance Image (MRI) analysis allows to find patterns on physiological and pathological processes. Nevertheless, extracting MRI information posses a challenge due to high-dimensional voxel-wise feature spaces. As an alternative, a new supervised kernelbased representation approach for MRI discrimination is proposed. In this sense, the shape features of brain structures is highlighted by the inherent inter-slice similarities of a 3D MRI volume. For testing the proposed approach, an SVM classifier is trained using the ADNI dataset and tested on the CADDementia images for dementia diagnosis. Attained classification results (86% average accuracy on ADNI and 83% on CADDementia) prove that our fully automatic methodology is able to discriminate dementia patients.

1 Introduction

The use of Brain Magnetic Resonance Images (MRI) allows analyzing the influence of physiological and pathological processes on structural or functional properties of brain regions [1]. Specifically, for Alzheimer's Disease (AD), which is the leading form of dementia, it is well-known that structures as grey matter and hippocampus tend to be reduced [6]. Such facts have raised the interest on the development automatic MRI discrimination tools for supporting the AD diagnosis.

Some approaches in MRI discrimination are the following: In [2], an introduced mean shift algorithm is employed to perform atlas stratification to determine whether each population is best represented the considered multi-modal distribution. In [1], a ranked atlas selection is performed by computing image similarities among subject images based on measures like sums of squared differences (SSD), cross-correlation, or mutual information. In any case, those estimators do not guarantee convergence due to the involved highly-dimensional spaces.

Here, to improve MRI discrimination, we propose a new kernel-based representation based on the computed inherent Inter-Slice Kernel (ISK) relationship that makes prominent brain structure distributions. Specifically, we compare three different types of ISK-based feature representation to estimate pairwise MRI similarities using generalized Euclidean metrics. We tune all needed metric parameters by means of a centered alignment approach, so that the obtained kernels resemble the most prior demographic information [4,3]. The proposed approach is tested on MRI data discrimination using dementia categories (namely, Normal Control (NC), Late Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD)).

2 Proposed Algorithm Description

2.1 MRI Representation based on Inter-Slice Similarities

A 3D Magnetic Resonance Image (MRI) volume comprises a spatially structured set of intensity voxels $\Psi = \{x_r \in \mathbb{R} : r = (i, j, k)\}$, where x_r is the magnetic field intensity measured at location $r \in \mathbb{R} \subset \mathbb{N}^3$. Provided this spatial structure, we describe the MRI volume as an ordered set of 2D slices along the axis views as:

$$\Psi = \{ \boldsymbol{X}_i^v \in \mathbb{R}^{L_{v'} \times L_{v''}} : i \in \{1, \cdots, L_v\} \}$$
(1)

where v is each one of the considered axes, noted as: axial, a, sagittal - s, and coronal - c, i indexes the slices, L_v corresponds to the volume size in the considered axis and $L_{v'}$, $L_{v''}$ are volume sizes in the remaining axes. The arrangement in Eq (1) provides a useful way to analyze MRIs by medical specialists since to read information on the whole 3D volume is harder than on a single 2D slice. So, we take advantage of this introduced slice view and propose the use of the Inter-Slice Kernel (ISK) to encode pairwise similarities of the image slice set in the vector, $s^v \in \mathbb{R}^{P_v}$, with elements described as: $s_{ij}^v = \kappa_{\mathbf{X}} \left\{ d_{\mathbf{X}} \left(\mathbf{X}_i^v, \mathbf{X}_j^v \right) : \forall i < j \right\}$, where $d_{\mathbf{X}} : \mathbb{R}^{L'_v \times L_v''} \times \mathbb{R}^{L'_v \times L_v''} \to \mathbb{R}$ is a used distance operator for implementing the positive definite kernel function $\kappa_{\mathbf{X}} \{\cdot\}$, and $P_v = L_v (L_v - 1)/2$, so s^v becomes the ISK representation of the image Ψ along each axis v. It is worth noting that the ISK representation becomes much smaller than the original image space, i.e., $L_v (L_v - 1)/2 \ll L_v L_{v'} L_{v''}$.

2.2 Learning MRI Similarities by Kernel Centered Alignment

We establish an MRI Similarity Kernel (MSK), $\mathbf{K}^{v} \in \mathbb{R}^{N \times N}$ from a set of MRI volumes $\{\Psi_{n} : n \in \{1, \ldots, N\}\}$, that is the ISK matrix version representing high-dimensional image information along the axes, where N is the number of considered MRIs. Specifically, we perform MRI similarities, for every axis v, by computing each pairwise relationship, $k_{nm}^{v} \in \mathbf{K}^{v}$, between the ISK-based features as:

$$k_{nm}^{v} = \kappa_{\boldsymbol{s}} \left\{ \mathrm{d}_{\mathrm{SA}} \left(\boldsymbol{s}_{n}^{v}, \boldsymbol{s}_{m}^{v} \right) \right\} : n, m \in \{1, \dots, N\}$$

$$(2)$$

where $d_{SA} : \mathbb{R}^{P_v} \times \mathbb{R}^{P_v} \to \mathbb{R}$ is a certain a distance operator implementing the positive definite kernel function $\kappa_s\{\cdot\}$. In order to reveal the main ISK relationships for learning MRI similarities, we rely on the Mahalanobis distance defined in P_v -dimensional space with inverse covariance matrix $\mathbf{A}^v \mathbf{A}^{v\top}$ as:

$$d_{\mathrm{SA}}^{2}(\boldsymbol{s}_{n}^{v},\boldsymbol{s}_{m}^{v}) = (\boldsymbol{s}_{n}^{v} - \boldsymbol{s}_{m}^{v}) \boldsymbol{A}^{v} \boldsymbol{A}^{v \top} (\boldsymbol{s}_{n}^{v} - \boldsymbol{s}_{m}^{v})^{\top}.$$
(3)

where matrix $\mathbf{A}^{v} \in \mathbb{R}^{P_{v} \times D_{v}}$ holds the linear projection $\mathbf{\nu}_{n}^{v} = \mathbf{s}_{n}^{v} \mathbf{A}^{v}$, with $\mathbf{\nu}_{n}^{v} \in \mathbb{R}^{D_{v}}$ and $D_{v} \leq P_{v}$. Moreover, we propose to learn the matrix \mathbf{A}^{v} based on the already estimated ISK-based feature similarities and by adding prior subject diagnosis information enclosed in the matrix $\mathbf{B} \in \mathbb{R}^{N \times N}$. Thus, we measure the dependence between both matrices \mathbf{K}^{v} and \mathbf{B} through the following kernel target centered alignment function [3,4]:

$$\rho\left(\boldsymbol{K}^{v},\boldsymbol{B}\right) = \frac{\langle \boldsymbol{H}\boldsymbol{K}^{v}\boldsymbol{H},\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}\rangle_{F}}{\left\|\boldsymbol{H}\boldsymbol{K}^{v}\boldsymbol{H}\right\|_{F}\left\|\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}\right\|_{F}},\,\rho\in\left[0,1\right]$$
(4)

where $\mathbf{H} = \mathbf{I} - N^{-1} \mathbf{1} \mathbf{1}^{\top}$, with $\mathbf{H} \in \mathbb{R}^{N \times N}$, is a centering matrix, $\mathbf{1} \in \mathbb{R}^{N}$ is an all-ones vector, and notations $\langle \cdot, \cdot \rangle_{F}$ and $\|\cdot, \cdot\|_{F}$ stand for the Frobenius inner product and norm, respectively. Generally, the centered version of the alignment coefficient in Eq. (4) gets better correlation estimates than its uncentered version [4,3].

Therefore, we propose to learn MRI similarities from ISK-based features taking advantage of the Kernel Center Alignment (KCA) cost function described in Eq. (4). In this sense, prior patient information, e.g., demographic data as age and gender, can be employed to reveal MRIs dependencies by learning the matrix \mathbf{A}^v that parameterizes a Mahalanobis distance between pairwise images (see Eq. (3)). Thereby, given a demographic-based similarity matrix \mathbf{B} , a KCA-based function can be formulated to compute the projection matrix \mathbf{A}^v in Eq. (2) as:

$$\boldsymbol{A}^{v*} = \operatorname{argmax}_{\boldsymbol{A}^{v}} \rho\left(\boldsymbol{K}_{\boldsymbol{A}^{v}}^{v}, \boldsymbol{B}\right), \tag{5}$$

where $K_{A^v}^v$ is the resulting MSK matrix for a provided A^v projection as given in Eq. (2). Consequently, we term each $K_{A^{v*}}^v$ as a Learned MRI Similarity Matrix (LMSK).

2.3 MRI Discrimination using LMSK

The proposed LMSK is learned from an MRI Mahalanobis distance as in Eq. (4). With regard to the needed kernel functions, because of its universal approximating capability [5], we choose the well-known Gaussian kernel noted as follows:

$$g \{ d_z(z, z'); \sigma \} \triangleq \exp \left(- d_z(z, z')^2 / (2\sigma^2) \right),$$

where $\sigma \in \mathbb{R}^+$ is the kernel bandwidth; $z, z' \in \mathbb{Z}$ is a sample pair in a given feature space \mathbb{Z} , and $d_z : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ is a distance operator in \mathbb{Z} . In this sense, we calculate each ISK-based feature vector s^v from MRI using the Frobenious norm:

$$s_{ij}^{v} = g\left(||\boldsymbol{X}_{i}^{v} - \boldsymbol{X}_{j}^{v}||_{F}; \sigma_{s_{v}}\right).$$

$$(6)$$

Afterwards, we calculate each \mathbf{K}^{v} matrix encoding pairwise MRI relationship as in Eq. (2), yielding:

$$k_{nm}^{v} = g\left(\mathrm{d}_{\mathrm{SA}}(\boldsymbol{s}_{n}^{v}, \boldsymbol{s}_{m}^{v}); \sigma_{SA_{v}}\right).$$

$$\tag{7}$$

We optimize the KCA-based cost function in Eq. (5) to learn A^v by a gradient descent solver, where the initial feasible solution is calculated by the Principal Component Analysis algorithm. In addition, the elements of the label kernel B are set as: $b_{nm} = \delta (c_n - c_m)$, being δ the delta function and where $c_n \in \{1, 2, \ldots, C\}$ is the label of Ψ_n . Namely, three classes are considered from the CADDementia Challenge dataset: Alzheimer's disease (AD), late mild cognitive impairment (MCI) and healthy controls (NC). It is worth noting that every kernel bandwidth in Eqns. (6) and (7) must be properly tuned. Since the variability of the Gaussian kernel $g(\cdot; \sigma)$ tends to zero whenever the kernel bandwidth tends to either zero or infinity to get an appropriate σ value spanning widely all similarity values, we propose to adjust the Gaussian kernel bandwidth employing the following criterion (Notation var(\cdot) stands for the variance operator):

$$\sigma^* = \arg\max_{\sigma} \left\{ \operatorname{var}(g(\cdot; \sigma)) \right\}$$
(8)

3 Dataset and Preprocessing

For training the proposed MRI discrimination approach, the ADNI dataset was employed. Specifically, a subset of 451 3T MRI volumes are considered from subjects aged from 55 to 90 years (148 NC, 205 MCI, and 98 AD). Provided images are bias filtered using the well-known N3 algorithm. As a further preprocessing stage, each image is registered to the MNI305 template by an affine transform to reference the whole dataset to the Talairach space.

4 Results

Figs. 1(a) to 1(c) show a concrete MRI example illustrating all three views. As seen in Figs. 1(d) to 1(e) displaying their corresponding estimated ISK representations, the red corner patches keep the MRI edges with no content, i.e., the background. Moreover, as the Sagittal ISK (see Fig. 1(e)) exhibits symmetry respect to the anti-diagonal, it is clear that such representation is able to keep the head sagittal symmetry. Therefore, due to the kernel shape varies accordingly to the brain structure distribution, we infer that proposed ISK suitably characterizes head shapes.

Using the above proposed feature extraction stage, a new MRI similarity matrix, comparing each pair of subjects, is learned using the LMSK approach. The prior diagnosis information matrix \boldsymbol{B} and the resulting similarity matrix $\boldsymbol{K}^{\boldsymbol{v}}$ are depicted in Fig. 2(a) and Fig. 2(b) using the coronal axis view (v = c), respectively. From both Figures, it can be seen how the resulting kernel resembles the most the prior label matrix. For the sake of visualization, resulting LMSK is decomposed using the well-known PCA and the first three eigen-components are shown in Fig. 2(c), where the three considered classes can be clearly identified.

For the sake of evaluating the proposed approach a Support-Vector-Machinebased classifier is trained over the LMSK representation using the whole ADNI dataset and tested on the CADDementia MRIs. Obtained class-wise Receiver



Fig. 1: Database subject and their estimated ISK representation per view



Fig. 2: LMSK results

Operating Characteristic (ROC) curve for both datasets is depicted in Fig. 3, while the confusion matrices are shown in Table 1. The Fig. 3 shows a lower area under the curve for the second class, as the Table 1 shows the lowest accuracy for that class. Both facts imply that MCI subjects are the most difficult to classify, which can be due to the wide spread class distribution (see green subjects in Fig. 2(c)). From a morphological perspective, the low accuracy in MCI subjects can be related to nature of such class. Since MCI is an intermediate class between Healthy and Alzheimer's Disease classes, those subjects tend to be more misdiagnosed than the ones belonging to NC and AD.

5 Conclusion

A new supervised kernel-based image representation is introduced for automatic MRI-based discrimination of dementia. The proposed approach encodes interslice similarities, which are related to the shape features of brain structures.



Fig. 3: Obtained ROC curve for training MRIs in the ADNI and CADDementia datasets.

Table 1: Dementia d	lassificati	on accuracy [%]
	(b)	CADDementia
(a) ADNI 86%	83.3%	

(a) ADNI 86%			83.3%				
\mathbf{Class}	\mathbf{CN}	MCI	AD	Class	\mathbf{CN}	MCI	AD
CN	87.8	9.8	4.1	CN	100	0	0
MCI	7.4	81.5	7.1	MCI	0	0.66	0.34
AD	4.8	8.7	88.8	AD	0	22.2	77.8

Table 2: Computation time per subject of all methodology stages

\mathbf{Stage}	Time [s]
Registration (Rigid)	21.60 ± 7.74
Feature extraction (LMSK)	0.05 ± 0.01
Classification (SVM)	0.65 ± 0.01
Total	22.3

Furthermore, an SVM is trained using the LMSK representation for classifying three dementia categories (NC, MCI and AD). Taking into account the obtained results over the ADNI and CADDementia datasets, our proposed representation proves to find the natural inherent distributions of MRI. The methodology achieves 86% average classification accuracy on the ADNI training set and 83% on the CADDementia, using the coronal axis view. It is important to highlight that all stages on the current methodology are fully automatic. As future work, we plan to evaluate axis view combination strategies and the inclusion of patient demographic data for enhancing the representation and class separability.

References

- Aljabar, P., Heckemann, R.a., Hammers, a., Hajnal, J.V., Rueckert, D.: Multiatlas based segmentation of brain images: atlas selection and its effect on accuracy. NeuroImage 46(3) (July 2009) 726-38
- Blezek, D.J., Miller, J.V.: Atlas stratification. Medical image analysis 11(5) (October 2007) 443–57
- 3. Brockmeier, A., Choi, J., Kriminger, E., Francis, J., Principe, J.: Neural decoding with kernel-based metric learning. Neural Computation **26** (2014) –
- Cortes, C., Mohri, M., Rostamizadeh, A.: Algorithms for learning kernels based on centered alignment. The Journal of Machine Learning Research 13 (2012) 795–828
- 5. Liu, W., Principe, J.C., Haykin, S.: Kernel Adaptive Filtering: A Comprehensive Introduction. Volume 57. John Wiley & Sons (2011)
- Rao, A., Lee, Y., Gass, A., Monsch, A.: Classification of Alzheimer's Disease from Structural MRI using Sparse Logistic Regression with Optional Spatial Regularization. Evaluation (3) (2011) 4499-4502

Classification of Alzheimer's disease using structural MRI

Chester V. Dolph, Manar D. Samad, Khan M. Iftekharuddin

Vision Lab, Old Dominion University, Norfolk, VA 23529

Abstract. We propose an Alzheimer's disease (AD) classification method that uses selected features in segmented brain tissue from individual MRI slices. The objective of our method is to aid physicians in classifying AD progression, particularly for subjects with Mild Cognitive Impairment. The proposed method captures atrophy of regions of the brain without extracting specific regions such as the hippocampus. All subjects are registered to an atlas designed for AD research obtained from Laboratory of Neuro Imaging (LONI) at University of Southern California (USC). Each slice is segmented into grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF) to obtain the number of instances of these tissues per slice. The features include the number of instances of a segmented tissue in a given slice and the ratio of WM to CSF. The proposed method uses existing tools for skull stripping, registration, segmentation, and feature classification. A 10-fold cross-validation on CADDementia training data of 30 MRI samples yields 80% classification accuracy in classifying the 3 different cognitive states such as Cognitive Normal (CN), Mild Cognitive Impairment (MCI), and Alzheimer's disease (AD). Our results show that all AD and CN patients are classified correctly, however, subjects with MCI are misclassified mainly as the CN class. Although a larger training data can enhance the classification result, we use this limited training data to obtain our test results for 354 patients for this challenge.

1 Introduction

The progression of AD is of great interest in medical research as 1 in 3 seniors in the United States dies with dementia [1]. Based on signs and symptoms, physicians usually track AD using the Clinical Dementia Rating (CDR) system. Using CDR, subjects are classified in three states such as CN, MCI, and AD. The progression of AD can be characterized by atrophy of GM and WM along with expansion of CSF volume. These structural changes in brain facilitate the distinction of an AD brain from a CN brain; however, the distinction between MCI and CN is subtle [2]. Previous studies on classifying AD from MRI data include features based on voxel intensity, volume, regional, and thickness of subcortical areas [3]. Considering individual voxels of entire volume of MRI data is computationally expensive. On the other hand, focusing on a particular region of the brain may miss important structural changes happening in other regions of the brain.



Fig. 1. Proposed algorithm pipeline.

Other methods include segmentation of the hippocampus [4]. For the MCI class, where the changes are subtle, considering only a particular region of the brain may not be helpful in classification. Rather than using individual voxels or a particular brain region, we consider total number of pixels per slice from the entire MRI volume for the purpose of feature extraction and classification of the 3 different cognitive states. From each slice, the proposed method is based on the total number of pixels per slice classified as GM, WM, and CSF and their proportional relationships. The benefit of this approach is dimensionality reduction where the structural changes in the entire MRI data are represented in a feature dimension equal to the number of slices.

Since AD progresses differently among subjects, our approach captures features from selected MRI slices while accounting for tissue atrophy. Starting with all the slices to probe for structural changes in different tissues, this work applies a trial and error method to select the slices that best represent the structural changes for the best performance of 3-class classification. The search for the best slices starts from the mid region of the brain where the slices encompass most of the brain tissues.

2 Methods

2.1 Training data

The training data used for this work is obtained from the non-uniformity corrected data from the CADDementia website [5]. This data set consists of 30 subjects structural MRI labeled AD, MCI, or CN. This training data set contains MR image data labeled for 9 AD, 9 MCI, and 12 CN patient classes.

2.2 Algorithm description

The data processing pipeline is shown in Figure 1. The pipeline consists of five major steps such as skull stripping, registration, segmentation, feature extraction from slices, and classification. The steps of the pipeline are discussed in details in the following paragraphs.

Data pre-processing Prior to feature extraction, the following pre-processing steps are employed within the pipeline.

Skull stripping The MRI from each subject is skull stripped using the Brainsuite software tool [6]. In general, default parameters are used in the skull stripping process. However, if any scalp remains after stripping with default parameters, the diffusion and iteration parameters are adjusted as necessary to remove the remaining skull.

Registration Subject MRI scans are registered to the Alzheimer's disease atlas obtained from the LONI at USC using Deformable Registration via Attribute Matching and Mutual-Saliency Weighting (DRAMMS) [7].

Segmentation The brain volumes are segmented into GM, WM, and CSF using the Brainsuite classification tool [8].

2.3 Feature Extraction

Feature extraction: The features for our classifier are based on the segmented tissues. GM and WM atrophy is expected of the AD class and to a less extent the MCI class, while CSF volume simultaneously expands . We sum each segmented class for each MRI slice to create features that capture atrophy of that specific segmented brain tissue class. The sum represents the total number of instances that a segmented tissue is classified as GM, WM, or CSF. For example, a slice could be described as having 100 pixels, which might be further decomposed as 40 GM pixels, 35 WM pixels, and 25 CSF pixels after the tissue segmentation step. Since we are expecting fewer instances of GM and slightly more instances of CSF for a typical AD subject compared to a CN subject, the ratio of total pixels for GM (given as slice GM) to total number of pixels for CSF (slice CSF) should yield a larger number than the corresponding ratio of a cognitively normal subject. We propose the features given as,

$$\frac{\text{slice WM}}{\text{slice CSF}}, \quad \frac{\text{slice GM}}{\text{slice CSF}}, \quad \text{slice CSF}, \quad \text{slice WM}, \quad \text{slice GM},$$
(1)

where slice CSF represents total number of pixel for tissue CSF in an MRI slice and so on. A feature selection step is involved in order to select the MRI slices that yield the best classification results. For 3-class classification, different multiclass classifiers such as support vector machine SVM (with both Polynomial, and RBF kernels) and random forest are evaluated for accuracy on the training data. A 10-fold cross-validation is performed to find the best classifier and the feature using the training data set. We use the combination of best performing feature and classifier to classify the 354 test MRI samples for this challenge.

3 Results and discussions

3.1 Algorithm performance

Computation of skull stripping and segmentation takes approximately 1 minute per subject, which is mainly the time to navigate on the graphical user (GUI) interface of Brainsuite tool. Registration time varies for subjects from about 15 to 20 minutes each for each volume. The registration is completed serially on an i7 processor with 8 gigabytes of RAM. Increasing the memory capacity of the machine has the potential of speedup, as currently, there is excessive memory swap. Analysis of the segmented tissues is performed serially in Matlab on a MacBook with i5 processor and 8 gigabytes of RAM. This process is parallel, thus it has potential for significant speedup. Since the proposed feature dimension is only 46 per subject, it took only several seconds to classify the test data with 354 subjects.

The proposed algorithm is semi-automatic. The first few pre-processing steps such as skull stripping and segmentation need manual intervention only to adjust some parameters using the Brainsuite software tool. The Brainsuite GUI requires only a few parameter modifications per subject for skull stripping. These adjustments typically require 30 seconds of time, however, roughly 45 subjects required 15 minutes of parameter selection to remove the entire scalp. The subject registration is batch scripted in DRAMMS and thus completely automatic. The remaining steps including per slice analysis of the segmented tissues, feature extraction, and classification are all automated as well. Overall, the proposed pipeline takes about 30 minutes per subject.

4 Classification performance

An MR image of a subject's brain is a 3D volumetric data consisting of 128 slices. Following Eq. (1), each feature value represents the ratio of a pair of tissue content in each slice. From pattern recognition perspective, not all slices may contain discriminatory information that may be useful for 3-class classification. There may be slices with redundant or irrelevant information, which may introduce ambiguity in different class boundaries under classification. Following a trial and error step, 46 slices from slice index 25 to 70 for each subject MR data are found to be the most effective set of slices for the proposed 3-class classification. Figure 2 shows the range of the slices in the 3D volume. Fig. 2 shows that the slices containing relevant subcortical areas such as hippocampus plays major role in the proposed classification scheme.

The features proposed in Section 2.2.2 are extracted and used from these 46 slices. For each type of feature, there are 46 attributes per subject. Therefore, for 30 subjects in the given training data set, the training data dimension is 30 by 46. 3-class classification is performed by training 3 possible binary classifiers. Results from these binary classifiers are combined to yield the final multi-class classification result. A 10-fold cross-validation is performed to evaluate the performance of our proposed features in classifying 3-class MR images of brain.



Fig. 2. Location of the cross-sectional slices used for classification.

Before testing a fold of data, the binary classifiers are trained by the remaining 9-fold data. The final 10-fold classification accuracy is obtained by averaging the accuracy from 10 individual folds of test data.

Table 1 shows the confusion matrix after 10-fold cross-validation using WM/CSF feature and SVM with RBF kernel (SVM-RBF) classifier. Note that all the AD and CN classes are correctly classified. However, MCI class is mostly confused with CN class. The 10-fold cross-validation result shows 80% classification accuracy wherein 24 out of 30 samples are correctly classified in multi-class classification. Table 2 shows area under the ROC curve for 3 different classes and different classifiers. For classification, an SVM-RBF classifier is found to yield the best

Classified As $->$	AD	MCI	CN
AD	9	0	0
MCI	1	3	5
$_{\rm CN}$	0	0	12

Table 1. Confusion matrix for training data set after 10-fold cross-validation.

classification accuracy. Out of five different features, we find that (WM/CSF) feature performs the best in classifying 3 classes. Therefore, we consider the same SVM-RBF classifier and WM/CSF feature to classify the test data set.

The absolute proportion of WM, GM, and CSF yields poor classification accuracy. This is understandable, since different subjects are likely to have different proportion of tissues regardless of their cognitive status. Therefore, to have a representative feature for a cognitive status regardless of the subject, the ratio of different tissues within a subject can be a subject independent attribute. Since subjects with cognitive impairment suffer atrophy in WM or GM with expansion in CSF region, a ratio between WM or GM with CSF can be representative feature for such classification. This is supported by our results shown above,
Classifier	AD	MCI	CN	Average
SVM-RBF	0.976	0.762	0.861	0.866
SVM-Poly	0.881	0.574	0.88	0.788
Random Forest	0868	0.624	0.861	0.792

Table 2. Area under the ROC curve for each class

which reveals a significant improvement in the classification performance using (WM/CSF) feature.

5 Conclusion

We eagerly await the classification accuracy of the testing data set, so that conclusions can be made on the performance of this proposed approach. The classification accuracy also depends on the success and consistency in pre-processing steps as well as on the training data size. Furthermore, the relatively poor classification between MCI and CN is due to the subtle differences between these two groups. We believe that a larger training data set will yield an improved performance in the classifying the test samples of MR images from 354 subjects.

Acknowledgments

This work is partially supported through grants funding from NIH (R15CA115464). Authors would like to thank Dr. Christos Davatzikos for his kind assistance on the DRAMMS tool.

References

- 1. Alzheimer's Association, 2013. Alzheimer's News 3/19/13, http://www.alz.org/news_and_events_facts_and_figures_report.asp.
- Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehricy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O., 2011. Automatic classification of patients with Alzheimers disease from structural MRI: A comparison of ten methods using the ADNI database. Neuroimage 56, 766781.
- Klöppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack Jr., C.R., Ashburner, J., Frackowiak, R.S.J., 2008. Automatic classification of MR scans in Alzheimer's disease. Brain 131 (3), 681689.
- 4. Chupin, M., Hammers, A., Liu, R.S., Colliot, O., Burdett, J., Bardinet, E., Duncan, J.S., Garnero, L., Lemieux, L., 2009a. Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. Neuroimage 46 (3), 749 761.
- 5. A standardized evaluation framework for Computer-Aided Diagnosis of Dementia based on structural MRI data, http://caddementia.grand-challenge.org/.

- Shattuck DW and Leahy RM (2002) BrainSuite: An Automated Cortical Surface Identification Tool Medical Image Analysis 8(2):129-142.
- DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting. Yangming Ou, Aristeidis Sotiras, Nikos Paragios, Christos Davatzikos. Medical Image Analysis, 15(4): 622-639, 2011.
- Shattuck DW, Sandor-Leahy SR, Schaper KA, Rottenberg DA, and Leahy RM (2001) Magnetic Resonance Image Tissue Classification Using a Partial Volume Model NeuroImage 13(5):856-876.

Detecting Alzheimer's disease by morphological MRI using hippocampal grading and cortical thickness

Simon F. Eskildsen¹, Pierrick Coupé³, Vladimir Fonov², D. Louis Collins² and the Alzheimer's Disease Neuroimaging Initiative

¹Center of Functionally Integrative Neuroscience, Aarhus University, Aarhus, Denmark. ²McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Canada. ³Laboratoire Bordelais de Recherche en Informatique, Unité Mixte de Recherche CNRS (UMR 5800), PICTURA Group, Bordeaux, France.

Abstract. Structural MRI is an important imaging biomarker in Alzheimer's disease as the cerebral atrophy has been shown to closely correlate with cognitive symptoms. Recognizing this, numerous methods have been developed for quantifying the disease related atrophy from MRI over the past decades. Special effort has been dedicated to separate AD related modifications from normal aging for the purpose of early detection and prediction. Several groups have reported promising results using automatic methods; however, it is very difficult to compare these methods due to varying cohorts and different validation frameworks. To address this issue, the public challenge on Computer-Aided Diagnosis of Dementia based on structural MRI data (CADDementia) was proposed. The challenge calls for accurate classification of 354 MRI scans collected among AD patients, subjects with mild cognitive impairment and cognitively normal control. The true diagnosis is hidden from the participating groups, thus making the validation truly objective. This paper describes our proposed method to automatically classify the challenge data along with a validation on 30 scans with known diagnosis also provided for the challenge.

1 Introduction

Neuronal injury is an important component of the pathophysiological process involved in Alzheimer's disease (AD). The cerebral atrophy resulting from the progressive neurodegeneration can be measured using magnetic resonance imaging (MRI) and is currently among the most important biomarkers of AD [1]. Optimizing such MRI-based biomarkers for detection and prediction of AD may have a significant impact on early diagnosis of patients as well as being valuable tools when designing therapeutic studies of individuals at risk of AD to prevent or alter the progression of the disease.

Hippocampal atrophy has long been recognized as an early feature of the degenerative process involved in AD [2]. Reductions in hippocampal volume appear to be correlated to early memory decline [3]. While sensitive to first stage of AD, hippocampal degeneration is involved in other dementias, such as vascular dementia [4], and is known to be part of non-pathological brain aging [5]. Thus, volumetric measurements of the hippocampus (HC) might be limited in their ability to predict the progression of AD [6-9]. Evidence suggests that the nature of degeneration in the HC and surrounding structures, such as the entorhinal cortex (ERC) and parahippocampal gyrus, is different in AD compared to other dementias and different from the changes occurring during normal aging [10]. We recently obtained results that support this finding since we showed that AD detection could be improved by considering the structural composition of the HC and its surrounding structures in the medial temporal lobe [8]. These results were obtained using a novel concept of measuring structural similarities, comparing the anatomy of a test subject to a library of AD patients and cognitive normal (CN) subjects.

Studies have shown that, apart from hippocampal and medial temporal lobe (MTL) atrophy, AD has a characteristic neocortical atrophy pattern [11, 12]. Cortical thinning of temporal and parietal lobe regions, the posterior cingulate and the precuneus seem to be involved at early stages of the disease [13]. In the advanced stages of the disease, atrophy spreads to almost the entire cortex sparing only the sensory-motor and visual cortex [14]. Recently, we showed [15] that if cortical thickness is measured in a consistent manner, patterns of cortical thinning can predict conversion to AD among mild cognitive impaired (MCI) subjects with higher accuracy.

In the current challenge the aim is to classify morphological MRI scans into the classes: AD, MCI and CN. For this purpose we propose to combine measurements of structural pathological patterns, measured by analyzing morphological alterations in key structures of the MTL with degenerative patterns of the neocortex, measured by cortical thickness.

2 Methods

2.1 Image data

Data used in the preparation of this article were obtained from the ADNI database (http://adni.loni.usc.edu/) and from the public challenge on Computer-Aided Diagno-Dementia based on structural MRI data (CADDementia, sis of http://caddementia.grand-challenge.org/). The ADNI database contains 1.5T and 3.0T T1w MRI scans for AD, MCI, and CN at several time points. For this study we used cohorts with 1.5T scans from the ADNI1 study and separate cohorts with 3.0T images from the ADNI2 study (see Table 1 for cohort statistics). The CADDementia data consist exclusively of 3.0T T1w MRI scans from three different sites in Europe. It should be noted that scanner equipment and protocols are not harmonized between the sites, which gives rise to additional variation in the data. In total 384 scans are provided of which 30 scans have known labels (Table 1). The remaining 354 scans are used for evaluating the proposed classification methods in a completely blinded fashion. The authors do not have access to the diagnostic labels behind the test scans.

Dataset	N (females)	Age±sd
ADNI1 AD	181 (88)	75.3±7.5
ADNI1 MCI	381 (139)	74.8±7.4
ADNI1 CN	222 (105)	75.9±5.0
ADNI2 AD	48 (16)	75.6±8.8
ADNI2 MCI	183 (69)	71.7±7.6
ADNI2 CN	73 (36)	75.6±6.2
CAD AD	9 (6)	66.1±5.2
CAD MCI	9 (4)	68.0±8.5
CAD CN	12 (3)	62.3±6.3
CAD test set	354 (141)	65.1±7.8

Table 1. Demographics of the datasets used in the analyses.

2.2 Image preprocessing

All images were processed using a fully automatic pipeline [16]. Images were denoised [17] using a Rician-adapted noise estimation [18], bias field corrected [19], and registered to MNI space using a 12 parameter affine transformation [20]. To enable robust registrations we used as registration target a population-specific template derived from the ADNI1 database constructed using a series of linear and non-linear registrations as described in [21]. The custom template was created from 50 AD patients and 50 CN subjects randomly selected. This template better reflects the anatomy of ADNI data compared to the conventional ICBM template build from young healthy adults. Image intensities were normalized to match the intensity profile of the template [22], and finally the images were skull stripped using BEaST [23].

2.3 Hippocampus and entorhinal cortex

Structural features of the hippocampal complex were estimated using SNIPE (Scoring by Nonlocal Image Patch Estimator) method [8, 24]. In this technique, the local structural information surrounding each voxel (i.e., 3D patch) of a test subject is compared to those in a training library of MRI datasets from ADNI1 and ADNI2 AD and CN subjects with segmentation of the considered structures (HC and ERC). In short, a small patch of MRI data around each voxel (e.g., 7x7x7 voxel patches) from the test subject is compared to the training library with the goal to find similar patches. The patch similarity is used to compute a weight for the match and used to determine two type of information. First, the weights are used to perform segmentation of both considered structures using label fusion strategy [25]. Second, the weights are used to obtain grading values reflecting the proximity of the current patch to both training populations (e.g., CN and AD) (see [24] for more details). As proposed in [24], the 50 closest subjects were first selected from each training population (i.e., 50

AD¹ and 50 CN) using SSD over an initialization mask. Then, the grading maps and the segmentations of the considered structures were obtained simultaneously using SNIPE. The ADNI1 and ADNI2 data were graded in a leave-one-out fashion, while the CADDementia data were graded using the entire ADNI1 and ADNI2 library. Finally, the 8 SNIPE-based features extracted were the average grading value over the left and right HC and ERC and the volume of the same four structures. Volumes were calculated in normalized space (MNI space).

2.4 Neocortex

Cortical thickness was calculated using FACE (Fast Accurate Cortex Extraction) [26-28] and mapped to the cortical surface of the population-specific average nonlinear anatomical template [21] using an iterative, feature-based algorithm [29]. MCI and CN subjects were used to generate a statistical map of group differences in cortical thickness. From this t-map, cortical thickness features were derived with the procedure described in [15] using the proportion of the cortical surface with the 15% largest t-values corresponding to a threshold of t=4.3 and t=1.0 for ADNI1 and ADNI2 respectively (see Figure 1). In brief, candidate ROIs were calculated using a multi-seed constrained surface based region growing algorithm initialized at local maxima of the 15% t-map. In [15] we found a cortical area of 10-15% to be suitable for searching for candidate ROIs. We used MCI/CN contrast to generate cortical features as opposed to the more conventional AD/CN contrast. This was to enable sensitivity to the three-way classification problem, acknowledging that the MCI classification is the most difficult task. A total of 90 cortical thickness ROIs were identified this way using ADNI1 data and 87 using ADNI2 (see Figure 2). Neocortical features comprised the mean cortical thickness within each of these ROIs measured in subject native space.

2.5 Classification

We used multinomial regression with lasso and L1/L2 elasticnet regularization for the classification. We applied the GLMNET matlab implementation for the purpose (http://www.stanford.edu/~hastie/glmnet_matlab/). During experiments we used ADNI1 or ADNI2 as training datasets. The proposed classification framework was based on ensemble learning approach [30]. In order to create the ensemble we used an iterative approach. For each iteration the following steps were performed:

1. Age correction based on the CN ADNI training population

As in [31], for each feature we estimated the age-related effect on the CN population using linear regression. The features for all the considered groups (ADNI and CAD groups) were then corrected using the estimated linear regression coefficients.

¹ For ADNI2 only 48 AD images were used in the segmentation and grading process.

2. Under-sampling and Over-sampling of the ADNI training dataset

In presence of imbalanced population sizes, the evaluation of the performance of classification algorithms might be biased [32]. Therefore, in this study, we used two different strategies to compensate for this bias. We first randomly under-sample majority classes (i.e., CN and MCI) to obtain similar number of subjects for the 3 classes. Then we used SMOTE (Synthetic Minority Oversampling Technique) to increase the number of samples in each classes. SMOTE creates new synthetic samples by performing interpolation of the nearby neighbors in the feature space

3. Grid search for optimal classifier parameters

In order to estimate the optimal parameters of the classifier, we performed a grid search using the CAD training dataset. When the accuracy of the classification was balanced between the 3 classes of CAD training dataset and higher than a given threshold, the model was applied on the CAD test dataset and the prediction (i.e., the 3 *a posteriori* probabilities for all the subjects) was stored in the model ensemble. Step 2 and 3 are iteratively performed until 25 good models were found.

4. Ensemble learning classification

Finally, at the end of the iterative process, all the selected models in the ensemble were fused using a non-weighted strategy. We estimated the mean of the 3 *a posteriori* probabilities for all the subjects over all the selected models. The maximum fused *a posteriori* probabilities were finally used to estimate the final label of the CAD test dataset.

This procedure is used in the four following scenarios:

- 1. Using ADNI1 as training dataset and SNIPE and FACE features (98 features in total)
- 2. Using ADNI1 as training dataset and SNIPE features (8 features)
- 3. Using ADNI2 as training dataset and SNIPE and FACE features (95 features in total)
- 4. Using ADNI2 as training dataset and SNIPE features (8 features)

Finally, the four scenarios were combined by using the grand mean of all a posterior probabilities from the scenarios. As in step four above, the maximum posterior probability was used to label the images.

Using the CADDementia training data (n=30), the correct classification rates of each of the three classes (AD/MCI/CN) were calculated for all five scenarios.



Fig. 1. Thresholded t-maps for ADNI1 MCI/CN contrast (left) and ADNI2 MCI/CN contrast (right). Notice the difference in statistical strength due to sample size differences.



Fig. 2. Cortical thickness ROIs generated by respectively ADNI1 cohorts (left) and ADNI2 cohorts (right).

3 Results

3.1 Computational time

The classification process is fully automatic. Using a single core (Intel Core i7 (@3.40Ghz) per subject the total computational time was approximately 55 minutes distributed on preprocessing (30 min), FACE (15 min), and SNIPE (10 min). Applying the classifier after it has been trained takes only a few seconds.

3.2 CADDementia training set

The final accuracy of the ensemble classifiers obtained on the CADDementia training set are presented in Table 2.

Table 2.	Classification	accuracies	of the	five	classifiers	when	applied	on	CADDementia	ı train-
ing data.	All numbers a	re in percen	tage (%	6).						

	Classification accuracy				
Classifier	CN	MCI	AD	Overall	
SNIPE/FACE ADNI1	75.0	66.7	66.7	70.0	
SNIPE ADNI1	75.0	77.8	77.8	76.7	
SNIPE/FACE ADNI2	58.3	66.7	77.8	66.7	
SNIPE ADNI2	66.7	66.7	77.8	70.0	
Combined	75.0	66.7	77.8	73.3	

4 Discussion

The results on the CADDementia training data may be inflated due to the grid search for optimal parameters. Nevertheless the results indicate the ranking of the different classifiers. It seems that training on ADNI1 data provides better results than training on ADNI2 data even though ADNI2 data should better represent the CADDementia data using only 3T images. The difference is most likely due to more available training data in the ADNI1 cohorts leading to well-defined classes. Thus the morphological variation of the three populations (AD/MCI/CN) is better represented in the larger sample. Adding cortical thickness features to the SNIPE features does not seem to improve results. In fact, it seems to impair the classifiers, possibly due to adding noise. Perhaps fewer and more carefully selected ROIs in the neocortex would have better complemented the MTL features and thus added discriminative information. Combining the four classifiers does not increase accuracy.

The AD group has consistently the highest (or tied with the highest) classification accuracy across all five classifiers. As expected the main difficulty in the three-way classification problem is to separate CNs and MCIs. A different strategy could have been to primarily focus on this problem by combining binary classifiers with the idea of "one class against all" applied in succession. Looking at the demographics there is a significant age difference between ADNI and CADDementia data. We addressed this issue by adjusting the features for linear age effects using the ADNI CN subjects. However, this may not sufficiently compensate for the age differences between the cohorts. Due to aging effects, such as demyelination of the brain tissues, the MRI signal is not independent of subject age. Thus features based on image intensity and texture, such as the SNIPE features, are affected by these properties of the MR signal. Cortical thickness estimates are also affected, however, most likely to a lesser extent. Similarly, differences in field strength between training and testing data will affect the classifier performance.

The sample size of the CADDementia training data is too small to be representative of the large morphological variation typically found in demented as well as healthy brains. Thus training on a separate dataset, in our case ADNI data, is absolutely necessary to obtain a well-balanced and unbiased classifier. When the underlying labels of the CADDementia testing data is revealed, we will know whether our strategy using ADNI data was the right choice.

Acknowledgements

The research was supported by the MINDLab UNIK initiative at Aarhus University, funded by the Danish Ministry of Science, Technology and Innovation, grant agreement number 09-065250. This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the Investments for the future Programme IdEx Bordeaux (ANR-10-IDEX-03-02), Cluster of excellence CPU and TRAIL (HR-DTI ANR-10-LABX-57).We acknowledge funding from Canadian Institutes of Health Research (MOP-111169) and les Fonds de Recherche Santé Québec.

References

1. Jack, C.R., Jr., Vemuri, P., Wiste, H.J., Weigand, S.D., Lesnick, T.G., Lowe, V., Kantarci, K., Bernstein, M.A., Senjem, M.L., Gunter, J.L., Boeve, B.F., Trojanowski, J.Q., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Knopman, D.S.: Shapes of the trajectories of 5 major biomarkers of Alzheimer disease. Archives of neurology 69, 856-867 (2012)

2. Ball, M.J., Fisman, M., Hachinski, V., Blume, W., Fox, A., Kral, V.A., Kirshen, A.J., Fox, H., Merskey, H.: A new definition of Alzheimer's disease: a hippocampal dementia. Lancet 1, 14-16 (1985)

3. De Leon, M., George, A., Stylopoulos, L., Smith, G., Miller, D.: EARLY MARKER FOR ALZHEIMER'S DISEASE: THE ATROPHIC HIPPOCAMPUS. The Lancet 334, 672-673 (1989)

4. Gainotti, G., Acciarri, A., Bizzarro, A., Marra, C., Masullo, C., Misciagna, S., Tartaglione, T., Valenza, A., Colosimo, C.: The role of brain infarcts and hippocampal atrophy in subcortical ischaemic vascular dementia. Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology 25, 192-197 (2004)

5. Driscoll, I., Hamilton, D.A., Petropoulos, H., Yeo, R.A., Brooks, W.M., Baumgartner, R.N., Sutherland, R.J.: The aging hippocampus: cognitive, biochemical and structural findings. Cereb Cortex 13, 1344-1351 (2003)

6. Chupin, M., Gerardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehericy, S., Benali, H., Garnero, L., Colliot, O.: Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. Hippocampus 19, 579-587 (2009)

7. Clerx, L., van Rossum, I.A., Burns, L., Knol, D.L., Scheltens, P., Verhey, F., Aalten, P., Lapuerta, P., van de Pol, L., van Schijndel, R., de Jong, R., Barkhof, F., Wolz, R., Rueckert, D., Bocchetta, M., Tsolaki, M., Nobili, F., Wahlund, L.-O., Minthon, L., Frölich, L., Hampel, H., Soininen, H., Visser, P.J.: Measurements of medial temporal lobe atrophy for prediction of Alzheimer's disease in subjects with mild cognitive impairment. Neurobiol Aging (2013)

8. Coupé, P., Eskildsen, S.F., Manjón, J.V., Fonov, V.S., Pruessner, J.C., Allard, M., Collins, D.L.: Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease. NeuroImage: Clinical 1, 141-152 (2012)

9. Wolz, R., Julkunen, V., Koikkalainen, J., Niskanen, E., Zhang, D.P., Rueckert, D., Soininen, H., Lotjonen, J.: Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. PLoS One 6, e25446 (2011)

10. Devanand, D.P., Bansal, R., Liu, J., Hao, X., Pradhaban, G., Peterson, B.S.: MRI hippocampal and entorhinal cortex mapping in predicting conversion to Alzheimer's disease. NeuroImage 60, 1622-1629 (2012)

11. Dickerson, B.C., Bakkour, A., Salat, D.H., Feczko, E., Pacheco, J., Greve, D.N., Grodstein, F., Wright, C.I., Blacker, D., Rosas, H.D., Sperling, R.A., Atri, A., Growdon, J.H., Hyman, B.T., Morris, J.C., Fischl, B., Buckner, R.L.: The Cortical Signature of Alzheimer's Disease: Regionally Specific Cortical Thinning Relates to Symptom Severity in Very Mild to Mild AD Dementia and is Detectable in Asymptomatic Amyloid-Positive Individuals. Cereb Cortex 19, 497-510 (2009)

12. McEvoy, L.K., Fennema-Notestine, C., Roddey, J.C., Hagler, D.J., Holland, D., Karow, D.S., Pung, C.J., Brewer, J.B., Dale, A.M.: Alzheimer Disease: Quantitative Structural Neuroimaging for Detection and Prediction of Clinical and Structural Changes in Mild Cognitive Impairment1. Radiology 251, 195-205 (2009)

13. Reiman, E.M., Jagust, W.J.: Brain imaging in the study of Alzheimer's disease. NeuroImage 61, 505-516 (2012)

14. Eskildsen, S.F., Fonov, V., Coupe, P., Collins, D.L.: Visualizing stages of cortical atrophy in progressive MCI from the ADNI cohort. In: Alzheimer's Association International Conference on Alzheimer's Disease. Alzheimer's & Dementia, (Year)

15. Eskildsen, S.F., Coupe, P., Garcia-Lorenzo, D., Fonov, V., Pruessner, J.C., Collins, D.L.: Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. NeuroImage 65C, 511-521 (2013)

16. Aubert-Broche, B., Fonov, V.S., Garcia-Lorenzo, D., Mouiha, A., Guizard, N., Coupe, P., Eskildsen, S.F., Collins, D.L.: A new method for structural volume analysis of longitudinal brain MRI data and its application in studying the growth trajectories of anatomical brain structures in childhood. NeuroImage 82C, 393-402 (2013)

17. Coupe, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C.: An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. IEEE Trans Med Imaging 27, 425-441 (2008)

18. Coupe, P., Manjon, J.V., Gedamu, E., Arnold, D., Robles, M., Collins, D.L.: Robust Rician noise estimation for MR images. Med Image Anal 14, 483-493 (2010)

19. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 17, 87-97 (1998)

20. Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C.: Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr 18, 192-205 (1994)

21. Fonov, V., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., Collins, D.L.: Unbiased average age-appropriate atlases for pediatric studies. NeuroImage 54, 313-327 (2011)

22. Nyul, L.G., Udupa, J.K.: Standardizing the MR image intensity scales: making MR intensities have tissue-specific meaning. pp. 496-504. SPIE, (Year)

23. Eskildsen, S.F., Coupe, P., Fonov, V., Manjon, J.V., Leung, K.K., Guizard, N., Wassef, S.N., Ostergaard, L.R., Collins, D.L.: BEaST: brain extraction based on nonlocal segmentation technique. NeuroImage 59, 2362-2373 (2012)

24. Coupe, P., Eskildsen, S.F., Manjon, J.V., Fonov, V.S., Collins, D.L.: Simultaneous segmentation and grading of anatomical structures for patient's classification: application to Alzheimer's disease. NeuroImage 59, 3736-3747 (2012)

25. Coupé, P., Manjón, J., Fonov, V., Pruessner, J., Robles, M., Collins, L.: Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. NeuroImage 54, 940-954 (2011)

26. Eskildsen, S.F., Ostergaard, L.R.: Quantitative comparison of two cortical surface extraction methods using MRI phantoms. Med Image Comput Comput Assist Interv 10, 409-416 (2007)

27. Eskildsen, S.F., Ostergaard, L.R.: Active surface approach for extraction of the human cerebral cortex from MRI. Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention 9, 823-830 (2006)

28. Eskildsen, S.F., Uldahl, M., Ostergaard, L.R.: Extraction of the cerebral cortical boundaries from MRI for measurement of cortical thickness. pp. 1400-1410. (Year)

29. Eskildsen, S.F., Ostergaard, L.R.: Evaluation of Five Algorithms for Mapping Brain Cortical Surfaces. Proceedings of the 2008 XXI Brazilian Symposium on Computer Graphics and Image Processing, pp. 137-144. IEEE Computer Society (2008)

30. Dietterich, T.: Ensemble Methods in Machine Learning. Multiple Classifier Systems, vol. 1857, pp. 1-15. Springer Berlin Heidelberg (2000)

31. Koikkalainen, J., Polonen, H., Mattila, J., van Gils, M., Soininen, H., Lotjonen, J.: Improved classification of Alzheimer's disease data via removal of nuisance variability. PLoS One 7, e31112 (2012)

32. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research 16, 321-357 (2002)

Dementia classification based on brain age estimation

Katja Franke¹, Christian Gaser¹

¹ Structural Brain Mapping Group, Departments of Neurology & Psychiatry, Jena University Hospital, Jena, Germany {katja.franke, christian.gaser}@uni-jena.de

Abstract. Early identification of neuroanatomical changes deviating from the normal agerelated atrophy pattern has the potential to improve clinical outcomes through early treatment or prophylaxis. Especially the pathological cascade of Alzheimer's disease (AD), the most common form of dementia, is widely linked to precocious and/or accelerated (brain) aging.

This work presents a novel magnetic resonance imaging (MRI)-based biomarker that indicates discrepancies in individual brain aging. By employing automatic preprocessing of structural MR images as well as high-dimensional pattern recognition methods, this approach uses the distribution of normal brain-aging patterns to estimate the individual brain age of a given new subject. The difference between the estimated brain age and the chronological age gives an individual deviation score, with positive values indicating the degree of acceleration in cerebral atrophy, which is considered a risk factor for AD. Here, this deviation score is used to classify the subjects as NO, MCI, and AD.

Keywords: MRI, relevance vector machines (RVM), support vector machines (SVM), regression, aging, brain disease

1 Introduction

The global prevalence of dementia is projected to rise sharply over the coming decades. By 2050, 1 in 85 persons worldwide will be affected by Alzheimer's disease (AD), the most common form of dementia (Brookmeyer et al. 2007). Manifold pathological changes begin to develop years or decades before the onset of cognitive decline (Jack et al. 2010), including premature changes in gene expression (Cao et al. 2010; Saetre et al. 2011), accelerated age-associated changes of the default mode network (Jones et al. 2011), and most obviously, abnormal changes in brain structures already at the mild cognitive impairment (MCI) stage (Driscoll et al. 2009; Spulber et al. 2010). Additionally, atrophic regions detected in AD patients were recently found to largely overlap with those regions showing a normal age-related decline in healthy control subjects (Dukart et al. 2011).

Early detection and quantification of abnormal brain changes is important for the prospective identification and subsequent treatment of individuals at risk for cognitive decline and dementia. At some point in the AD disease course accelerated neurodegeneration takes place, preceding accelerated cognitive decline (Jack et al. 2010). Although brain atrophy in general is not specific for AD, MRI-detected atrophy was found to retain the closest relationship with cognitive decline (Jack et al. 2010; Vemuri et al. 2009a, b) suggesting a crucial role for structural MRI in predicting future conversion to AD (Frisoni et al. 2010; Jack et al. 2010).

Assuming AD to be preceded by precocious and/or accelerated brain aging (Bartzokis 2011; Driscoll et al. 2009; Spulber et al. 2010), a straightforward and efficient solution is to model healthy aging on the one hand, and to identify accelerated (thus pathological) brain atrophy on the other. Consequently, in order to recognize faster brain atrophy, a model of healthy and normal brain aging is needed. A straightforward and efficient solution is to model age regression based on normal brain anatomy, such that an individual's brain age can be accurately estimated from his/her brain scan alone.

The approach presented here takes into account the widespread but sequential age-related brain tissue loss. Based on single time-point structural MRI the complex, multidimensional aging patterns across the whole brain are aggregated to one single value, i.e. the estimated brain age. Consequently, although using only a standard MRI scan, the deviation in brain atrophy from normal brain aging can be directly quantified.

2 Methods

2.1 Subjects/Database

To train the age estimation framework, MRI data from the publicly accessible IXI cohort (http://www.brain-development.org) were used. In September 2011, the IXI database contained T_1 -weighted images from 561 healthy subjects [250 male] aged 20-86 years [mean (SD) = 48.6 (16.5) years], which were collected on three different scanners (Philips 1.5T, General Electric 1.5T, Philips 3T). The distributions of age and gender within the training sample are shown in Figure 1. Table 1 shows the characteristics of the *CAD Dementia* training and test samples.



Figure 1: Age distribution in the male (left) and female (right) IXI sample used for modeling normal brain aging.

CAD Dementia	train s	sample	test sample		
	male	female	male	female	
No. subjects	17	13	213	141	
Age mean (SD)	64.2 (6.7)	66.5 (7.4)	65.7 (7.4)	64.2 (8.3)	
Age range	54 - 79	57 - 80	49 - 88	46-88	
No. NO subjects	9	3	-	-	
No. MCI subjects	5	4	-	-	
No. AD subjects	3	6	-	-	

Table 1: Characteristics of the subjects in the CAD Dementia training and test samples.

2.2 Preprocessing of MRI data and data reduction

Preprocessing of the T_1 -weighted images was done using the SPM8 package (http://www.fil.ion.ucl.ac.uk/spm) and the VBM8 toolbox (http://dbm.neuro.uni-jena.de), running under MATLAB. All T1-weighted images were corrected for bias-field inhomogeneities, then spatially normalized and segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) within the same generative model (Ashburner and Friston 2005). The segmentation procedure was extended by accounting for partial volume effects (Tohka et al. 2004), by applying adaptive maximum a posteriori estimations (Rajapakse et al. 1997), and by using a hidden Markov random field model (Cuadra et al. 2005; Gaser 2009). The images were processed with affine registration and smoothed with (*i*) 4-mm full-width-at-half-maximum (FWHW) or (*ii*) 8-mm FWHW smoothing kernels. Spatial resolution was set to (*i*) 4 mm or (*ii*) 8 mm, respectively. For further data reduction, principal component analysis (PCA) was performed on the training sample with subsequently applying the estimated transformation parameters to the test sample. PCA was done using the 'Matlab Toolbox for Dimensionality Reduction' (http://ict.ewi.tudelft.nl/~lvandermaaten/Home.html).

2.3 Relevance vector regression (RVR)

To capture the complex, multidimensional aging pattern across the whole brain, the brain age estimation framework utilizes relevance vector regression (RVR; Tipping, 2001). Relevance vector machines (RVM) were introduced as a Bayesian alternative to support vector machines (SVM) for obtaining sparse solutions to pattern recognition tasks. The main idea behind SVMs is the transformation of training data from input space into high-dimensional space - the feature space - via a mapping function Φ (Bennett and Campbell 2003; Schölkopf and Smola 2002). For the purpose of classification, the hyperplane that best separates the groups is computed within this feature space, resulting in a nonlinear decision boundary within the input space. The best separating hyperplane is found by maximizing the margin between the two groups. The data points lying on the margin boundaries are called support vectors since only these are used to specify the optimal separating hyperplane. For the case of real-valued output functions (rather than just binary outputs as used in classification), the SV algorithm was generalized to regression estimation (Bennett and Campbell 2003; Schölkopf and Smola 2002). In support vector regression (SVR), a function has to be found that fits as many data points as possible. Analogous to the margin in classification, the regression line is surrounded by a tube. Data points lying within that tube do not influence the course of the regression line. Data points lying on the edge or outside that tube are called support vectors.

In contrast to the support vectors in SVM, the relevance vectors in RVM represent the prototypical examples within the specified classification or regression task, instead of solely representing separating attributes. Furthermore, severe overfitting associated with the maximum likelihood estimation of the model parameters was avoided by imposing an explicit zero-mean Gaussian prior (Ghosh and Mujumdar 2008; Zheng et al. 2008). This prior is a characteristic feature of the RVM, and its use results in a vector of independent hyperparameters that reduces the data set (Faul and Tipping 2002; Tipping 2000; Tipping 2001; Tipping and Faul 2003). Therefore, in most cases the number of relevance vectors is much smaller than the number of support vectors. Furthermore, in SVR additional parameters have to be determined or statistically optimized (e.g. with cross-validation loops) in order to control for model complexity and model fit. To control the behavior of the RVR, only the type of kernel has to be chosen, whereas all other parameters are automatically estimated by the learning procedure itself. More details can be found in (Bishop 2006; Schölkopf and Smola 2002; Tipping 2000).

2.4 Brain age estimation framework

The brain age estimation framework utilizes RVR to capture the complex, multidimensional aging pattern across the whole brain and to subsequently estimate individual brain ages based on T1-weighted images. As suggested in Franke et al. (2010), the kernel was chosen to be a polynomial of degree 1, since age estimation accuracy was shown to not improve when choosing non-linear kernels. Thus, parameter optimization during the training procedure was not necessary.

In general, the model is trained with preprocessed whole brain structural MRI data (as described in 2.2) of the training sample, resulting in a complex model of healthy brain aging (Figure 2A, left panel). Put in other words, the algorithm uses those whole-brain MRI data from the training sample that represent the prototypical examples within the specified regression task (i.e., healthy brain aging). Additionally, voxel-specific weights are calculated that represent the importance of each voxel within the specified regression task (i.e., healthy brain aging). For an illustration of the most important features (i.e., the importance of voxel locations for regression with age) that were used by the RVR to model normal brain aging and more detailed information please refer to Franke et al. (2010).

Subsequently, the brain age of a test subject can be estimated using the individual tissue-classified MRI data, aggregating the complex, multidimensional aging pattern across the whole brain into one single value (Figure 2A, right panel). In other words, all the voxels of the test subject's MRI data are weighted by applying the voxel-specific weighting matrix. Then, the brain age is calculated by applying the regression pattern of healthy brain aging and aggregating all voxel-wise information across the whole brain. The difference between estimated brain age and the true chronological age will reveal an individual deviation score. Consequently, this deviation score directly quantifies the amount of acceleration or deceleration in individual brain aging (Figure 2B). For example, if a 70 years old individual has a deviation score of +5 years, this means that this individual shows the typical atrophy pattern of a 75 years old individual. To compute the final age regression model as well as to predict the

individual brain ages, we used the freely available toolbox "The Spider" (http://www.kyb.mpg.de/bs/people/spider/main.html).



Figure 2. (A) The model of healthy brain aging is trained with the chronological age and preprocessed structural MRI data of a training sample (left; with an exemplary illustration of the most important voxel locations that were used by the age regression model). Subsequently, the individual brain ages of previously unseen test subjects are estimated, based on their MRI data (blue; picture modified from Schölkopf and Smola (2002)). **(B)** The difference

between the estimated and chronological age results in the deviation (i.e., BrainAGE) score. Consequently, positive deviation scores indicate accelerated brain aging. (Image reproduced from Franke et al. (2012), with permission from Hogrefe Publishing, Bern).

The age estimation model was separately trained on male and female subjects in the IXI training sample. Furthermore it was trained using *(i)* preprocessed GM images, *(ii)* preprocessed WM images, and *(iii)* the linear combination of preprocessed GM and WM images. Subsequently, the brain ages of the subjects in the *CAD Dementia* training and test samples were estimated based on their non-uniformity corrected (nuc) MRI data. The difference between the estimated and the true age resulted in *(i)* GM, *(ii)* WM, and *(iii)* GMWM deviation scores, respectively. Sample-specific set-offs and slopes were corrected for the whole *CAD Dementia* sample using linear regression. The whole brain age estimation framework works automatically. All data preprocessing, model training, and brain age estimation was done using MATLAB.

2.5 Classification based on brain age estimation

Classification of the CAD Dementia subjects is based on the individual deviation scores. To explore the best classification accuracies in the CAD Dementia training sample, the brain age estimation model was run with all combinations of parameter variation during preprocessing (i.e. 4 mm & 8 mm FWHM smoothing kernel) and segmented brain tissue (i.e., GM, WM & GMWM). Subsequently, in each of those six brain age models two linear thresholds (i.e., NO vs. MCI and MCI vs. AD) were searched for to classify the subjects as NO, MCI, or AD based on individual deviation scores. Accuracy rates and receiver operating characteristics (ROC) for two-class classifications (i.e., NO vs. MCI, NO vs. AD, MCI vs. AD) were computed in the CAD Dementia training sample, resulting in the area under the ROC curve (AUC). Based on these results, the final brain age estimation models for classification were chosen. To differentiate between NO vs. MCI and MCI vs. AD the model using the linear combination of preprocessed GM and WM images with a 4 mm FWHM smoothing kernel showed best performances. To further refine the differentiation between NO vs. MCI we additionally used WM deviation scores (8 mm FWHM smoothing kernel). Then, overall classification accuracies were calculated following (Hand and Till 2001), as recommended in the CAD Dementia challenge rules. Subsequently, the final thresholds that reveal the highest overall classification accuracies in the CAD Dementia training sample were applied to classify the CAD Dementia test subjects.

3 Results

3.1 Performance of the age estimation model

To analyze the performance of the specific models used for training of the typical age patterns in healthy subjects, leave-one-out was used in the IXI training sample. For the GMWM age estimation

model (with 4 mm FWHM smoothing kernel) the correlation between estimated brain age and chronological age was r = 0.92 in the male [n= 250; mean (SD) age: 46.8 (16.4) years] as well as in the female subsample [n= 311; mean (SD) age: 48.6 (16.5) years]. The mean absolute error (MAE) was 5.1 years in both subsamples. For the WM age estimation model (with 8 mm FWHM smoothing kernel) the correlation between estimated brain age and chronological age was r = 0.88 in the male and r = 0.90 in the female subsample. The MAE in the male subsample was 6.4 years and 5.8 years in the female subsample.

3.2 Classification based on brain age estimation in the CAD Dementia training sample

Overall classification accuracies as calculated following (Hand and Till 2001), was 90% in the *CAD Dementia* training sample. Regarding the GMWM deviation score the thresholds are set at 2.6 to differentiate between NO vs. MCI and 4.5 to differentiate between MCI vs. AD. To further refine the differentiation between NO vs. MCI an additional threshold was set at -2.9 in the WM deviation score (Figure 3).



Figure 3: Classification scheme for the *CAD Dementia* training (colored dots) and test subjects (black dots) depending on the thresholds in WM and GMWM deviation scores. Subjects in the green area are classified as "NO", subjects in the blue area are classified as "MCI", and subjects in the red area are classified as "AD".

The whole approach is fully automatic. Preprocessing the MRI data (as described in 2.2) takes about 8 minutes per brain image. The whole process of training the brain age estimation models with about 560 subjects from the IXI sample, subsequent estimation of the brain ages for all 384 *CAD Dementia* subjects, and final classification of the *CAD Dementia* training subjects takes about 110 seconds in total on MAC OS X, Version 10.5.8, 2.4 GHz Intel Core 2 Duo.

4 Discussion

For estimating the brain age from T_1 -weighted MRI scans, we propose a fast and fully automatic framework that includes preprocessing of the images, dimension reduction via PCA, training of a RVM for regression with a polynomial kernel of degree 1, and finally estimating the brain age of the *CAD Dementia* subjects. This age estimating framework turns out to be a straightforward method to accurately and reliably estimate brain age with as little preprocessing and parameter optimization as possible. Using MRI data of about 560 healthy subjects aged between 20 and 86 and scanned on three

different scanners, the age estimation with RVR showed excellent performance, with an overall MAE of only 5 years and a correlation of r = 0.92 between the chronological and the estimated brain age.

Using structural MRI data, our fully automated brain age estimation model aggregates the complex, multidimensional aging patterns across the whole brain to one single value (i.e. the deviation score) and finally identifies subtle pathological brain aging in the *CAD Dementia* training MCI and AD subjects, with increasing deviation scores at indicating an increased risk for AD.

In conclusion, our brain age estimation framework could potentially help to recognize and indicate advanced brain atrophy on an individual level, thus contributing to an early diagnosis of neurodegenerative diseases like AD and facilitate early treatment or a preventative intervention.

References

Ashburner J, Friston KJ (2005) Unified segmentation. NeuroImage 26 (3):839-851

Bartzokis G (2011) Alzheimer's disease as homeostatic responses to age-related myelin breakdown. Neurobiology of Aging 32 (8):1341-1371

Bennett KP, Campbell C (2003) Support vector machines: hype or hallelujah? SIGKDD Explorations 2:1-13

Bishop CM (2006) Pattern Recognition and Machine Learning. Springer, New York, NY

Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM (2007) Forecasting the global burden of Alzheimer's disease. Alzheimers Dement 3 (3):186-191

Cao K, Chen-Plotkin AS, Plotkin JB, Wang LS (2010) Age-correlated gene expression in normal and neurodegenerative human brain tissues. PLoS One 5 (9):e13098

Cuadra MB, Cammoun L, Butz T, Cuisenaire O, Thiran JP (2005) Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images. IEEE Transactions on Medical Imaging 24 (12):1548-1565

Driscoll I, Davatzikos C, An Y, Wu X, Shen D, Kraut M, Resnick SM (2009) Longitudinal pattern of regional brain volume change differentiates normal aging from MCI. Neurology 72 (22):1906-1913

Dukart J, Schroeter ML, Mueller K (2011) Age correction in dementia--matching to a healthy brain. PLoS One 6 (7):e22193

Faul A, Tipping ME (2002) Analysis of sparse Bayesian learning. In: Dietterich TG, Becker S, Ghahramani Z (eds) Advances in Neural Information Processing Systems 14. MIT Press, pp 383–389

Franke K, Gaser C, for the Alzheimer's Disease Neuroimaging Initiative (2012) Longitudinal changes in individual *BrainAGE* in healthy aging, mild cognitive impairment, and Alzheimer's disease. GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry 25 (4):235 - 245

Franke K, Ziegler G, Klöppel S, Gaser C, the Alzheimer's Disease Neuroimaging Initiative (2010) Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. NeuroImage 50 (3):883-892

Frisoni GB, Fox NC, Jack CR, Jr., Scheltens P, Thompson PM (2010) The clinical use of structural MRI in Alzheimer disease. Nat Rev Neurol 6 (2):67-77

Gaser C (2009) Partial volume segmentation with Adaptive Maximum a Posteriori (MAP) approach. NeuroImage 47 (S1):S121

Ghosh S, Mujumdar PP (2008) Statistical downscaling of GCM simulations to streamflow using relevance vector machine. Advances in Water Resources 31 (1):132-146

Hand D, Till R (2001) A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Machine Learning 45 (2):171-186

Jack CR, Jr., Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, Petersen RC, Trojanowski JQ (2010) Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 9 (1):119-128

Jones DT, Machulda MM, Vemuri P, McDade EM, Zeng G, Senjem ML, Gunter JL, Przybelski SA, Avula RT, Knopman DS, Boeve BF, Petersen RC, Jack CR, Jr. (2011) Age-related changes in the default mode network are more advanced in Alzheimer disease. Neurology 77 (16):1524-1531

Rajapakse JC, Giedd JN, Rapoport JL (1997) Statistical approach to segmentation of single-channel cerebral MR images. IEEE Transactions on Medical Imaging 16 (2):176-186

Saetre P, Jazin E, Emilsson L (2011) Age-related changes in gene expression are accelerated in Alzheimer's disease. Synapse 65 (9):971-974

Schölkopf B, Smola A (2002) Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA

Spulber G, Niskanen E, MacDonald S, Smilovici O, Chen K, Reiman EM, Jauhiainen AM, Hallikainen M, Tervo S, Wahlund LO, Vanninen R, Kivipelto M, Soininen H (2010) Whole brain atrophy rate predicts progression from MCI to Alzheimer's disease. Neurobiology of Aging 31 (9):1601-1605

Tipping ME (2000) The Relevance Vector Machine. In: Solla SA, Leen TK, Müller K-R (eds) Advances in Neural Information Processing Systems 12. MIT Press, Cambridge, MA, pp 652-658

Tipping ME (2001) Sparse bayesian learning and the relevance vector machine. Journal of Machine Learning Research 1 (3):211-244

Tipping ME, Faul AC Fast marginal likelihood maximisation for sparse Bayesian models. In: Bishop CM, Frey BJ (eds) Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, Key West, FL, Jan 3-6 2003.

Tohka J, Zijdenbos A, Evans A (2004) Fast and robust parameter estimation for statistical partial volume models in brain MRI. NeuroImage 23 (1):84-97

Vemuri P, Wiste HJ, Weigand SD, Shaw LM, Trojanowski JQ, Weiner MW, Knopman DS, Petersen RC, Jack CR, Jr. (2009a) MRI and CSF biomarkers in normal, MCI, and AD subjects: diagnostic discrimination and cognitive correlations. Neurology 73 (4):287-293

Vemuri P, Wiste HJ, Weigand SD, Shaw LM, Trojanowski JQ, Weiner MW, Knopman DS, Petersen RC, Jack CR, Jr. (2009b) MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change. Neurology 73 (4):294-301

Zheng Y-T, Neo S-Y, Chua T-S, Tian Q (2008) Probabilistic optimized ranking for multimedia semantic concept detection via RVM. Paper presented at the Proceedings of the 2008 international conference on Content-based image and video retrieval, Niagara Falls, Canada,

Alzheimer's disease state classification using structural volumetry, cortical thickness and intensity features

Christian Ledig¹, Ricardo Guerrero¹, Tong Tong¹, Katherine Gray¹, Alexander Schmidt-Richberg¹, Antonios Makropoulos^{1,2}, Rolf A. Heckemann^{3,4}, Daniel Rueckert¹

¹Department of Computing, Imperial College London, UK ²Division of Imaging Sciences and Biomedical Engineering, King's College London, UK ³MedTech West, Institute of Neuroscience and Physiology, University of Gothenburg, Sweden ⁴Division of Brain Sciences, Faculty of Medicine, Imperial College London, UK

Abstract. Evaluating algorithms for the image-based classification of dementia on the basis of common data and using common metrics is essential for an objective comparison of different approaches. It is the aim of the Computer-Aided Diagnosis of Dementia based on structural MRI data (CADDementia) challenge to address this need by providing the opportunity of objectively evaluating individual approaches. In this paper, a classification framework is presented, in which four different sets of features extracted from structural MR images are examined with respect to their discriminative abilities. These features are based on volumetric and morphologic parameters, image intensities and patch similarities. Moreover, a combined feature set is employed to analyse the amount of complementary information contained in the features. For the three-class classification problem - Alzheimer's Disease (AD) vs. Mild Cognitive Impairment (MCI) vs. Control (CN) - a classification rate on a subset of the ADNI1-2 databases between 51% and 59% is achieved with all five feature sets. The combined feature set leveraging the potential of all four methods leads to only a minor improvement over the individual sets.

1 Introduction

The *Computer-Aided Diagnosis of Dementia based on structural MRI data* (CADDementia) challenge [1] is aimed at evaluating different methods for the image-based diagnosis of dementia. Held in the course of the MICCAI 2014 conference, the challenge provides a standardised platform for objectively testing different approaches on a common set of image data, which allows an objective comparison.

A crucial aspect of the computer-aided diagnosis based on MRI data is the extraction of features from the images. These features have to be meaningful for the diagnostic purpose, that means they are designed to contain valuable information about the state and the progression of the disease. In the case of Alzheimer's disease, for example, features proposed in the literature range from concrete and well-known clinical biomarkers like the hippocampal volume [2] or the cortical thickness [3] to abstract parameters derived directly from the image intensities by manifold learning approaches [4].

In this contribution, we aim at comparing different feature extraction methods that have

2 C. Ledig, R. Guerrero, T. Tong, K. Gray, et al.

all been shown to be of high potential for the differential diagnosis of Alzheimer's disease and related forms of dementia. These features comprise:

- Volumetric features (VOL): A total of 134 distinct brain volumes are automatically segmented (see Sec. 3.1).
- Morphologic features (CORT): Based on the brain segmentations, morphological features such as cortical thickness and cortical surface measurements are computed (see Sec. 3.2).
- Manifold-based learning features (MBL): Manifold-based learning is used to map intensity texture descriptors of all subjects into a *d*-dimensional space, such that similarities between the images are maintained. The *d* coordinates of this space (or a subset of them) are then considered as features (see Sec. 3.3).
- Patch-based grading features (GRAD): A patch-based approach is employed to find similar intensity patterns in scans of other subjects. Features are then computed as the weighted average of the labels associated with similar patches (see Sec. 3.4).

The special focus of this work is on A) comparing the performance of the features on a set of unseen image data provided by the CADDementia challenge, and B) assessing the degree of complementary information contained in the feature sets. Therefore, as well as analysing the four groups of features independently, a joint feature set (ALL) is tested that combines all available features.

For classification, Random Forests were employed consistently in all experiments as they have been shown to be powerful in particular for multi-class classification [5] (see Sec. 4). They were trained on the Alzheimer's Disease Neuroimaging Initiative (ADNI) [6] database using the five different feature sets and the classification results were submitted to the CADDementia challenge.

In Sec. 5, classification results for a 10-fold cross validation on the ADNI database and for the classification of the 30 training cases hosted by the CADDementia challenge are additionally provided. These experiments showed that a high performance of all feature sets for the classification of the Alzheimer's disease state, with three-class classification accuracies ranging between 51% and 59% for the cross validation. The intensity based features MBL and GRAD slightly outperformed the other features. Only a small improvement was reached with the combined feature set, which suggests that the amount of complementary information contained in the features is limited.

2 CADDementia data and common preprocessing

The challenge data is comprised of 30 plus 354 T1-weighted images for training and testing data respectively. The images were acquired at three different sites at 3T magnetic field strength. More details can be found at [1]. Both training and testing data were preprocessed using the same pipeline. All T1-weighted MR scans with in-place resolution below 0.5 mm were resampled and their in-plane resolution doubled. The images were further corrected for potential intensity inhomogeneities employing the N4 algorithm [7]. Brain masks were then calculated for both training and testing images using pincram [8]. As atlas database for pincram, 64 subjects of the ADNI1 cohort were

AD classification using volumetry, cortical thickness and manifold learning

3

chosen. After visual inspection $\approx 10\%$ of the brain masks for the CADDementia testing dataset did not meet our quality criterion. An updated pincram version was rerun on these subjects adding 3T scans from the ADNI2 study as well as successfully extracted subjects from the CADDementia training set to the atlas database. The extended database lead to an improved segmentation quality sufficient for a further analysis, such that no manual editing was required. It was observed that scans acquired at EMC Rotterdam were particularly challenging to extract.

3 Feature Extraction

In the following we present four approaches to extract biomarkers that have been shown to have potential for Alzheimer's disease state classification. The focus is on providing a brief description of the individual approaches and how they are applied. For further details we refer to the papers describing the whole brain segmentation approach (VOL, [9]), the cortical thickness measurement (CORT, [10]), the extraction of intensity features (MBL, [11]) and the calculation of patch-based grading values (GRAD, [12]).

3.1 Volumetric features from multi-structure whole brain segmentation (VOL)

Training data To automatically parcellate the provided MR scans into anatomical regions, the 30 brain atlases (excluding repeat scans) provided through the "MICCAI Grand Challenge on Multi-Atlas Labeling 2012" [13] were employed. This atlas database consists of 30 T1-weighted MR scans from the OASIS database that were annotated by expert raters¹ into 134 distinct brain regions.

Method description We employed the multi-atlas label propagation method described in [9]. In this approach, all 30 brain atlases are aligned to an unsegmented subject MRI using a robust nonrigid registration approach based on multi-level free form deformations [14–16]. The individual atlas label maps are then transformed to the image space of the unsegmented image using the calculated transformation and a nearest neighbour interpolation scheme. The 30 propagated atlas label maps are fused into a consensus probabilistic segmentation using a local weighting approach. The obtained probabilistic segmentation is further refined using a method that exploits image intensities in an expectation maximisation framework [9]. For each subject we finally extract 135 volumetric features, including background, based on the segmentations.

3.2 Cortical morphology features (CORT)

Training data Measurements of cortical morphology were obtained based on the segmented regions of the cortex. Cortical surface area, curvature and thickness features were calculated for the whole cortex and its 98 regional subdivisions (in total 591 features).

¹ provided by Neuromorphometrics, Inc. (http://Neuromorphometrics.com/) under academic subscription.

4 C. Ledig, R. Guerrero, T. Tong, K. Gray, et al.

Cortical thickness estimation Cortical thickness was estimated as described in [10]. In this approach, a potential field from the GM-WM to the GM-CSF interface is determined by solving the Laplace's equation. Voxel-wise thickness was then calculated as the sum of the voxel's distance to the WM and to the CSF, following the direction perpendicular to the potential field. Cortical sulci correction was performed similarly to [17]. We calculated cortical thickness (mean and standard deviation) for each cortical region (98+98 features).

Cortical surface measurements Cortical surface meshes were obtained by triangulation of the CGM-WM isosurface of each hemisphere with the marching cubes algorithm [18]. The surfaces were smoothed with Laplacian smoothing [19] for an even distribution of the mesh vertices. An example cortical surface is presented in Figure 1.



Fig. 1: Cortical surface of the CGM-WM interface with overlaid segmentation.

Surface area and curvature measures of the cortex were computed from the meshes. We adopted a number of area-independent curvature measures from [20] with T-normalisation that are not sensitive to the surface area. The surface measures included in this study were: surface area in the whole cortex and each region (1+98), relative surface area (98), mean curvature L^2 norm (1+98) and Gaussian curvature L^2 norm (1+98).

3.3 Manifold-based learning for multi-level variable selection (MBL)

Training data Data used was obtained from the ADNI database. In this work, a subset of 292 ADNI-1 subjects with baseline 1.5T MR images and that did not have 1.5T MR images available at 12 or 24 month follow-up, were used for training a multilevel variable selection scheme based on sparsity. The remaining 1.5T and 3T ADNI-1, -GO and -2 baseline images (as of November 2013) were used to evaluate and tune the proposed framework. In total 1701 images were used, from which 292 were used for multilevel variable selection and 1409 for evaluation and tuning.

Multilevel relevant variable selection The goal of variable selection is to reduce the amount of input variables to those that are relevant for a specific task. In this work a sparse regression was used in a similar fashion as the method described in [11] to select

AD classification using volumetry, cortical thickness and manifold learning

5

relevant variables. Here, independent variables are the MR image intensities, while the mini mental state examination (MMSE) score acts as dependent variable [11]. Furthermore, as disease specific imaging characteristics might manifest at different alignment levels, each of the *N* subjects has associated *R* images that have been created by aligning the original scan to the MNI152 template at *R* different levels using [15]. A matrix **X** is built, where each column represents a location in MNI space at a certain alignment level. Each row in matrix **X** represents a subject $n \in N$ via concatenating its vectorised MR images at multiple levels $r \in R$. The algorithm then selects a subset of *D* variables from **X**, that correspond to column indices of **X**. This yields a 4D mask, where the first three dimensions are coordinates in MNI space and the fourth is the alignment level of the image to the template.

Local binary patterns MR images acquired using different acquisition protocols have different intensity appearance and thus cannot be easily combined into a single framework. In this work we extract local binary patterns (LBP) [21] around the 26-connected neighbourhood of each selected voxel and encode them as binary vectors. This transforms MR intensity features to an augmented binary space where the images lie in the same space and thus can be combined, assuming that the original acquisitions protocols are reasonably similar (e.g. both are T1-weighted).

Dimensionality reduction The aim of this work is to produce a three-class classifier. For this purpose, it was found in [11] that the learned 4D mask is still relatively high-dimensional and that reducing the dimensionality generally improves classification results. In this work principal component analysis (PCA) [22] was used to reduce the dimensionality of the data.

3.4 Patch-based disease grading features (GRAD)

Training data ADNI baseline scans of 629 subjects (233 CN, 231 MCI, 165 AD) acquired at 3T and 30 training images from the challenge were utilised as training dataset. For each image, 150 grading features were calculated for classification. Then, an optimised classifier was trained on the training dataset and used to predict the class labels of the testing images from the challenge.

Method description After preprocessing, non-rigid registration, based on B-spline free-form deformation [15] with a final control point spacing of 5mm, was performed to align all images to the MNI152 template space. The intensity was normalised between each image and the template using the approach proposed in [23]. Then, the sparse regression method proposed in [24] was used to calculate a probability map for selecting patches. Finally, 150 patches with the highest probabilities are extracted from each image for calculating grading features.

A patch-based approach was proposed in [25] to calculate grading features for classification of AD. Our method is an extension of this method by introducing sparse representation techniques. To calculate a grading value for each target patch p_t , corresponding patches in the training images are extracted to form a training patch library

6 C. Ledig, R. Guerrero, T. Tong, K. Gray, et al.

 P_L . The patch library P_L typically contains thousands of training patches. The relationship between the target patch p_t and the patches in the training library P_L is modelled by a weighting function in [25]. In our work, the Elastic Net sparse regression model [12] is used to model the relationship between the target patch and the training patches. Since the pathological status of the training patches are known, we can propagate the pathological status to the target patch by using the patch relationship. The grading value of the target patch p_t can then be estimated as:

$$\begin{cases} \hat{a}_{t} = \min_{a_{t}} \frac{1}{2} \| p_{t} - P_{L} a_{t} \|_{2}^{2} + \lambda_{1} \| a_{t} \|_{1} + \frac{\lambda_{2}}{2} \| a_{t} \|_{2}^{2} \\ g_{t} = \frac{\sum_{j=1}^{N} \hat{a}_{t}(j) s_{j}}{\sum_{j=1}^{N} \hat{a}_{t}(j)} \end{cases}$$
(1)

where \hat{a}_t are the coding coefficients for the target patch p_t . Most of the coefficients in \hat{a}_t are zero due to the sparsity constraint. If the coefficient in \hat{a}_t is not zero, it indicates that the corresponding training patch has been selected to propagate their pathological information to the target patch. N is the number of the training patches in the patch library P_L . The CN and AD training groups were selected from the training dataset to propagate pathological status to the target image as suggested in [25]. The pathological status of the training patch $P_L(j)$ is denoted as s_j . If the training patch is extracted from CN subjects, s_j is set to 1. $s_j = -1$ is used for patches extracted from AD subjects. Finally, a grading value g_t can be calculated for each target patch p_t . Since 150 patches are extracted from each image, 150 grading features are calculated for classification.

4 Classification based on Random Forests

For all classification experiments, we applied the random forests algorithm [26] implemented in scitkit-learn ([27]; http://scikit-learn.org/). This method is able to predict both binary labels and class probabilities, and can be directly applied for three-class classification. As in the originally proposed algorithm, no maximum tree depth was specified, and a bootstrap sample of the training data was passed to each tree. Tree nodes were split based on an entropy criterion, and 100 trees were included in each forest. The number of features randomly sampled at each tree node was set to \sqrt{n} for all experiments, following the recommendation of Liaw and Wiener [28].

5 Results

5.1 Training data

Features generated by different methods seek to model differences in subgroups in a distinct way and thus may hold complementary information between each other. In order to asses this, a subset of subjects that overlap for all of the proposed methods was used for training. Features generated by those methods were combined into a single classification framework, where if complementary information exists, overall classification results would be expected to improve. In total, 734 subjects from the ADNI1-2 datasets where included in this analysis. See Table 1 for the demographics of the data.

AD classification using volumetry, cortical thickness and manifold learning

 0.59 ± 0.04

7

Table 1: Subject groups mean age, sample size, MMSE scores, gender, CDR scores and Magnetic strength from ADNI1-2.

	N	Age	MMSE	Men (#)	CDR	1.5T (3T)
AD 1	170	74.77±7.62	23.12 ± 2.06	46% (78)	0.79 ± 0.27	110 (80)
MCI 2	288	73.79±7.47	27.28 ± 1.79	55% (158)	0.50 ± 0.00	185 (103)
CN 2	276	74.75±5.82	29.07 ± 1.16	47% (131)	0.00 ± 0.00	156 (120)

We performed classification experiments in two variations. Firstly, we performed a 10fold cross validation approach within this subset of the ADNI database. Secondly, we trained the random forest classifier using this ADNI subset, and classified the provided CADDementia training set. The second experiment is the same setup that was used for calculating the submitted results based on the CADDementia testing data.

For each classification task we employed a random forest classifier that was trained on five subsets of the available features. We used features provided by each of the four individual methods (VOL, CORT, MBL, GRAD) as presented in 3.1-3.4. We furthermore combined all available features (ALL={VOL, CORT, MBL, GRAD}) provided by the presented methods to investigate whether they provide complementary information. We did not use available meta information such as age, field strength and gender.

The mean accuracies obtained in the ADNI cross-validation experiment are provided in Table 2. The mean accuracies obtained by 10 classification runs for the CADDementia training data are summarised in Table 3.

When using ALL features, we further investigated which features were most informative by extracting feature importances from the random forest classifier trained for the three-class classification task. We found that the first manifold coordinate of the MBL features was most important. Otherwise the top 50 features were, except left and right Amygdala volume (#18, #24), constituted by exclusively grading features provided by GRAD.

cross validation.							
Туре	# Feat.	AD vs. CN	AD vs. MCI	MCI vs. HC	AD vs. MCI vs. HC		
VOL	135	0.83±0.05	$0.68 {\pm} 0.04$	$0.67 {\pm} 0.05$	$0.54{\pm}0.04$		
CORT	591	$0.80{\pm}0.05$	$0.65 {\pm} 0.06$	$0.63 {\pm} 0.04$	$0.51 {\pm} 0.05$		
MBL	20	$0.89 {\pm} 0.05$	$0.67 {\pm} 0.07$	$0.70 {\pm} 0.05$	$0.58 {\pm} 0.03$		
GRAD	150	$0.86 {\pm} 0.04$	$0.67 {\pm} 0.04$	$0.69 {\pm} 0.04$	$0.56 {\pm} 0.04$		

 0.72 ± 0.05

 0.68 ± 0.04

ALL

896

 0.87 ± 0.03

Table 2: Overview of the classification results for the 10-fold cross validation on the subset of the ADNI1-2 cohort. Mean classification accuracy (\pm SD) based on 10-fold cross validation.

8 C. Ledig, R. Guerrero, T. Tong, K. Gray, et al.

Table 3: Overview of the classification results obtained on CADDementia training data. Mean classification accuracy (\pm SD) based on 10 classification runs.

Туре	# Feat.	AD vs. CN	AD vs. MCI	MCI vs. HC	AD vs. MCI vs. HC
VOL	135	$0.86 {\pm} 0.03$	$0.73 {\pm} 0.05$	$0.68{\pm}0.05$	$0.56 {\pm} 0.08$
CORT	591	$0.91 {\pm} 0.05$	$0.67 {\pm} 0.09$	$0.65 {\pm} 0.05$	$0.58 {\pm} 0.07$
MBL	20	$0.94{\pm}0.02$	$0.62{\pm}0.04$	$0.75 {\pm} 0.04$	$0.66 {\pm} 0.01$
GRAD	150	$0.88 {\pm} 0.03$	$0.75 {\pm} 0.06$	$0.76 {\pm} 0.03$	$0.67 {\pm} 0.05$
ALL	896	$0.92{\pm}0.02$	$0.78{\pm}0.05$	$0.75 {\pm} 0.04$	$0.68{\pm}0.05$

5.2 Computation times

The approximate computation times per subject are summarised in Table 4. None of the presented methods requires manual interaction.

Table 4: Overview of the approximate computation times per subject.

Task	Runtime	Implementation	Automatic
N4 bias correction	< 30 minutes	single core	yes
pincram brain extraction	< 1 hour	parallel	yes*
registration of the 30 atlases (VOL)	< 2 hours	parallel	yes
atlas fusion (VOL)	< 20 minutes	single core	yes
cortical thickness (CORT)	< 15 minutes	single core	yes
variable selection, 292 images (MBL)	< 2.5 hours	single core	yes
local binary patterns (MBL)	< 1 second	single core	yes
dimensionality reduction, ~1800 subjects (MBL)	< 10 seconds	parallel	yes
Grading feature extraction (GRAD)	< 5 minutes	single core	yes
classification	< 1 second	single core	yes

*(manual quality control)

6 Discussion

We have presented four independent approaches for extracting both structural and intensity based biomarkers from MR images with the goal of Alzheimer's disease state classification. Cross-validation experiments on the ADNI database suggest that the volumetric features (VOL) and in particular the intensity (MBL) and grading features (GRAD) are competitive to published state-of-the-art classification results. We have found that combining these features by training a single random forest for all features jointly seems to have no great impact on the classification performance. This finding suggests that the presented biomarkers provide little complementary information, which was unexpected but indicates the difficulty of an accurate three-way classification on AD classification using volumetry, cortical thickness and manifold learning

9

the ADNI database. Classification results on the training data provided through the CADDementia challenge were substantially higher than those obtained on the ADNI database. While this finding is, due to the small sample size of the training dataset, not definitive it is highly interesting as it suggests that the subjects in the challenge data show a clearer group separation than the ADNI cohort.

A Author's contributions

CL preprocessed and brain extracted all available scans. CL visually inspected all brain masks and calculated the multi-class whole brain segmentations. AM and CL calculated the cortical morphology features based on these segmentations. RG performed the multi-level variable selection using intensity features. TT calculated the gradient based grading features. KG combined all features and performed the final classification experiments. ASR supported the work sharing experience on this type of challenge and image registration in particular. RAH developed and assisted in the application of pincram. DR continuously supervised the project and enabled this work by providing the necessary resources. The individuals drafted the corresponding paragraphs of the manuscript.

References

- "Computer-Aided Diagnosis of Dementia based on structural MRI data (CADDementia)." http://caddementia.grand-challenge.org/.
- R. Wolz, P. Aljabar, J. V. Hajnal, A. Hammers, D. Rueckert, and Alzheimer's Disease Neuroimaging Initiative, "LEAP: learning embeddings for atlas propagation," *NeuroImage*, vol. 49, pp. 1316–1325, Jan. 2010.
- R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, O. Colliot, and Alzheimer's Disease Neuroimaging Initiative, "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," *NeuroImage*, vol. 56, pp. 766–781, May 2011.
- S. Gerber, T. Tasdizen, P. Thomas Fletcher, S. Joshi, R. Whitaker, and Alzheimers Disease Neuroimaging Initiative (ADNI), "Manifold modeling for brain population analysis.," *Med Image Anal*, vol. 14, pp. 643–653, Oct. 2010.
- K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, D. Rueckert, and Alzheimer's Disease Neuroimaging Initiative, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *NeuroImage*, vol. 65, pp. 167–175, Jan. 2013.
- S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett, "The Alzheimer's Disease Neuroimaging Initiative," *Neuroimaging Clinics of North America*, vol. 15, no. 4, pp. 869–877, 2005. Alzheimer's disease: 100 years of progress.
- N. Tustison, B. Avants, P. Cook, Y. Zheng, A. Egan, P. Yushkevich, and J. Gee, "N4ITK: Improved N3 bias correction," *IEEE Transactions on Medical Imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.
- R. A. Heckemann, C. Ledig, P. Aljabar, K. R. Gray, D. Rueckert, J. V. Hajnal, and A. Hammers, "Label propagation using group agreement DISPATCH," pp. 75–78, MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling, 2012.
- C. Ledig, R. Wolz, P. Aljabar, J. Lötjönen, R. A. Heckemann, A. Hammers, and D. Rueckert, "Multi-class brain segmentation using atlas propagation and EM-based refinement," *Proceedings of ISBI 2012*, pp. 896–899, 2012.

- 10 C. Ledig, R. Guerrero, T. Tong, K. Gray, et al.
- S. E. Jones, B. R. Buchbinder, and I. Aharon, "Three-dimensional mapping of cortical thickness using laplace's equation," *Human Brain Mapping*, vol. 11, no. 1, pp. 12–32, 2000.
- R. Guerrero, R. Wolz, A. W. Rao, and D. Rueckert, "Manifold population modeling as a neuro-imaging biomarker: Application to ADNI and ADNI-GO.," *NeuroImage*, vol. 94C, pp. 275–286, 2014.
- T. Tong, R. Wolz, P. Coupé, J. V. Hajnal, and D. Rueckert, "Segmentation of mr images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling," *NeuroImage*, vol. 76, pp. 11–23, 2013.
- B. Landman and S. K. Warfield, "MICCAI 2012 Workshop on Multi-Atlas Labeling. https://masi.vuse.vanderbilt.edu/ workshop2012/images/c/c8/miccai_2012_workshop_v2.pdf," 2012.
- R. A. Heckemann, S. Keihaninejad, P. Aljabar, D. Rueckert, J. V. Hajnal, and A. Hammers, "Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation," *NeuroImage*, vol. 51, no. 1, pp. 221–227, 2010.
- D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, and D. J. Leach, M. O.and Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999.
- M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin, "Fast free-form deformation using graphics processing units," *Computer Methods and Programs in Biomedicine*, vol. 98, no. 3, pp. 278–284, 2010.
- X. Han, D. L. Pham, D. Tosun, M. E. Rettmann, C. Xu, and J. L. Prince, "CRUISE: cortical reconstruction using implicit surface evolution," *NeuroImage*, vol. 23, pp. 997–1012, Nov. 2004.
- W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," in *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '87, (New York, NY, USA), pp. 163–169, ACM, 1987.
- 19. L. R. Herrmann, "Laplacian-isoparametric grid generation scheme," *Journal of the Engineering Mechanics Division*, vol. 102, pp. 749–907, Oct. 1976.
- C. E. Rodriguez-Carranza, P. Mukherjee, D. Vigneron, J. Barkovich, and C. Studholme, "A framework for in vivo quantification of regional brain folding in premature neonates," *NeuroImage*, vol. 41, pp. 462–478, June 2008.
- T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- H. Hotelling, "Analysis of a complex of statistical variables into principal components," Journal of Educational Psychology, vol. 24, pp. 417–441, 1933.
- 23. L. G. Nyu and J. K. Udupa, "On standardizing the MR image intensity scale," *Magnetic Resonance in Medicine*, vol. 42, no. 6, p. 1072, 1999.
- E. Janousova, M. Vounou, R. Wolz, K. Gray, D. Rueckert, and G. Montana, "Biomarker discovery for sparse classification of brain images in Alzheimer's disease," *Annals of the British Machine Vision Association (BMVA)*, vol. 2012, no. 2, pp. 1–11, 2012.
- P. Coupé, S. F. Eskildsen, J. V. Manjón, V. S. Fonov, J. C. Pruessner, M. Allard, and D. L. Collins, "Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease," *NeuroImage: Clinical*, vol. 1, no. 1, pp. 141–152, 2012.
- 26. L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
- 27. F. Pedregosa, G. Varoquaux, A. Gramford, *et al.*, "Scikit-learn: Machine Learning in Python," *The Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- 28. A. Liaw and M. Wiener, "Classification and regression by randomForest.," *R news*, vol. 2, pp. 18–22, 2002.

MRI based dementia classification using semi-supervised learning and domain adaptation

Elaheh Moradi¹, Christian Gaser², Heikki Huttunen¹, Jussi Tohka¹ and Alzheimer's Disease Neuroimaging Initiative^{**}

 Department of Signal Processing , Tampere University of Technology, P.O. Box 553 FIN-33101 Tampere, Finland jussi.tohka@tut.fi
Departments of Neurology and Psychiatry, Jena University Hospital

Jena, Germany

Abstract. We propose a method for dementia classification based brain magnetic resonance images (MRIs). The method learns to recognize patients with Alzheimer's disease or Mild Cognitive Impairment from healthy controls. The features used are extracted with sparse logistic regression from a large pool of voxel-wise gray matter densities computed based on MRIs registered to stereotactic space. The classifier uses a Low Density Separation algorithm, which can take advantage of both labeled and unlabeled samples. The differences between the training and test sets are compensated based on an algorithm for unsupervised domain adaptation. The method is fully automatic. The proposed method participated in the 2014 CADDementia competition, with an estimated accuracy of 0.767 for the public test data. The training data was extracted from ADNI database.

Key words: Semi-supervised learning, Alzheimer's disease, mild cognitive impairment, domain adaptation, low density separation

1 Introduction

Alzheimers disease (AD) is the most common form of dementia. More than 30 million people worldwide suffer from AD and, due to the increasing life expectancy, this number is expected to triple by 2050 [1]. Therefore, it is extremely important to identify subjects in a risk of getting the disease.

In this paper, we propose an approach to automatically categorize subjects into three classes: Subjects with Alzheimer's disease (AD), subjects with mild

^{**} Data used in preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc. edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

cognitive impairment (MCI) and cognitively normal subjects (NC). Our method relies on voxel-based morphometry (VBM) style preprocessing [2]. The MRI features used for the classification are selected from a larger pool of features using sparse logistic regression [3, 4]. Then, relying on these features, we construct a hierarchical classification framework utilizing binary classifiers as component classifiers. Each binary classifier is trained in a semi-supervised manner, meaning that they can utilize unlabeled data in addition to labeled data. The labeled data is obtained from the ADNI database and CADDementia data is used as unlabeled data. The semi-supervised learning is performed by the Low Density Separation (LDS) algorithm [5], and we have demonstrated its efficiency for classifying subjects with stable and progressive MCI [6, 7]. Here, we extend the methods [4, 7] to the three-category classification problem (AD vs. MCI vs. NC). Moreover, because our training and test sets have different characteristics we utilize an unsupervised domain adaptation method to normalize the samples [8]. The method is fully automatic.

The rest of this paper is organized as follows. Section 2 describes the image data used for the training and validating the classifier. Also the image preprocessing and the used features are described in Section 2. Section 3 introduces the classification scheme. Section 4 presents experimental validation results with CADDementia data and Section 5 concludes the paper.

2 Material

2.1 Training data: ADNI

Data used in this work to train the classifier is obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database http://adni.loni.usc.edu/. The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and nonprofit organizations, as a 60 million US dollar, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in

the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

Data used include baseline MRI of 835 subjects (T1-weighted MP-RAGE sequence at 1.5 Tesla, typically 256 x 256 x 170 voxels with the voxel size of approximately 1 mm x 1 mm x 1.2 mm). 835 subjects are grouped as

- 1. AD (Alzheimers disease), if diagnosis was Alzheimers disease at baseline (n = 200);
- 2. NC (Normal Cognitive), if diagnosis was normal at baseline (n = 231);
- sMCI (stable MCI) if diagnosis was MCI at all available time points, but at least for 36 months (n = 100);
- 4. *pMCI (progressive MCI)*, if diagnosis was MCI at baseline but conversion to AD was reported after baseline within 1, 2 or 3 years, and without reversion to MCI or NC at any available follow-up (n = 164);
- 5. *uMCI (unknown MCI)*, if diagnosis was MCI at baseline but they are not diagnosed at the end of the project. These subjects' data were not used for the classifier training.

We used different training sets based on this labeling to build the component classifiers of our hierarchical scheme. We explain the details in Section 3.4.

2.2 Test data:CADDementia

The method was validated with 30 labeled images from CADDementia data described in the challenge homepage ¹. These images were not used for training nor parameter tuning. The essential difference between this data and the training data is that the CADDementia images were acquired with 3 Tesla scanners.

2.3 Image preprocessing and features

All the images (train and test) were preprocessed in a fully automatic manner by a pipeline similar to that described in [2]. Preprocessing of the T1-weighted images was performed using the SPM8 package² and the VBM8 toolbox³, both running under MATLAB. All T1-weighted images were corrected for bias-field inhomogeneities, then spatially normalized and segmented into grey matter (GM), white matter, and cerebrospinal fluid (CSF) within the same generative model [9]. The segmentation procedure was further extended by accounting for partial volume effects [10], by applying adaptive maximum a posteriori estimations [11], and by using an hidden Markov random field model [12] as described previously [13]. Only the GM images were used. Note that these images represent GM tissue fractions in each voxel. Following the pipeline proposed by [14], the GM images

¹ http://caddementia.grand-challenge.org/home/

² http://www.fil.ion.ucl.ac.uk/spm/

³ http://dbm.neuro.uni-jena.de/

were processed with affine registration and smoothed with 8-mm full-width-athalf-maximum smoothing kernels. After smoothing, images were resampled to 4 mm spatial resolution. See Figure 1 for an example slice of the original image and preprocessed image.

Masking of the GM tissue fraction images results in aligned GM tissue fractions from 29852 voxels. As discussed in [4, 7], the number of voxels significantly exceeds the number of the available training data. Although our classifier is relatively tolerant to high-dimensional data, it is still unable to process this high number of features. Therefore, we initially reduce the number of features by using a regularized logistic regression classifier, that has an inherent feature selection property, on data from AD and NC classes (ADNI) [4, 7, 3]. The classifier produces a set of good candidate subsets with different cardinalities, and the most appropriate subset is selected by cross-validation. As a result, 309 voxels were selected. Finally, age related effects were removed from the data by using linear regression [15, 6].



Fig. 1. An example slice of the original image (a) and preprocessed image (b)

3 Classifier

3.1 Overview

A simplified flowchart of the training of the classifier is shown in Figure 2. Details of this procedure are explained in Section 3.4. Important algorithmic components, low density separation classifier and domain adaptation are briefly described in Sections 3.2. and 3.3, respectively.

Our classification scheme is hierarchical. In the first step, we aim to separate AD and NC subjects to different classes without caring to which class the MCI subjects are classified. We term these disease classes (AD + MCI and NC + MCI + MCI

MCI) as the first level classes. Before the classification, we perform the domain adaptation by using only AD and NC subjects in both datasets. In second step, the first level classes are further divided to AD and MCI classes (AD + MCI) and MCI and NC classes (NC + MCI) giving us the desired three class classification. Note that the label of the subjects classified as MCI is independent of whether the first level class was AD + MCI or NC + MCI. The rationale of this hierarchical scheme is that the MCI is a transitional stage between the AD and normal aging, however, not all the MCI subjects convert to AD and this makes the MCI class very heterogeneous.



Fig. 2. Classifier structure. DA stands for domain adaptation. LDS stands for low density separation.

3.2 Low density separation

We apply a semi-supervised learning (SSL) technique called Low Density Separation (LDS) to train the three required component classifies in our hierarchical scheme. We next explain the main ideas of the algorithm briefly; see [5] for further details. Semi-supervised learning (SSL) approaches are able to use unlabeled data in conjunction with labeled data in a learning procedure for improving the classification performance. LDS is a semi-supervised learning algorithm which relies on the assumption that there is low density region with little (if any) data, which is where the decision boundary should lie.

The algorithm consists of two stages. First, it constructs a graph distance derived kernel with the aim of increasing class separability and on the other hand to increase the clustering within the classes. Heuristically, the distance between the two nearest neighbors in feature space is incremented if they are far from each other and decremented if they are close to each other. The definition of the distance depends on the parameter ρ which we tune by cross-validation (see [5] for details).

The second step consists of training a transductive support vector machine with the graph-distance derived kernel to obtain the parameters for the dis-

criminant function $y = sign(\mathbf{w}^T \mathbf{x} + b)$, where **w** is the weight vector (in the transformed space), **x** is the image to be classified (in the transformed space), $y = \{-1, 1\}$ is the label of **x** and *b* is the bias of the classifier. Vector **w** and bias *b* are found by minimizing

$$\frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{n=1}^N L(y_n(\mathbf{w}^T \mathbf{x}_n - b)) + C^* \sum_{n=N+1}^{N+M} L(|\mathbf{w}^T \mathbf{x}_n - b|).$$

where $(\mathbf{x}_n, y_n), n = 1, ..., N$ are the labeled data, $\mathbf{x}_n, n = N + 1, ..., N + M$ are the unlabeled data, $L(\cdot)$ is the hinge loss function, and C and C^* are scalar parameters. The value of C is selected by cross-validation within the training data, and C^* is set as in [5].

3.3 Domain adaptation

It is unlikely that the data from ADNI and CADDementia would follow the same distributions conditioned on the disease labels. For example, the ADNI data we use has been acquired with 1.5 Tesla scanners while the CADDementia data has been acquired with 3 Tesla scanners. Techniques for addressing learning problems with mismatched distributions are often referred as *domain adaptation* or *transfer learning*. The idea of these algorithms is try to improve the similarity of the data from source (ADNI in our case) and target domains (CADDementia in our case). When there is no labeled data from the target domain to help learning classifiers, the problem setting is termed *unsupervised domain adaptation*. Here, we utilize an information theoretic approach for the unsupervised domain adaptation [8]. We did not compensate for possible differences between the different acquisition sites in the CADDementia data.

3.4 Detailed procedure

In the first level, the ADNI (AD and NC) subjects are used as source data and all CADDementia data are used as target data for domain adaptation. After domain adaptation, the AD and NC subjects from ADNI are used as training data for AD + MCI and NC + MCI classes, respectively. In order to design the first level classifier and the most important LDS parameters (C and ρ) are tuned using 10-fold cross-validation inside the training (ADNI) data. All CADDementia data are used as unlabeled data for this classifier and eventually divided into two groups. This is repeated 101 times to obtain the best possible parameter values and the final class of the subjects is decided based on the majority vote. Based on this procedure all the CADD data are divided into two groups, i.e., AD + MCI group and NC + MCI group.

In the second level, to design AD vs. MCI classifier, the AD and pMCI subjects of ADNI are used as the source data and the CADDementia subjects classified to the AD + MCI class during the first level are used as target data for the domain adaptation. After the domain adaptation the AD and pMCI

subjects of ADNI are used as labeled data for training in order to design the AD vs MCI classifier and the LDS parameters (C and ρ) are tuned using 10-fold cross-validation inside training data. The CADDementia subjects classified to the AD + MCI class during the first level are used as unlabeled data for the LDS classifier and subsequently classified to AD and MCI groups. Again this is repeated 101 times and the final label of the subject is decided based on the majority vote.

In the second level, to design NC vs. MCI classifier, the sMCI and pMCI subjects of ADNI are used as the source data and the CADDementia's NC + MCI class from the first level are used as the target data for the domain adaptation. After the domain adaptation the sMCI and pMCI subjects of ADNI are used for training in order to design the NC vs. MCI classifier and the LDS parameters (C and ρ) are tuned using cross-validation inside training data, again repeating the procedure for 101 times. The classifier divides CADDementia's NC + MCI class into two subclasses, i.e. NC and MCI subclasses. The decision to use sMCI subjects' data as the training data for NC class was made based on experimental grounds. In this phase, we balanced the numbers of pMCI and sMCI subjects in the ADNI data by resampling in both domain adaptation and classification phases because the number of sMCI subjects (100) is smaller than the number of pMCI (164) subjects in the ADNI data.

3.5 Computation time

The total running time was approximately 9 minutes per image with a Matlab based implementation. The computationally most heavy part was the image pre-processing implemented in VBM8 that required approximately 8 minutes per image. The domain adaptation required, on average, 29.73 seconds per 100 images (domain adaptation cannot be performed for a single image). The LDS classification and age removal required under one second per image. The domain adaptation were performed twice for each image (in the first and second level).

4 Results

The classification accuracy on 30 labeled examples in the CADDementia dataset was 0.767. Since we did not use the label information on the CADDementia data, we consider this to be unbiased estimate of the classification accuracy. The confusion matrix with this data is shown in Table 1. The confusion matrix shows that there were no NC subjects mislabeled as AD subjects or vice versa. However, there were mislabelings between MCI and AD and NC and MCI. This is consistent with the MCI being a transitional stage between normal aging and AD.
8 Semi-supervised AD classification

True class	Pre	dicted	l class
	NC	MCI	AD
NC	9	3	0
MCI	2	7	0
AD	0	2	7

Table 1. Confusion matrix

5 Discussion

We have proposed a method for dementia classification based brain MRIs. The fully automated method recognized patients with AD or MCI from healthy controls. The features for the classification were extracted from voxel-wise gray matter densities computed based on aligned MRIs [2]. The classification method was hierarchical, utilizing the fact that the MCI is a translational stage between the AD and normal aging. The main novelties of the classifier were the utilization of unlabeled data in additional to labeled data in order to improve the classification [5] and the use of unsupervised domain adaptation to try to compensate between the differences of the training and test data [8]. We have previously demonstrated the utility of unlabeled data for predicting MCI-to-AD conversion [4]. The method trained with ADNI data achieved an accuracy of 0.767 with the public part of the CADDementia test data.

Acknowledgments. Financially supported by Academy of Finland grants 130275 and 263785. 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: Alzheimers Association; Alzheimers Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research &Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuti-cals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; And Takeda Pharmaceutical Company. 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 Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

- Barnes, D.E., Yaffe, K.: The projected effect of risk factor reduction on Alzheimer's disease prevalence. The Lancet Neurology 10(9) (2011) 819–828
- Gaser, C., Franke, K., Klöppel, S., Koutsouleris, N., Sauer, H., et al.: BrainAGE in mild cognitive impaired patients: predicting the conversion to alzheimers disease. PloS ONE 8(6) (2013) e67346
- Friedman, J.H., Hastie, T., Tibshirani, R.: Regularization paths for generalized linear models via coordinate descent. J. Stat. Software 33(1) (2010) 1–22

Semi-supervised AD classification

- Moradi, E., Gaser, C., Tohka, J.: Semi-supervised learning in MCI-to-AD conversion prediction - when is unlabeled data useful? In: IEEE Pattern Recognition in Neuro Imaging, IEEE (2014) 121 – 124
- Chapelle, O., Zien, A.: Semi-supervised classification by low density separation. In: AISTATS. (2005) 57–64
- Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J.: Semi-supervised learning for early MRI-based MCI-to-AD conversion prediction. In: Organization for human brain mapping annual meeting. (2014)
- Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J.: Semi-supervised learning and random forests for MRI-based prediction of AD conversion in MCI subjects. Submitted (2014)
- Shi, Y., Sha, F.: Information-theoretical learning of discriminative clusters for unsupervised domain adaptation. ICML 2012 (2012)
- Ashburner, J., Friston, K.J.: Unified segmentation. Neuroimage 26(3) (2005) 839– 851
- Tohka, J., Zijdenbos, A., Evans, A.: Fast and robust parameter estimation for statistical partial volume models in brain MRI. Neuroimage 23(1) (Sep 2004) 84–97
- Rajapakse, J.C., Giedd, J.N., Rapoport, J.L.: Statistical approach to segmentation of single-channel cerebral mr images. Medical Imaging, IEEE Transactions on 16(2) (1997) 176–186
- Cuadra, M.B., Cammoun, L., Butz, T., Cuisenaire, O., Thiran, J.P.: Comparison and validation of tissue modelization and statistical classification methods in t1weighted MR brain images. IEEE Trans Med Imaging 24(12) (Dec 2005) 1548–1565
- Gaser, C.: Partial volume segmentation with adaptive maximum a posteriori (MAP) approach. NeuroImage 47 (2009) S121
- Franke, K., Ziegler, G., Klöppel, S., Gaser, C.: Estimating the age of healthy subjects from t1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. Neuroimage 50(3) (2010) 883–892
- Dukart, J., Schroeter, M.L., Mueller, K., et al.: Age correction in dementia– matching to a healthy brain. PloS one 6(7) (2011) e22193

Evaluation of morphometric descriptors of deep brain structures for the automatic classification of patients with Alzheimer's disease, mild cognitive impairment and elderly controls

Alexandre Routier^{1,2,3,4,5,6}, Pietro Gori^{5,1,2,3,4}, Ana B. Graciano Fouquier^{1,2,3,4,5}, Sophie Lecomte^{1,2,3,4,5,6}, Olivier Colliot^{1,2,3,4,5,6}, Stanley Durrleman^{5,1,2,3,4,6} and the Alzheimer's Disease Neuroimaging Initiative

¹ Sorbonne Universités, UPMC Univ Paris 06, UMR S 1127, ICM, F-75013, Paris ² Inserm, U1127, 75013, Paris, France

³ CNRS, UMR 7225, 75013, Paris, France

⁴ Institut du Cerveau et de la Moelle épinière, ICM, 75013, Paris, France ⁵ Inria Paris-Rocquencourt, 75013, Paris, France

⁶ Centre d'Acquisition et de Traitement des Images (CATI), Paris, France

1 Introduction

Our participation in the MICCAI 2014 CADDementia challenge aims at evaluating the performance of morphometric descriptors in multi-class classification tasks for the prediction of Alzheimer's disease and Mild Cognitive Impairment from structural Magnetic Resonance Images (MRIs).

We used the method for the construction of population-specific atlases that is described in [6, 5]. The method takes as input a set of segmented brain structures, which take the form of the union of labelled 3D surface meshes, called shape complexes. The method estimates an anatomical model, called template, which is representative of the shape complexes within a group of subjects. The variability in shape within the group is captured by 3D space deformations of the ambient space, which warps the anatomical model to the anatomical shape complex of each subject. The method estimates the anatomical model together with the deformation parameters.

The method requires to use the same set of homologous structures for all subjects. We choose a subset of 12 deep brain structures that were segmented from MRIs: caudate nucleus, putamen, pallidum, thalamus, hippocampus and amygdala of each hemisphere. We do not include the lateral ventricles because of a large variability in the segmentation of the horns of the ventricles, which could have masked other patterns of shape variability in the statistical analysis. We do not include the cortical surface because of the subject-specific gyrification.

Deformation parameters are seen as a multi-variate descriptor, which encodes the differences in shape between each subject's anatomical configuration and the anatomical model. This descriptor encodes different patterns such as the shift of the caudate nucleus due to the ventricular enlargement and the hippocampal atrophy, for instance. The residual shape, namely the difference between the deformed template and the subject's shape complex, is considered as noise. The combination of the two terms gives the likelihood of a given anatomical shape complex, which will be used in classification.

We use a sub-set of the Alzheimer's Disease Neuroimaging Initiative (ADNI) database to build the anatomical models. We build three anatomical models considering a group of Cognitively Normal (CN) subjects, subjects with Mild Cognitive Impairment (MCI) and patients with Alzheimer's disease (AD). Once the models are built, we test any new subject by registering each model to the shape complex of this subject and computing its likelihood. We then classify according to the maximum likelihood. We test our classifiers on another sub-set of the ADNI database and the CADDementia database.

The method is fully automatic. The atlas construction method uses the concept of varifolds [3] for mesh comparison and therefore does not require specific mesh pre-processing. The method is indeed robust with respect to changes in topology between meshes, small holes, spikes, irregular sampling and inconsistency in normal orientations. We do not perform quality control of the segmentations as small errors in the position of the boundaries are likely to be averaged out in this kind of shape analysis. The use of smooth 3D deformations also acts as low-pass filter which smoothes out irregularities in the boundaries of the structures. Few important failures in segmentations are likely to be considered as outliers in the statistical analysis. We use the implementation of the method in the software Deformetrica, which is freely available at www.deformetrica.org.

Building the anatomical models took 3 days, 15 hours on average (with a parallelization on 40 threads). Registering the anatomical models to test subjects took 10 hours and 20 minutes on average, with a standard deviation of about 1 hour and 30 minutes. The computations were made on a computer cluster wich is composed of two types of machines. The first one (with 32 computing nodes) is running on an Intel®Xeon®Processor X5650 (2x6 Cores, 2.66 GHz) and 12x4GB 1333MHz DDR3 Memory and the second one (with 2 computing nodes) is running on an Octo-processor Intel®Xeon®Processor X7550 (8x2x8 Cores, 2 GHz) and 128x2GB 1066MHz DDR3 Memory.

2 Material and Methods

2.1 Data sets

We use the baseline images from the ADNI database to build the statistical models. We choose the same set of 509 subjects as the ones selected in [4], decomposed into 162 cognitively normal controls (CN), 210 patients with Mild Cognitive Impairment (MCI) and 137 patients diagnosed with Alzheimer's disease (AD) at baseline. We split the data set into a training set of 50 CN, 50 MCI and 50 AD, the rest being our test set.

We perform the same pre-processing to all ADNI and CADDementia data. The atlas construction is performed only on the training sub-set of the ADNI data. Classification are performed on the test set of the ADNI data and the CADDementia data.

2.2 Data pre-processing

The data pre-processing consists of the following steps:

- We run FreeSurfer¹ on the T1 MRI data [7] with default parameters. The output is volumetric segmentation of various structures. At this stage, we exclude from the ADNI dataset, 2 subjects for which the FreeSurfer pipeline failed.
- We run a marching cube algorithm (as implemented in FreeSurfer) to reconstruct 3D triangular meshes from the volumetric segmentation of the 12 selected structures on the RAS coordinate system (Right, Anterior, Superior). We do not perform any other processing on the meshes, although they have holes, spikes and irregular meshing.
- We register all images to the image of a control young adult from the ADNI training data set (126_S_0405_S14635_I38828) using FSL software² [9]. We use rigid and scaling transformation with 7 degrees of freedom. The transformations are then applied to the meshes. The transformed meshes are the inputs given to the software Deformetrica.

Additionally, we build a naive prototype initialization for the anatomical models to give as input of Deformetrica. We build this prototype by mapping a sphere to each structure of the reference subject with very smooth parameters. The corresponding initial anatomical model is shown in Fig. 1-left.

2.3 Atlas construction on ADNI training data

We use the Deformetrica software to build the anatomical models and estimate the deformation parameters. The method minimizes the following criterion (see [5] for details):

$$E(\mathbf{X}_0, \mathbf{c}, \boldsymbol{\alpha}_0, \dots, \boldsymbol{\alpha}_N) = \sum_{i=1}^N \left\{ \sum_{k=1}^{12} \frac{1}{2\sigma_k^2} \left\| \phi^{\boldsymbol{\alpha}_i}(\mathbf{X}_{0,k}) - S_k^i \right\|_W^2 + \boldsymbol{\alpha}_i^T \mathbf{K}_V \boldsymbol{\alpha}_i \right\}$$
(1)

where

- $\mathbf{X}_0 = {\mathbf{X}_{0,k}}_{k=1,\dots,12}$ denotes the position of the vertices of the anatomical model with 12 components, one for each anatomical structure,
- **c** denotes a set of control points which are supposed to move to the most variable parts of the anatomical model,
- $\{\alpha_i\}_{i=1,...,N}$ denotes momentum vectors attached to the control points which parameterize the deformations of the anatomical model to each subject's anatomical configuration (among N the number of subjects),
- $-S_k^i$ denotes the mesh of the k-th structure of the i-th subject,

¹ http://surfer.nmr.mgh.harvard.edu

² http://fsl.fmrib.ox.ac.uk

- $\{\phi^{\alpha_i}\}_{i=1,...,N}$ denotes the smooth 3D deformation from the anatomical model to the *i*-th subject,
- $\ \|.\|_W$ denotes the varifold norm,
- $-\sigma_k^2$ denotes the variance of the noise of the *k*-th structure in the space of varifolds,
- \mathbf{K}_V is the deformation kernel matrix, so that $\boldsymbol{\alpha}_i^T \mathbf{K}_V \boldsymbol{\alpha}_i$ measures the squared norm of the initial velocity of the deformation

We choose the following parameters, using the rationale detailed in [5]:

- deformation kernel width: $\sigma_V = 10 \text{ mm},$
- varifold kernel width: $\sigma_W = 5 \text{ mm}$,
- variance of noise: $\sigma_k^2 = 16$ for all structures,
- template kernel width $0.5\sigma_V$,

other parameters being the ones by default in Deformetrica.

2.4 Classification of ADNI test data and CADDementia data

Any test image is transformed into a set of sub-cortical structures after the preprocessing steps explained in 2.2. We then register each atlas to this subject's shape complex. The registration is performed by minimizing the following criterion, which is essentially (1) for N = 1 and keeping fixed the atlas parameters: the template shape \mathbf{X}_0 and the control points \mathbf{c} :

$$E(\boldsymbol{\alpha}) = \sum_{k=1}^{12} \frac{1}{2\sigma_k^2} \left\| \phi^{\boldsymbol{\alpha}}(\mathbf{X}_{0,k}) - S_k \right\|_W^2 + \boldsymbol{\alpha}^T \mathbf{K}_V \boldsymbol{\alpha},$$
(2)

where the S_k 's denotes the test subject shapes.

The value of the criterion E at convergence is an approximation of the loglikelihood of the test data [1,2]. In order to take into account the covariance of the deformation parameters, we replace the matrix \mathbf{K}_V by the inverse of the regularized empirical covariance matrix of the momentum vectors $\boldsymbol{\alpha}_i$. This corrected value of the criterion is used in classification.

3 Results

3.1 Results on the ADNI data

In Fig. 2, the 3 estimated template shape complexes are shown. The template of the MCI class falls in-between the template of the CN and AD classes. These shapes show the shift of the caudate nucleus toward the lateral parts of the brain due to a larger and larger ventricular enlargement. We notice also a greater and greater atrophy of the hippocampus.

The confusion matrix of the classification performed on the test sub-set of the ADNI database is shown in Table 1. The accuracy, assuming the probability of 1/3 for each class, is 51% (i.e. $\frac{1}{3}\sum_{k=1}^{3} n_{k,k}/n_k$ where $n_k = \sum_{i=1}^{3} n_{i,k}$ is the

total number of samples of the class k). We notice that our classifier tends to empty the MCI class, and to classify MCI subjects as either CN or AD with equal probability. This may be explained by the fact that our descriptors of MCI subjects overlap the descriptors of CN and AD classes, as if there is a continuum between the three classes. In other words, our classifier does not detect shape patterns that are specific to MCI subjects. This conclusion is corroborated by the visualization of the 3 template shapes complexes in Fig. 2.

The ROC curves of pairwise classification are shown in Fig. 3. As expected, the AD versus CN classification has overall better performance than classification of AD or CN against MCI.



Fig. 1. Initial prototype given as input of Deformetrica (left) and an instance of estimated atlas given as output (right): template shapes are representative of the group and momenta arrows parameterize template-to-subject deformations.



Fig. 2. Superimposition of the 3 template shapes for the CN, MCI and AD classes in green, blue and red respectively. Anterior view (left) and posterior view (right)





Fig. 3. ROC curves of pairwise classification on the ADNI database.

Results on the CADDementia training data 3.2

We test our classifier on the 30 subjects of the training database of CADDementia. Table 2 shows the confusion matrix using the thresholds that maximize the accuracy of the classifier on the ADNI data set, for which the accuracy is 50%. These two thresholds determine the position of the boundaries between the three classes. The optimization of these two thresholds on the given 30 subjects of the CADDementia database yields the confusion matrix in Table 3 and an accuracy of 73%. Differences in optimum thresholds between the two databases may come from differences in patients, differences in age distribution, differences in clinical practice for the diagnosis of mild cognitive impairment and dementia. Optimizing the thresholds on only 30 subjects is also not ideal, as they might not generalize well to the rest of the data set. For these reasons, we decided to submit two predictions for each subject: one using the thresholds estimated from

the ADNI data set and the other one using the thresholds estimated from the CADDementia training data set.

		Tru	e cla	ISS			Tr	ue cla	SS
	A	AD N	1CI	CN			AD	MCI	CN
A A	D	4	3	0	ized	AD	9	5	2
M ilass	CI	3	0	1	thes	MCI	0	4	1
di C	N	2	6	11	Hype	CN	0	0	9

Table 2. Confusion matrix for the classi-
Table 3. Confusion matrix for the classification of the CADD
ementia training set, sification of the CADD
ementia training using the thresholds that are optimum for set after optimizing the thresholds for this
the ADNI data set.

4 Discussion and conclusion

This work evaluates the performance of the Deformetrica software in classification tasks. The software computes shape descriptors for anatomical shape complex of sub-cortical structures that are known to be markers of disease progression. The approach is essentially multi-variate and combine different shape patterns such as the effect of hippocampal atrophy and ventricular enlargement on the shape of the sub-cortical structures. Our results suggest that the method does not find shape features that are characteristic of MCI subjects. The method tends to position the anatomy of MCI subjects, as an intermediate stage of disease progression. This fact may come from the method itself, which does not capture characteristics of such non-demented subjects. It may also come from the heterogeneity of the MCI group.

Our goal was to use the software Deformetrica "out of the box" as a test case, whereas several improvements could be made such as the estimation of the covariance of deformation parameters and noise variance during the training phase along the lines of [1, 8]. We could have determined also the best thresholds using cross-validation on the ADNI database. Correction for age and sex could also have improved classification performance.

Acknowledgements

This work has been supported by the "Centre d'Acquisition et de Traitement des Images" (CATI) and the program "Investissements d'Avenir" ANR-10-IAIHU-06. Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database³. As such, the investigators

³ http://www.loni.ucla.edu/ADNI

within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. ADNI investigators include (complete listing available at http://www.loni.ucla.edu/ ADNI/Collaboration/ADNI Author ship list.pdf).

References

- Allassonnière, S., Amit, Y., Trouvé, A.: Towards a coherent statistical framework for dense deformable template estimation. Journal of the Royal Statistical Society Series B 69(1), 3–29 (2007)
- Allassonnière, S., Kuhn, E., Trouvé, A.: Construction of bayesian deformable models via a stochastic approximation algorithm: A convergence study. Bernoulli Journal 16(3), 641–678 (2010)
- Charon, N., Trouvé, A.: The varifold representation of non-oriented shapes for diffeomorphic registration. SIAM J. Imaging Sci. 6(4), 25472580 (2013), accepted for publication
- Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehricy, S., Habert, M., Chupin, M., Benali, H., Colliot, O., ADNI: Automatic classification of patients with alzheimer's disease from structural mri: a comparison of ten methods using the adni database. Neuroimage 56(2), 766–81 (2011)
- Durrleman, S., Prastawa, M., Charon, N., Korenberg, J., Joshi, S., Gerig, G., Trouvé, A.: Morphometry of anatomical shape complexes with dense deformations and sparse parameters. Neuroimage (2014), under minor revision
- 6. Durrleman, S., Prastawa, M., Korenberg, J.R., Joshi, S., Trouvé, A., Gerig, G.: Topology preserving atlas construction from shape data without correspondence using sparse parameters. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) Med Image Comput Comput Assist Interv. Med Image Comput Comput Assist Interv., vol. LNCS 7512, pp. 223–230. Springer (2012)
- Fischl, B., Salat, D., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355 (2002)
- Gori, P., Colliot, O., Worbe, Y., Marrakchi-Kacem, L., Lecomte, S., Poupon, C., Hartmann, A., Ayache, N., Durrleman, S.: Bayesian atlas estimation for the variability analysis of shape complexes. In: Proc. Med Image Comput Comput Assist Interv. vol. LNCS 8149, pp. 267–274 (2013)
- Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., Luca, M.D., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., Stefano, N.D., Brady, J.M., Matthews, P.M.: Advances in functional and structural {MR} image analysis and implementation as {FSL}. NeuroImage 23, Supplement 1(0), S208 – S219 (2004)

Advanced Feature Selection in Multinominal Dementia Classification from Structural MRI Data

Alessia Sarica¹ *, Giuseppe Di Fatta², Garry Mark Smith^{2,3}, Mario Cannataro¹, and James Douglas Saddy³ and for the Alzheimer's Disease Neuroimaging Initiative^{**}

¹ Department of Medical and Surgical Sciences, Magna Graecia University of Catanzaro, Italy, sarica, cannataro@unicz.it ² School of Systems Engineering, University of Reading, UK, g.difatta,g.m.smith@reading.ac.uk

³ Centre for Integrative Neuroscience and Neurodynamics (CINN), University of Reading, UK, j.d.saddy@reading.ac.uk

Abstract. Recent studies have shown that features extracted from brain MRIs can successfully discriminate Alzheimer's disease from Mild Cognitive Impairment. This study describes a method that sequentially applies advanced feature selection techniques for finding the best subset of features in terms of binary classification accuracy. The classifiers that provide the highest accuracy, are then used for solving a multi-class problem by the one-versus-one strategy. Although several approaches based on Regions of Interest (ROIs) feature extraction exist, the predictive power of these features has not yet been investigated by comparing filter and wrapper techniques. The findings of this work suggest that (i) the IntraCranial Volume (ICV) normalization can lead to overfitting and can worsen the predictive accuracy on data originated by different studies and (ii) the combined use of a correlation filter and a Random Forest-based filter improves the accuracy of classification.

Keywords: Dementia, MRI, Multinominal classification, Feature selection

^{*} The author is funded by the European Commission, European Social Fund and Regione Calabria. The author had full control of the design of the study, methods used, outcome parameters and results, analysis of data and production of the manuscript.

^{**} Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

1 Introduction

Alzheimer's disease (AD) is one of the most frequent neurodegenerative condition that causes loss of cognitive abilities and memory and, due to its morbidity, it represents a growing health problem. In prodromal stages of AD, patients are usually classified as having an amnestic Mild Cognitive Impairment (MCI), but not all patients affected by MCI will convert in AD. The criteria for distinguishing subjects affected by Alzheimer's disease from Mild Cognitive Impairment are usually based on clinical examination and neuropsychological assessment [1], but analyses of MRI neuroimages have been proposed for the early diagnosis of these two diseases too [2].

Approaches for extracting features from MRI neuroimages are usually based on different type of features from MRIs: the voxel-based approach considers the probability maps of different tissue, the vertex-based approach considers the vertex-level on the cortical surface and the ROI-based approach typically includes only the hippocampus volume and/or shape [3].

A major challenge is related to the discovery of the best subset of biomarkers that could improve the accuracy in discriminating AD from MCI. Furthermore, an important drawback is the lack of a general solution that is independent from the acquisition methods, the scanners, the pre-processing techniques of the neuroimages and the software used for this purpose. The CADDementia challenge¹ was launched with the aim of comparing computer-aided diagnosis methods for dementia based on MRI brain data. In the context of the challenge, this study proposes a fully automated method based on advanced feature selection techniques over the statistics generated by FreeSurfer [4], a tool for performing the segmentation and reconstruction of MRI neuroimages.

The use of FreeSurfer statistics for an analogue problem was evaluated in [3], where the authors adopted a given subset of predictors according to domain knowledge and based only on a linear regression criterion. On the contrary, this study applies and combines a number of data-driven methods for estimating the prediction power of features by sequentially applying and combining feature selection techniques on the three binary classification problems: controls versus ADs, controls versus MCIs and ADs versus MCIs. The binary models are evaluated in terms of estimated accuracy and the best three models are combined into a final *one-versus-one* multi-class classification.

Section 2 describes the adopted MRI datasets from the CADDementia challenge and the Alzheimer's disease Neuroimaging Initiative (ADNI) database and how they have been processed for extracting ROI features. A detailed description of the proposed method is provided in Section 3. Section 4 reports the results of feature selection, binary classification and multi-class classification. Finally, Section 5 presents some conclusions and directions of future work.

¹ http://caddementia.grand-challenge.org

Description	Name	Class	Nr. of subjects	Female%	$\begin{array}{c} Age\\(mean\pm std)\end{array}$
CADDementia training set	D1	HC AD MCI	12 9 9	25% 66.6% 44.4%	$\begin{array}{c} 62.33{\pm}6.26\\ 66.11{\pm}5.21\\ 68{\pm}8.54 \end{array}$
ADNI	D2	HC AD MCI	70 70 70	$\begin{array}{c} 48.6\% \\ 52.8\% \\ 45.7\% \end{array}$	73.6 ± 5.49 74.15 ± 8.07 72.6 ± 7.78

Table 1. Descriptive statistics of the CADDementia and ADNI data sets. Age values (years) are mean±standard deviation and include both female and male subjects.

2 MRI Datasets and Feature Generation

Given a set of MRI brain images, the classification task considered in this work, consists in building a multinominal classification model over the three classes AD, MCI and healthy control subjects (HC). A pre-processing step is used to generate a set of features from each MRI image. Feature extraction from MRIs is performed by means of FreeSurfer² [4]. In particular, the FreeSurfer recon-all script is applied with the option hippo-subfields to obtain cortical surfacebased measures, cortical and subcortical volume-based measures and volumes of the hippocampus subfields. The features used in this study consists of 45 volumes of subcortical structures, 34 mean thickness and 34 cortical volumes for each hemisphere, and 8 hippocampus volume subfields for each hemisphere, for a total of 197. Including the diagnosis, gender and age the total number of features per subject is 200. FreeSurfer tools store the generated features in a number of files organised per subject and per study. In order to manage the feature generation and selection task, we have adopted the Konstanz Information Miner (KNIME)³, a popular data analytics framework, and have developed K-Surfer⁴ [5], a novel KNIME plugin for a fully automated extraction of selected features from the numerous FreeSurfer output files. The characteristics of the data sets used in this work are reported in Table 1 and briefly discussed in the following sections.

2.1 The CADDementia dataset (D1)

The CADDementia challenge provides two sets of MRI scans, which include subjects from different studies. The competition training data is a small labelled data set. The competition testing data is a larger unlabelled data set. Hence,

² http://freesurfer.net

³ http://www.knime.org

⁴ https://sourceforge.net/projects/ksurfer/

for the aim of this work, only the competition training data are used to evaluate the classification models. We will refer to this dataset as D1, which consists of 30 MRI scans (T1w and acquired by 3T scanners) of controls, without any dementia syndrome, Alzheimer's disease (AD) affected patients and Mild Cognitive Impairment (MCI) affected patients. Additional information about the data characteristics, can be found on the official web site of the competition. Feature generation from the CADDementia MRIs was performed by using FreeSurfer 5.3.

2.2 The ADNI dataset (D2)

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is an international project that collects and validates neurological data such as MRI and PET images, genetics or cognitive tests. Besides imaging resources, ADNI provides cortical reconstruction and volumetric segmentation generated by FreeSurfer.

For the aim of this work, a subset of the ADNI data has been selected to match the characteristics of the CADDementia data. Only MRIs acquired by 3T scanners, weighted in T1 and processed by FreeSurfer, have been selected. The selected data include those subjects (a) diagnosed as HC, AD and late Mild Cognitive Impairment (late MCI) that (b) completely passed the quality test, and include (c) only *Non-Accelerated T1* scans related to the baseline visit. Data entries with missing values have been excluded.

Among all available subjects complying with the above constraints, the three class groups have been randomly sampled in order to obtain a balanced dataset. This dataset, which we refer to as D2, consists of 210 subjects (70 for each class).

3 Feature Selection and Classification Model Inference

This study employs a fully automatic feature selection approach and considers all the 197 volume and thickness features generated by FreeSurfer as described in the previous section: no manual feature selection is applied, e.g. by exploiting a priori clinical and/or domain knowledge.

High dimensionality of the data may affect the computational performance (processing time) and, worse, it may lead to a wrong estimation and identification of the relevant predictors. Feature selection can reduce dimensionality, thus mitigating performance issues and improving the classification accuracy.

Finding an optimal solution to the feature selection problem would require an exhaustive search over the feature subsets and is intractable. A suboptimal feature selection is typically solved with heuristic methods. Several techniques exist [6]: filters methods are applied independently from the chosen classification method, wrappers methods are strictly associated to the classification method, and embedded methods are performed within the classification process. In this work two filter methods and a wrapper method have been combined to improve the accuracy of the subsequent classification process. The overall data analysis workflow is shown in Figure 1 and is described in the next section.



Fig. 1. Diagram of the adopted workflow based on a combination of feature selection techniques.

3.1 The data analysis workflow

The adopted workflow is composed by five steps (Figure 1): (i) IntraCranial Volume normalization; (ii) Feature Selection with three techniques; (iii) Z-Score normalization; (iv) binary classification and (v) multi-class classification.

The core of the proposed method consists in sequentially applying a Correlation filter, a Random Forest (RF) filter and a Support Vector Machines (SVM) wrapper on the training dataset, to identify a subset of features that provides the highest binary classification accuracy. Four different combinations of these feature selection techniques are considered and tested.

ICV normalization • The aim of the IntraCranial Volume (ICV) normalization is to take into account differences between subjects in ROIs due to the size of the head and to the gender. It is performed by dividing each volumetric feature by the total intracranial volume of the subject. This normalization is widely used in the literature [7]. The proposed workflow has also been tested without it, as it could hide subtle differences in small areas.

Correlation-based filter • This filter discards any feature that is highly correlated to a feature already selected. A correlation constrain |r| < 0.7 is typically adopted, although more restrictive (0.4) and less restrictive (0.85) thresholds have also been used [8]. Since no conclusive knowledge about correlation filtering of dementia predictors exists, correlated features have been removed considering a very conservative threshold of 0.90.

Filter • A drawback associated with correlation-based filters is that they do not take into account the relations between features, which can actually improve the classification performance when considered together. One commonly-used data mining technique for filtering is Random Forests (RF), which is able to measure the importance of features w.r.t. the classification outcome, thanks to

the hierarchical decision tree structure that can model non-linear associations. A random forest-based filter is applied after removing highly correlated features and the features with an importance greater than 0.50 are selected.

Wrapper • As stated in [6], filters are tipically faster compared to wrappers, they can effectively reduce space dimensionality and they overcome overfitting. However, filters are independent from the chosen learning algorithm and take into account the prediction power of individual features, not of subsets of features. Wrapper methods solve this issue, searching the space for those subsets that provide the highest prediction accuracy. The Recursive Feature Elimination (RFE) method [9] performs backward feature elimination, with the aim of sequentially and iteratively removing the most irrelevant features. In this study, RFE has been applied with an SVM radial predictor as a filter.

Z-Normalization • A Z-score transformation is applied before performing classification to prevent that range differences in the features could have a negative effect.

Binary classification • Many classification methods have been applied to neuroscience data. Among those, it has been shown that Support Vector Machines (SVM) overcome the limitations of other techniques and have been successfully applied to discriminate a variety of neurological conditions [10]. SVMs are based on kernel functions, interpreted as a measure of *similarity* between two inputs. In this study a Radial Basis kernel is used. A 10 - fold cross-validation method is used for evaluating the accuracy of classification and for finding the optimal cost (C) hyperparameter of the SVM classification model. Multiple SVM models are generated from the training dataset and are applied to the test dataset to compute the prediction accuracy on the three pair-wise classification problems.

Multi-class classification • In general, distinguishing between two classes is an easier task and the most common strategy for multi-class classification is based on the aggregation of binary classifiers. The *one-versus-one* (OVO) method divides the problem into a number of binary problems equal to all possible combinations of pairs of classes [11]. The final output of the OVO method is derived from the probabilities calculated by each binary classifier as reported in a *score matrix*. Different aggregation techniques can be used, such as the *Voting Strategy* (VOTE) or the *Weighted Voting strategy* (WV). In this study, since no previous work has been done to determine the most appropriate aggregation strategy in the case of dementia, the VOTE method has been chosen for its simplicity and robustness [11]. The binary models with the highest accuracy are used for the multi-class classification. Where different models presented similar accuracy, the one with a smaller number of features is preferred.

4 Results

Three binary problems, HCvsAD, HCvsMCI and ADvsMCI, have been investigated by applying the proposed approach. Considering the four alternative feature selection combinations, eight SVM models are generated for each binary problem, four with ICV normalization and four without it. The best binary classification models are finally combined to provide a multinominal classification.

4.1 Binary classification

ADNI data (D2) are used for training the binary models and their accuracy is estimated with both ADNI data (10-fold cross validation) and CADDementia data (D1). The detailed results for the binary problems are reported in Table 3. The accuracy for the two cases with and without ICV normalization are provided. The highest accuracies on the ADNI data (D2) for each binary classification problems resulted to be as follows.

1) HCvsMCI: 77.1% with ICV normalization, RF filter + SVM wrapper; 2) HCvsAD: 90% with ICV normalization and Corr. filter; 3) ADvsMCI: 62.1% with no ICV normalization and RF filter.

The accuracy on the CADDementia data (D1) is computed with a simple hold-out method and the highest values were obtained without ICV normalization in all the three binary problems.

1) HCvsMCI: 66.7% with no ICV normalization and Corr. filter; 2) HCvsAD: 95.2% with no ICV normalization and RF filter; 3) ADvsMCI: 88.9% with no ICV normalization and RF filter.

4.2 Multi-class classification

The three best binary classification models have been chosen in terms of the highest accuracy on D1 and are indicated in bold in Table 3. These models were used for performing the multi-class classification (OVO). In particular, where equal accuracy has been obtained, the model with less features has been preferred. Where the number of features is equal too, it has been preferred the model with less steps applied. The multi-class accuracy is computed with and without ICV normalization. In the case of ICV normalization, the chosen models are: for HCvsMCI RF+wrapper (42.9%), for HCvsAD RF filter (42.9%), for ADvsMCI RF filter (50%). The chosen models with no ICV normalization are: HCvsMCI Corr. filter (66.7%), HCvsAD RF filter (95.2%), ADvsMCI RF filter (88.9%).

These models are used to compute the prediction probabilities on D1 and to generate a score matrix. The VOTE strategy applied to the score matrix of each test data entry provides the final multinominal classification. The best accuracy of 70% was obtained with *No ICV*, while ICV normalization seems to introduce a significant degradation in the classification process with an accuracy of only 30%.

The training dataset (D2) and the test data set (D1) originated from different studies and, in this work, have not been combined to train the models: the predictive models obtained from D2 have been tested on D1 to provide an indication of the level of generalisation that can be achieved when dealing with heterogeneous data sources.

4.3 Computation time

The average computation time for each stage of the proposed algorithm is reported in Table 2. The processing, filtering and analysis of the datasets was performed on a 2.4GHz Intel Core i7 with 8GB RAM running Mac OS X 10.7.5. Apart from the preprocessing time for the ROI features generation from MRIs, the most expensive phase of the process is the application of the SVM wrapper (119.790s). Less time is taken for the SVM wrapper applied after the RF filter (91.790s) because of the fewer number of features.

Table 2. Mean computation time for each step of the proposed approach.

Step	Time
ROI feature extraction	5 hours per subject
Correlation filter	0.836 s
Random Forest filter	7.744 s
SVM wrapper	112.790 s
RF filter + SVM wrapper	91.790 s
OVO classification	0.496 s

5 Conclusions

The present study was designed for searching the best subsets of ROI features extracted by FreeSurfer from brain MRIs, by applying advanced methods for feature selection. The main goal was to obtain the highest accuracies on three binary problems, HC vs MCI, HC vs AD and AD vs MCI so to maximize the accuracy of multi-class classification by OVO strategy. The proposed workflow sequentially applies a correlation filter, a Random Forest filter and a Support Vector Machines wrapper on the data with and without IntraCranial Volume normalization. The findings suggest that (i) ICV normalization can lead to overfitting and worse accuracy and (ii) even though the SVM wrapper is more complex and slower than the Random Forest filter, it does not provide better accuracy. However in order to provide conclusive results, a more extensive analysis is required.

One limitation of the proposed method is that the same correlation and variable importance threshold has been used for each binary problem, while, due to the subtle differences between AD and MCI, distinct values may help to improve the overall accuracy.

Thus, further work is required in order to determine the best thresholds for the correlation filter and the best variable importance value for the Random Forest filter for each binary sub-problems independently.

References

- Frank Jessen, Steffen Wolfsgruber, Birgitt Wiese, Horst Bickel, Edelgard Mösch, Hanna Kaduszkiewicz, Michael Pentzek, Steffi G Riedel-Heller, Tobias Luck, Angela Fuchs, et al. Ad dementia risk in late mci, in early mci, and in subjective memory impairment. Alzheimer's & Dementia, 10(1):76–83, 2014.
- Yue Cui, Wei Wen, Darren M. Lipnicki, Mirza Faisal Beg, Jesse S. Jin, Suhuai Luo, Wanlin Zhu, Nicole A. Kochan, Simone Reppermund, Lin Zhuang, Pradeep Reddy Raamana, Tao Liu, Julian N. Trollor, Lei Wang, Henry Brodaty, and Perminder S. Sachdev. Automated detection of amnestic mild cognitive impairment in community-dwelling elderly adults: A combined spatial atrophy and white matter alteration approach. *NeuroImage*, 59(2):1209 – 1217, 2012.
- 3. Rémi Cuingnet, Emilie Gerardin, Jérôme Tessieras, Guillaume Auzias, Stéphane Lehéricy, Marie-Odile Habert, Marie Chupin, Habib Benali, and Olivier Colliot. Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *NeuroImage*, 56(2):766 781, 2011. Multivariate Decoding and Brain Reading.
- 4. Anders M Dale, Bruce Fischl, and Martin I Sereno. Cortical surface-based analysis: I. segmentation and surface reconstruction. *Neuroimage*, 9(2):179–194, 1999.
- 5. Alessia Sarica, Giuseppe Di Fatta, and Mario Cannataro. K-Surfer: A KNIME extension for the management and analysis of human brain MRI FreeSurfer/FSL data. In Proceedings of the International Conference on Brain Informatics and Health, Special session on Neuroimaging Data Processing Strategies, Warsaw, Poland, 11–14 August 2014, in press, LNAI. Springer.
- Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. J. Mach. Learn. Res., 3:1157–1182, March 2003.
- Randy L Buckner, Denise Head, Jamie Parker, Anthony F Fotenos, Daniel Marcus, John C Morris, and Abraham Z Snyder. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *Neuroimage*, 23(2):724–738, 2004.
- 8. Carsten F. Dormann, Jane Elith, Sven Bacher, Carsten Buchmann, Gudrun Carl, Gabriel Carré, Jaime R. García Marquéz, Bernd Gruber, Bruno Lafourcade, Pedro J. Leitão, Tamara Münkemüller, Colin McClean, Patrick E. Osborne, Björn Reineking, Boris Schröder, Andrew K. Skidmore, Damaris Zurell, and Sven Lautenbach. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1):27–46, 2013.
- Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3):389–422, 2002.
- Mario Cannataro, Pietro Hiram Guzzi, and Alessia Sarica. Data mining and life sciences applications on the grid. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3(3):216–238, 2013.
- Mikel Galar, Alberto Fernández, Edurne Barrenechea, Humberto Bustince, and Francisco Herrera. An overview of ensemble methods for binary classifiers in multiclass problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition*, 44(8):1761 – 1776, 2011.

Best	
D1).	
(D2/)	
l D1	
) and	
2/D2	
02 (D	
on L	
lated	
estin	
Iracy	
accu	
2 and	
on D	
med	
erfor	
ting p	
train	
tion:	
ssifica	
y clas	bold
binar	ed in
f the	ldicat
ults of	are in
Rest	D1
le 3.	lts on
Tab	resul

$\begin{array}{c ccccccccccccccccccccccccccccccccccc$																									
$\frac{ICV}{Nr.\ D2/D2} \frac{No\ ICV}{D2/D1} \frac{No\ ICV}{Nr.\ D2/D2} \frac{No\ ICV}{D2/D1} \frac{Nc\ D2/D2}{Nr.\ D2/D2} \frac{Nc\ D2/D2}{D2/D1} \frac{Nc\ D2/D2}{Nr.\ D2/D2} \frac{Nc\ D2/D2}{D2} \frac{Nc\ D2/D2}{D2} \frac{Nc\ D2/D2}{Nc\ D2/D1} \frac{Nc\ D2/D2}{Nr.\ D2/D2} \frac{Nc\ D2/D2}{$				Corr. $ r <$. filta 20.90	er			Ra	ndom F	oresi	t filter				n MMS	rapp	юr		Ranc	dom Fo	rest filt	er +	^{n}MAS	vrappeı
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $			ICV			No IC	Ai		ICV			No IC	Λ,		ICV			No IC	2		ICV			No IC	Λ
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$		Nr.	D2/D2	D2/D1	Nr.	D2/D2	D2/D1	Nr.	D2/D2	D2/D1	Nr.	D2/D2	D2/D1	Nr.	D2/D2	D2/D1	Nr.	D2/D2	D2/D1	Nr.	D2/D2	D2/D1	Nr.	D2/D2	D2/D1
HCvsMCI 133 71.4% 42.9% 140 67.9% 66.7% 71 71.4% 42.9% 41 65.7% 52 HCvsAD 133 90% 42.9% 140 81.4% 95.2% 109 86.4% 42.9% 95 82.1% 95 ADvsMCI 133 58.6% 50% 139 58.6% 83.3% 47 60% 50% 44 62.1% 88		feat.	. 10xval	holdout	feat.	. 10xval	holdout	feat.	10xval	holdout	feat.	10xval	holdout	feat.	10xval	holdout	feat.	10xval	holdout	feat.	10xval 1	holdout	feat.	10xval	holdout
HCvsAD 133 90% 42.9% 140 81.4% 95.2% 109 86.4% 42.9% 95 82.1% 95 ADvsMCI 133 58.6% 50% 139 58.6% 83.3% 47 60% 50% 44 62.1% 88	DVsMC	[133	71.4%	42.9%	140	67.9%	66.7%	71	71.4%	42.9%	41	65.7%	52.4%	133	71.4%	42.9%	140	67.9%	66.7%	30	77.1%	42.9%	41	65.7%	52.4%
A DvsMCI 133 58.6% 50% 139 58.6% 83.3% 47 60% 50% 44 62.1% 88	CvsAD	133	%06	42.9%	140	81.4%	95.2%	109	86.4%	42.9%	95	82.1%	95.2%	133	%06	42.9%	140	81.4%	95.2%	109	86.4%	42.9%	95	82.1%	95.2%
	DvsMC	[133	58.6%	50%	139	58.6%	83.3%	47	80%	50%	44	62.1%	88.9%	133	58.6%	50%	139	58.6%	83.3%	47	80%	50%	44	62.1%	88.9%

Global Disease Index, a novel tool for MTL atrophy assessment.

Francesco Sensi¹, Luca Rei¹, Gianluca Gemme¹, Paolo Bosco², Nicola Amoroso³, Andrea Chincarini¹, and The Alzheimer's Disease Neuroimaging Initiative^{*}

 ¹ National Institute of Nuclear Physics (INFN), Branch of Genoa, Genoa, Italy
 ² LENITEM Laboratory of Epidemiology, Neuroimaging, and Telemedicine - IRCCS Centro S. Giovanni di Dio - FBF, Brescia, Italy

³ National Institute of Nuclear Physics (INFN), Branch of Bari, Bari, Italy

Abstract. Hippocampi and medial temporal lobe (MTL) structures are notoriously among the first anatomical districts to be troubled by Alzheimer's Disease (AD). Accurate atrophy quantification for temporal and cortical brain structures is considered a promising marker for prodromal AD, thus the urge upon finding suitable automatic tools to perform voxel-based-morphometry tasks such as anatomical structures segmentation, shapes outlining and features selection.

We propose an original neuroanatomical approach, called "Global Disease Index" (GDI), stemming from the methodology appeared in [1]. It is a profound reworking of that procedure, based on local analysis of 9 MRI regional volumes of interest (VOIs) containing relevant MTL structures. These VOIs are filtered by means of a Random Forest classifier in order to enhance peculiar image features found to be the most significant to discriminate between cognitively normal subjects and Alzheimer's Disease patients. These features are subsequently processed with a Random Forest and a Support Vector Machines classifiers, providing an assessment of MTL atrophy in the form of a classification index.

The procedure proved to be a robust and reliable tool, able to distinguish with fine accuracy CN, AD and MCI. On MICCAI's CADDementia Grand Challenge provided train data, GDI has demonstrated a detection power of 93.5%, 68.9%, 92.6% of AUC for CN, MCI and AD cohorts respectively, supporting the reliability of the overall algorithm.

Keywords: Alzheimer's Disease, MRI, MTL atrophy

Abbreviations: GDI: Global Disease Index, AD: Alzheimer's Disease; CN: Cognitively Normal; MCI: Mild Cognitive Impairment; MCI-conv: MCI converter; MCInonconv: MCI non converter; VOI: Volume of Interest; RF: Random Forest; SVM: Support Vector Machines; AUC: Area Under Curve;

* Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators is available at http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Authorship_list.pdf. 2

1 Introduction

The past 10 years have seen a growing consensus that the atrophy of medial temporal lobe structures can potentially be crucial to capture early AD onset and disease progression by means of T1-weighted MR images analysis. In particular atrophy of the hippocampus induced by AD may reasonably be a specific biomarker of pathology progression, known that volume loss in hippocampi is strictly connected to increasing impairment in cognitive performances of afflicted the subjects [2]. This fact justifies the need of defining reliable quantification methods to assess changes in MTL morphology. In the case of hippocampal volume, a commonly accepted and clinically validated automatic segmentation procedure would quickly replace traditional time-consuming and rater-dependent manual tracing protocols.

Automatic approaches are introduced, as the one proposed in [3], permitting accurate extraction of desired Volumes of Interest (VOIs) from MRI, on which voxel-based morphometry indicators can be computed.

We present in this work an analysis pipeline build to automatically assess the progression of brain atrophy brought by AD in medial temporal lobe structures.

The technique, named GDI, is based on major improvements of the feature selection and classification procedure seen in [1], to study intensity and textural characteristics of a set selected volumes surrounding MTL structures to discriminate AD and CTRL subjects.

The core of the methodology is the local analysis of 9 MRI regions of interest containing relevant MTL anatomical structures. These VOIs are filtered by means of a Random Forest (RF) classifier in order to select, in the target MRI, image features previously found to be significant to discriminate between CN subjects and AD patients. These voxels are are processed with classifiers (RF plus SVM), providing an index assessing overall MTL atrophy.

2 Materials & Methods

We propose an improvement of the procedure designed in [1]. The procedure is fully automated and needs an average time of 45 minutes to complete a single subject scan analysis, on an average single core computer with 2.27 GHz, 64 bit system. We noticed that GDI analysis speed depends more on computer clock-cycle than on its memory or file system characteristics. The most resource demanding steps in the pipeline are registrations and volumes extractions.

2.1 Subjects

Data used in the preparation of this work consist in a sample of 551 baseline MRI, of just as many subjects, downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) public database

(http://www.loni.ucla.edu/ADNI/Data). Images were acquired all with 1.5 T

scanners, whilst scanners' type can be different and used with different acquisition protocols.

Statistical information on this training cohort is summarized in Table 1.

Table 1. ADNI train sample.

Cohort	Size	Gender	Age
$_{\rm CN}$	190	95 M, 95 F	76.6 ± 5.5
AD	195	66 M, $79\ {\rm F}$	76.6 ± 7.8
MCI	166	71 M, 45 F	75.5 ± 7.4

AD cohort includes 50 MCI subjects who converted to AD after a follow-up of 2 years (MCI-conv), while MCI cohort includes only MCI subjects who retained the same clinical assessment after a follow-up of 2 years (MCI-nonconv). The ground for the decision to merge the MCI-conv into the AD cohort, was to keep the MCI cohort, with only MCI-nonconv, distinct from the AD [6,7].

The 30 MRI provided by CADDementia organizers are employed to fine tune the procedure free parameters. 13 of these scans come from the EMC: Erasmus MC center (Rotterdam, the Netherlands), 3 scans from the UP: University of Porto - Hospital de São João, and 14 from the VUMC: VU University Medical Center (Amsterdam, the Netherlands).

2.2 MRI analysis

The main steps of GDI workflow are image preprocessing (noise reduction and registration), multiple template-based anatomical structure registration, extracted volumes intensity normalization, features enhancement, and Random Forest plus Support Vector Machines features classification.

Each target image is processed with a pyramid noise-filtering algorithm [4] to promote image uniformity across sites and machines. The difference with respect to previous paper is that the 3 thresholds N_t necessary to extend to 3D the algorithm are no longer automatically calibrated for every image and every direction based on the Structural Similarity Index (SSI) curve, but are replaced with a single fixed value corresponding to the mean value of all N_t calculated on training data in the original procedure. The reason behind this choice is that dynamic threshold denoise was found to much invasive and image dependent. A fixed threshold sensibly reduces running time of denoising module.

De-noised scans are then registered and re-sampled onto the Montreal Neurological Institute (MNI) ICBM152 reference, with a $1mm^3$ isotropic grid [5]. The 3-fold registration process of the ancestor paper has been substituted with a faster, single registration process implemented with Insight Toolkit (ITK, www.itk.org/). Each incoming image is subjected to a rigid registration with

7 degree of freedom (similarity registration) and to an affine registration. This simplified and faster workflow has comparable performances to the original one.

Once preprocessed in this way, each MRI is sampled with 9 VOIs with different dimensions placed around pertinent biological entities in MTL and cortex (refer to Figure 1), to reduce analysis burden. This VOIs are chosen to include those temporal lobe structures that are known to be affected in early AD, such as the entorhinal, perirhinal cortex, hippocampus and parahippocampal gyri, irrespective of normal inter- and intra-individual variability. Two additional VOIs are chosen as control volumes, in regions known to be relatively spared in early AD.

This extraction operation, producing parallelepiped-shaped volumes containing the desired anatomical structures, is carried on, in order to preserve accurate anatomical correspondence, by means of a rigid registration using references; i.e. a registration of several predefined VOIs onto the subject MNI-normalized brain. There are at least 8-10 references for each contra-lateral target object.

These template VOIs are designed to capture the morphological differences among subjects showing varying degrees of neurodegeneration, ranging from healthy elderly to severe AD. Details on the generation of VOI references can be found in [3].

VOI extraction step has the advantage of providing a reliable method to circumscribe noticeable structures and nearby tissues with reasonably high accuracy and reproducibility among subjects and machineries.

The intensity normalization operation is now applied on registered boxes. Mean values of CSF/Gray Matter/White Matter within the target VOI are obtained with k-means cluster analysis [8].

New intensity values are obtained by non-linear matching of these 3 values to the 3 mean segmented cluster values found for a $n = 50 \times 120 \times 50$ voxel region extracted around the corpus callosum of the MNI template. This mapping is extended to intermediate intensities with a smooth piece-wise polynomial curve.

All 9 normalized VOIs from each MRI are now filtered to highlight a reduced set of relevant voxels.

We used 18 different filters (Gaussian mean, standard deviation, range, entropy and Mexican-hat filters calculated on different voxel neighbourhoods), therefore the feature set for each subject under analysis consists of the ensemble of all voxels of the filtered VOIs extracted from its MRI. These features are subsequently pruned by means of RF, keeping the 85% most significant ones in terms of training set CN vs AD distinction [1].

On the output restricted collection of MRI features a Random Forest and a Support Vector Machines classifiers are built.

GDI value is then calculated combining the outcome of the 2 classifiers: a weighted mean of the two values is computed, considering the GDI intervals in which every classifier is more reliable.





Fig. 1. (*Above Left*) VOI size and positioning displayed on the MNI reference image. VOI n. 1,2: red; VOI n. 3,4: green; VOI n. 5,6: yellow; VOI n. 7: cyan; VOI n. 8,9: magenta. (*Above Right*) Main gray matter structures captured in the VOIs; [a] potentially significant regions; [b] control regions. (*Below Left*) Intensity normalization VOI size and positioning displayed on the MNI reference image. Such a region serves as basis for the histogram matching procedure, following the segmentation into CSF/GM/WM. (*Below Right*) Example of hippocampal VOI registration on a test subject.

2.3 Classification

Each subject is given a membership probability to each group (CN, MCI and AD). Probability Distributions, depicted in Figure 2, are generated with the GDI values coming from the classification of the 30 CADDementia train images.

The GDI index of a new image is evaluated as member of all PDF curves, producing 3 probability values in CN, MCI and AD distributions.

These values are normalized to one and the final class is assigned to the subject with a winner-takes-all scheme (the class with the greatest probability is the winning one).



Fig. 2. Probability density estimates calculated from the 30 CADDementia training images. The restricted number of samples is the reason why curves show some bumps. The integral of each curve equals 1.



Fig. 3. Performances for the distinction of CN from AD (black curve), CN from MCI (blue curve) and MCI from AD (magenta curve) on CADDementia training sample (*Left*) and on ADNI sample data (*Right*).

3 Results

GDI index discriminates, in Leave-20-Out cross-validation, CN from AD, CN from MCI and MCI from with AUC values of 0.97 (sensitivity = 0.94 @ specificity = 0.90), 0.71 (sensitivity = 0.77 @ specificity = 0.78) and 0.85 (sensitivity 0.86 @ specificity = 0.64).

The classification results on CADDemetia train dataset stands at AUC values of 0.93 for CN vs AD discrimination (sensitivity = 0.92 @ specificity = 0.88), AUC = 0.78 for CN vs MCI (sensitivity = 0.92 @ specificity = 0.67) and 0.8 for MCI vs AD distinction (sensitivity = 0.89 @ specificity = 0.62).

This outcomes can also be seen in terms of One-versus-All classes detection. In this case CN subjects are identified with AUC of 0.935, MCI with AUC of 0.689 and AD with AUC of 0.926. The accuracy of the CN/MCI/AD classification on the CADDementia training data is 0.733. Performances are represented in Figures 3 and 4.

On the other hand blind classification of the test population delivers 146 subjects as cognitively normal, their GDI index centered on a mean value of 0.85 with a standard deviation of ± 0.06 , 125 as mild cognitive impairment (GDI in 0.51 ± 0.16) and 83 as Alzheimer's Disease (GDI in -0.31 ± 0.29).

Classification index strictly depends on the goodness of the registration outcome. At the end of the automatic process we visually checked the registered images finding that a very little percentage of them has been poorly registered. This fact had not prevented the algorithm to proceed, so that we have no missing data in our analysis. However we reckon that this fact produces an uncertainty on our results which can be quantified: 3% on ADNI and 6 subjects in 354 on the CADD test sample.

4 Discussion

GDI procedure needs an average cpu-time of 45 minutes to process a 1.5T MRI and provides an index that is an assessment of MTL atrophy progression. GDI index reliability strongly depends on image registration process, and we noticed that approximately 3% of processed image is not properly registered. Nevertheless this index has been used to detect CN, MCI and AD cohorts in CADDementia train dataset with fine accuracy (93.5%, 68.9%, 92.6% respectively).

Incidentally, we have developed a filter testing registration accuracy by means of simple correlation coefficient with the reference template, in a way that we've been able, during training phase with ADNI data, to reject little correlated registration outcomes or to re-register them with more suitable parameters.

For what regards classification results, the not so optimal performances on CN vs MCI distinction can be explained looking at the consistent overlap of their probability distributions. On the contrary, MCI and AD present a delicate area of overlap (around 0.2), while CN and AD populations are almost completely separated in terms of GDI index (Figure 2).



Fig. 4. (*Above*) ROC curves for the detection of CN, MCI and AD for all training data of CADDementia challenge together, and (*Below*) considered separately for each of the three provided database subsets EMC, UP, VUMC (see 2.1).

5 Conclusions

In the current study we have shown an automatic system providing an index working as an objective measure of hippocampal and temporal lobe atrophy and its performance on the training and test data available in CADDementia.

The algorithm characteristics - such a speed and required computational resources - make it suitable to work on grid environment or to be provided as remote web-service.

References

- Chincarini, A., Bosco, P., Calvini, P., Gemme, G., Esposito, M., Olivieri, C., Luca Rei, Squarcia, S., Rodriguez, G., Bellotti, R., et al.: Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. Neuroimage. 58(2), 469–480 (2011)
- Frisoni, G.B., Fox, N.C., Jack, C.R., Scheltens P. and Thompson, P.M.: The clinical use of structural MRI in Alzheimer's disease. Nature Reviews s Neurology. 6(2), 67–77 (2010)
- Calvini, P., Chincarini, A., Gemme, G., Penco, M.A., Squarcia, S., Nobili, F., Rodriguez, G., Bellotti, R., Catanzariti, E., Cerello, P., De Mitri, I., Fantacci, M.E.: Automatic analysis of medial temporal lobe atrophy from structural MRIs for the early assessment of Alzheimer disease. Medical Physics. 36 (2009).
- Castleman, K., Schulze, M., Wu, Q.: Simplified design of steerable pyramid filters. Proceedings of the 1998 IEEE International Symposium. 329–332 (1998)
- 5. Mazziotta, L.: A probabilistic atlas of the human brain: theory and rationale for its development. Neuroimage. 2, 89–101 (1995)

- Petersen, R.C., Doody, R., Kurz, A., Mohs, R.C., Morris, J.C., Rabins, P.V., Ritchie, K., Rossor, M., Thal, L., Winblad, B.: Current concepts in mild cognitive impairment. Archives of neurology. 58(12), 1985–92 (2001)
- 7. Petersen, R.C.: Mild cognitive impairment as a diagnostic entity. J Intern Med. 256(3), 183–94 (2004)
- 8. Seber G.: Multivariate Observations. Wiley, (1984)

Towards the Computer-aided Diagnosis of Dementia based on the Geometric and Network Connectivity of Structural MRI Data

Garry M. Smith¹², Zhivko V. Stoyanov¹², Danica V. Greetham³, Peter Grindrod⁴, J. Doug Saddy², and for the Alzheimers Disease Neuroimaging Initiative^{*}

¹ School of Systems Engineering, University of Reading, Reading, RG6 6AY, UK,

 $^2\,$ Centre for Integrative Neuroscience and Neurodynamics, University of Reading, RG6 6AH, UK

³ Centre for the Mathematics of Human Behaviour, Department of Mathematics and Statistics, University of Reading, RG6 6AX, UK

⁴ Mathematical Institute, University of Oxford, Oxford, OX2 6GG, UK

Abstract. We present an intuitive geometric approach for analysing the structure and fragility of T1-weighted structural MRI scans of human brains. Apart from computing characteristics like the surface area and volume of regions of the brain that consist of highly active voxels, we also employ Network Theory in order to test how close these regions are to breaking apart. This analysis is used in an attempt to automatically classify subjects into three categories: Alzheimer's disease, mild cognitive impairment and healthy controls, for the CADDementia Challenge.

Keywords: MRI, dementia, mild cognitive impairment, voxel, automatic, diagnosis, graph Laplacian, Network Theory

1 Introduction

The UK government reports⁵ that there are currently 800,000 dementia sufferers in the UK alone, with the disease costing the economy £23 billion per year. By 2040 the government estimates the costs associated with the disease

^{*} Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

⁵ https://www.gov.uk/government/policies/improving-care-for-people-with-dementia

will triple as the number of dementia sufferers increases to 1.6 million people. The government is responding through a number of initiatives including: The National Dementia Strategy⁶ published in 2009; The Prime Minister's Dementia Challenge⁷ launched in 2012; and by increasing the annual funding of dementia research to $\approx \pounds 66$ million by 2015. As well as gaining a better understanding of the disease, the government aims to increase diagnosis rates so that they are among the best in Europe. Clearly the risk of neurological diseases, such as Alzheimer's disease (AD), to public health is of international importance and as such cross-cutting, interdisciplinary research combining ideas from across fields has the potential to contribute to the future well-being of global health.

The CADDementia Challenge⁸ was established by the Biomedical Imaging Group Rotterdam, Erasmus MC, Rotterdam, in order to provide a standardised evaluation framework for the Computer-Aided Diagnosis of Dementia, based on T1-weighted structural Magnetic Resonance Image (MRI) data. The primary aim of the CADDementia Challenge is to objectively validate the different imagebased diagnosis/classification methods that are emerging from research centres, such that suitably robust techniques may be identified as candidates for clinical use. The CADDementia competition requires participants to⁸: i. use a common dataset for training algorithms, as well as ii. a previously unseen multi-centre test dataset (to avoid over-training), and iii. to perform a multi-class diagnosis of Alzheimer's disease, Mild Cognitive Impairment (MCI) and controls.

We present an intuitive geometric algorithm for analysing the structure and fragility of MRI data. Apart from computing characteristics like the surface area and volume of regions of the brain that consist of highly active voxels, we also employ Network Theory[2] in order to test how close these regions are to breaking apart. This analysis is used in an attempt to classify structural MRI scans into three categories: CN (controls), MCI and AD.

The algorithm presented in this paper was executed over MRI data from the CADDementia and ADNI datasets using up to 120 computer CPU cores simultaneously to perform the analysis. However, with minor modifications to the job submission and data handling mechanisms currently employed, the same algorithm could be scaled up to massive numbers of CPU cores, provisioned on demand, through private and/or public Cloud providers, thereby potentially, allowing health authorities to offer screening as part of routine health checks to a larger proportion of the population and at greater frequencies.

2 The geometric and network structure of MRI data

Mathematically, MRI data can be considered as a collection of small cuboids (voxels) and for each of them we have a non-negative value which represents tissue properties. A central point in our approach is that in AD part of the neural

⁶ https://www.gov.uk/government/publications/living-well-with-dementia-anational-dementia-strategy

⁷ http://dementiachallenge.dh.gov.uk/

⁸ http://grand-challenge.org/site/caddementia/home/

mass degenerates progressively, therefore it is reasonable to assume that negative changes in the T1 signal gradients would be a feature of neural degeneration (this would be reverse for T2). We use these changes in voxel values as markers of degeneration in order to trace a path of similarity over long distances in the brain.

A 3D T1-weighted MRI image consists of $n_1 \times n_2 \times n_3$ voxels and $f(i, j, k) \ge 0$ is the level of T1-weighted signal recorded in voxel (i, j, k). Then we define

$$M := \max\{f(i, j, k) \mid 1 \le i \le n_1, 1 \le j \le n_2, 1 \le k \le n_3\}.$$

Further, for each voxel, instead of the recorded signal, f(i, j, k), we consider a normalised signal,

$$g(i,j,k) := \frac{f(i,j,k)}{M},\tag{1}$$

where we now have $0 \leq g(i, j, k) \leq 1$ for all voxels; we normalise the signal according to (1) for each brain. We can think of this as a way of introducing a similar scaling across all brains. The idea is to focus on the part of the brain (i.e. on those voxels) for which the signal is above a certain threshold, θ . Mathematically, this set is denoted by

$$A_{\theta} := \{ (i, j, k) \mid g(i, j, k) \ge \theta, 1 \le i \le n_1, 1 \le j \le n_2, 1 \le k \le n_3 \}.$$

We work with $\theta = 0.6, 0.61, 0.62, \ldots, 0.8$. This follows from both computational and physiological reasons: computing eigenvalues for a whole brain matrix is computationally intractable, but more importantly starting at a 0.6 threshold allows us to generate connectivity networks based on primarily white matter values (see Figures 1 & 2).

For each brain, we consider the 3D set A_{θ} and compute its surface area, S_{θ} , and its volume, V_{θ} ; we also compute a measure of the fragility of its structure, f_{θ} , i.e how close A_{θ} is to "breaking" apart into smaller components.

Apart from being a geometrical 3D object, we can think of A_{θ} as a network, denoted by N_{θ} , in which two voxels are connected if they share a face or an edge (but not a corner).

The advantage of interpreting A_{θ} , as a graph, or a network N_{θ} , is that we can apply certain techniques from Spectral Graph Theory. Each graph/network can be represented with a matrix. Computing eigenvalues of such a matrix gives us a spectrum - an array of values that describes some structural characteristics of the given graph. The most widely used matrices assigned to graphs are adjacency, Laplacian or normalised Laplacian matrices. For a review of using spectra of graphs in computational biology see [1]. A comprehensive study of a normalised Laplacian spectrum with detailed definitions and many examples of its applications is given in [2]. A useful property of the eigenvalues of the normalised Laplacian matrix is that they are all real and are between 0 and 2 for any number of vertices and edges. Furthermore, the number of zero eigenvalues correspond to the number of connected components of the corresponding graph [2]. The smallest positive eigenvalue of the Laplacian matrix is called *algebraic connectivity* [4] and is an indicator of the robustness of the graph [7] to



Fig. 1. A histogram showing the distribution of the intensities across a brain; the two peaks roughly indicate the range of intensities for white and grey matter. We can see that by choosing $\theta \geq 0.6$ we predominantly select white matter.



Fig. 2. Coronal view of a single brain depicting the tissue that is selected as θ increases from 0.6 to 0.8. Here, for illustrative purposes, the fraction θ is over the 95th percentile value, whereas in reality we work with a fraction θ of the absolute maximum. The main disadvantage of the latter approach is that it is sensitive to noise, but one of the advantages is that the size of the brain that is left after thresholding becomes computationally manageable.

vertex and edge failures and to betweenness in networks, which can help with identifying communities [8]. The eigenvector corresponding to algebraic connectivity, but also to the second smallest normalised Laplacian eigenvalue is used for spectral clustering [10].

Now, if A_{θ} is split into *m* disjoint parts, this will correspond to N_{θ} consisting of *m* connected components, which in turn corresponds to *m* eigenvalues equal to zero in the normalised Laplacian spectrum of N_{θ} . The eigenvalues close to zero (around the second smallest normalised Laplacian eigenvalue) give us an indication of the fragility of A_{θ} . The larger the number of eigenvalues that are close to zero, the more fragile (i.e. sensitive to breaking apart) A_{θ} is. Hence, given a threshold, θ , we denote by f_{θ} the number of eigenvalues that are close to zero in that particular N_{θ} , and call this *fragility*. Here, by 'close to zero' we mean those eigenvalues that are less than 0.001, a number which we determined experimentally.

Additionally, we have to compute the surface area, S_{θ} , and the volume, V_{θ} ; since the values of n_1 , n_2 and n_3 (defined at the beginning of this section) can be different (for example, between the three centres EMC, UP and VUMC in the training and test sets) we assumed in our computations that the edges of a single voxel are equal to $\frac{1}{n_1}$, $\frac{1}{n_2}$ and $\frac{1}{n_3}$, respectively. From here one can compute the area of each face of a voxel, as well as its volume (the latter is equal to $\frac{1}{n_1 n_2 n_3}$). Once re-scaled the surface area and volume of A_{θ} can be computed.

We calibrated the algorithm against the CADDementia training set by combining S_{θ} (surface area), V_{θ} (volume) and f_{θ} (fragility), with the age of the subject and used these four features (numbers) as *predictors* for the stage of neural degeneration (CN, MCI or AD). We firstly used gender to split the subjects apart into two groups.

For illustration purposes, let us consider the group of 13 females (out of 30 subjects) from the training set. For a fixed threshold, θ , we use multinomial logistic regression, which is discussed in detail in [5], [9] and [3]. Specifically, MATLAB is used to compute the multinomial logistic regression (the function mnrfit) with predictors $X_{\theta} = [\text{age}, S_{\theta}, V_{\theta}, f_{\theta}]$ and the responses, Y, are the labels for the diagnoses of the subjects (0 = CN (control); 1 = MCI; and 2 = AD). As an output of mnrfit we get a matrix of coefficient estimates, B_{θ} . This derives B_{θ} , we then remove the labels, Y (diagnoses); using only B_{θ} and (the same) predictors, X_{θ} , we compute the probabilities for each subject being diagnosed with CN, MCI or AD. The latter probabilities are computed using the MATLAB function mnrval. The output for $\theta = 0.66$ is given in Table 1 (columns 4, 5 and 6). Amongst the probabilities p_{CN} , p_{MCI} and p_{AD} we choose the highest, and this determines the class to which a subject is assigned (the third column of Table 1).

For each θ , as in Table 1, we compare our predictions with the set of diagnoses and choose the values of θ for which we get best agreement. In all the tests we performed on the CADDementia training set as well as with a large subset of the ADNI dataset, we consistently found, across both male and female groups, that $\theta = 0.63, 0.64, 0.65, 0.66$ and $\theta = 0.71, 0.72$ gave the best fit between the diagnoses and our predictions. However, in our submission for the CADDementia challenge we chose $\theta = 0.71$ for females and $\theta = 0.63$ for males because these were the most stable values we were getting, that is, small changes in θ did not lead to significant changes in our predictions.

Table 1. The output from the MATLAB function mnrval applied to the group of female subjects on the training CAD dataset, $\theta = 0.66$. The second column, that is, the diagnosis of each subject, is only given as a reference here. The input parameters to the the function mnrval are only the matrix B_{θ} and the (matrix of) predictors, X_{θ} . The function mnrval outputs the probabilities p_{CN} , p_{MCI} and p_{AD} .

subject ID	diagn.	predict.	p_{CN}	p_{MCI}	p_{AD}
train_emc_002	2	1	0	0.78	0.22
train_emc_003	0	0	0.99	0.005	0.003
train_emc_008	0	0	0.89	0.0008	0.1
train_emc_009	2	2	0	0	1
train_emc_011	1	1	0	0.87	0.13
train_up_001	2	2	0	0.004	0.995
train_up_002	1	1	0	0.64	0.36
train_vumc_004	2	2	0	0.03	0.97
$train_vumc_005$	0	0	0.98	0.01	0.01
train_vumc_008	2	2	0	0	1
train_vumc_010	1	1	0.004	0.92	0.08
$train_vumc_012$	1	1	0.05	0.74	0.21
train_vumc_013	2	2	0.1	0.001	0.90

Given our value for θ we compute the matrix B_{θ} on female subjects in the CADDementia training set. Further, we can find the corresponding predictors, [age, $S_{\theta}, V_{\theta}, f_{\theta}$], for the female subjects from the CADDementia test data set. Therefore, we can use B_{θ} with those new predictors as input parameters to the function mnrval and derive the corresponding probabilities, p_{CN}, p_{MCI} and p_{AD} .

3 Materials

The algorithm described in §2 was tested against the CADDementia training set as well as suitable data from the ADNI database as described next.

3.1 CADDementia Data

The CADDementia dataset⁹ comprises of 384 T1-weighted 3T MRI scans in gzipped Nifti format of subjects with AD, MCI and healthy controls, that were captured from multiple different centres. The original data (6.2 Gbytes¹⁰) as well as a non-uniformity corrected version (16 Gbytes¹⁰) is provided and all data is reported as being clinically-representative. A training subset comprising of 30 scans that are reported to be equally distributed over the originating centres, as well as the corresponding diagnostic labels are provided. The demographic metadata comprises of age and gender. Note that we used the non-uniformity-corrected version of the dataset.

 $^{^9~{\}rm http://grand-challenge.org/site/caddementia/download_data$

¹⁰ gzip compressed.

3.2 ADNI Data

The Alzheimers Disease Neuroimaging Initiative (ADNI) database¹¹ was launched in 2003 by the National Institute on Ageing (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organisations, as a \$60 million, 5-year public-private partnership. 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 Alzheimers disease (AD). The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date, the ADNI Website reports that these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD.

Only T1-weighted structural 3T MRI data was selected for use in our investigation as contained in the following ADNI collections:

- $AD-\{bl,m06,m12,m24\}-3.0T$
- ADNI1:{Baseline,Annual 2Yr,Complete{1,2,3}Yr}
- MCI-{bl,m06,m12,m18,m24,m36}-3.0T
- $Normal-m{06, 12, 24, 36}-3.0T$

resulting in multiple scans being retrieved for each of 189 subjects (33 Gbytes non-compressed), with age ranging from 58 to 93 years old.

4 Compute platform

The computers used for this study (see Table 2) comprised of workstations from our Analysis Laboratory (AL) as well as our small Infiniband-connected compute cluster (IB), all of which combine to report a total of 120 Intel CPU cores, to a HTCondor¹² batch queue. The Gentoo Linux¹³ operating system, with kernel 3.14.4-gentoo x86_64 was used across all machines. The notation 'S-R' in the *Disk* column indicates that the system disk is SSD and the scratch disk is rotational. The *HT* column indicates if hyper-threading¹⁴ was enabled.

5 Workflow and execution times

Conceptually the algorithm and all processing presented in this paper is entirely automated, however in terms of our current workflow there are presently a number of steps between the stages, that we invoke manually (e.g. instruct a script to execute).

¹¹ http://adni.loni.usc.edu

¹² HTCondor version 7.8.8, see http://research.cs.wisc.edu/htcondor

¹³ http://www.gentoo.org

¹⁴ http://en.wikipedia.org/wiki/Hyper-threading
$\# \operatorname{Grp}$	Nodes(s)	Qty Model	GHz	#	#	HT	${\rm Mem}~{\rm GPGPU}$	Disk	Network
				CPU	Cores		(GB)		
$1 \mathrm{AL}$	$ncpc{50-62}$	13 i7-4770	3.40	1	8	\checkmark	32 C2050	S-S	GigE
2 AL	ncpc5	1 i 7 920	2.67	1	4	\times	24 C2050	S-S	GigE
3 IB	ncpc139	1 X3460	2.80	1	4	\times	16 -	S-R	QDR IB
4 IB	$ncpc14\{1,2\}$	2 i 7 950	3.07	1	4	\times	24 -	S-R	QDR IB
6 IB	$ncpc14{5-9}$	5 X 5570	2.93	2	8	\times	24 -	S-R	QDR IB
$7 \mathrm{IB}$	ncpc150	1 X5690	3.46	2	12	\times	$48.3 \times C2050$	S-S	QDR IB
$10 \ \mathrm{SRV}$	$ncsrv\{1,2,3,4\}$	4 -	-	-	-	-		R-R	QDR IB

Table 2. The computational resources used during this study

The workflow is described thus:

- 1. Compile Matlab code to a standalone binary The Matlab code described in §2 is compiled to a standalone binary using the Matlab Compiler in preparation for license-free parallel execution¹⁵ across the compute nodes.
- 2. Download the CADDementia and ADNI Data Given the CADDementia data consisted of relatively few files, we elected to download manually, although this could potentially have been scripted. We obtained the ADNI data via¹⁶.
- 3. Unpack data This operation is scripted and can either be performed sequentially from a single machine or in parallel using, e.g. our HTCondor installation.
- 4. Brain extraction Brains are extracted from all the scans using the following FMRIB FSL¹⁷ roi and bet commands in a data parallel way on our cluster: standard_space_roi <in_file> <intermediate_file> -b; bet <intermediate_file> <output_file> -f 0.15
- 5. **Process the data** The compiled Matlab code is executed in a data parallel way (each data file is submitted as an independent task to the batch queue) via our computational resources. Results are written back to a single location on our network file system.
- 6. Final classification The results from the previous step (together with some of the demographics) are used as *predictors* (discussed in §2) in order to do the final classification of the subjects (into the following classes: CN, MCI and AD). The classification is done in MATLAB, using the functions mnrfit and mnrval.

The preprocessing task in workflow step #4 (brain extraction) takes less than 90 seconds for a single brain scan, and ≈ 15 minutes for all 354 CADDementia test brains to be extracted in parallel across the cluster (sequentially this same operation could take up to ≈ 8 hours to complete on a single computer).

¹⁵ http://www.mathworks.co.uk/products/compiler/mcr

¹⁶ https://ida.loni.usc.edu/login.jsp

¹⁷ FMRIB Software Library, http://fsl.fmrib.ox.ac.uk/fsl/fslwiki

The length of the computation (workflow step #5) of a single MRI scan depends, mostly, on the amount of voxels that were left in the set A_{θ} (defined in $\S2$) and also, in the computation of the spectrum of the normalised Laplacian matrix, on how fragile or connected A_{θ} is as a 3D structure. Therefore, MRI scans in which the set A_{θ} was rather large took longer to compute (the length being dependant on the complexity of the MATLAB eigenvalue solver, eigs). For example, workflow step #5 takes anywhere between 6 and 24 minutes for the majority of the brain scans processed. However, twenty six of the 354 CADDementia test scans took considerably longer with, for example, stripped_test_vumc_116.nii, requiring ≈ 60 hours of actual compute time and a memory footprint of 4.1Gbytes. Moreover, despite having an otherwise empty batch queue there were more data items to be processed from the CADDementia test dataset than the available processors (exposed by the batch queue) across our local compute resources, hence the overall end-to-end wall clock time to getting a result for the scan was in fact $87\frac{1}{2}$ hours, after taking into account the time the job spent languishing in the queue. The final classification (workflow step #6) takes less than a minute on a single machine.

The condor_submit files for workflow steps #3,4,5 are automatically generated by a BASH¹⁸ script that dynamically identifies compute tasks based on the previously downloaded data.

6 Conclusions

The results that we observed while testing our method with the CADDementia training set (consistently, less than 20% incorrect predictions) and the ADNI dataset (consistently, less than 35% incorrect predictions) appear promising. The approach we have presented here is intuitive and easy to implement. We believe it is a potential step towards employing Network Theory in the analysis and classification of neural diseases, and as such it can be extended to include more sophisticated techniques from Network Theory.

This technique is agnostic to underlying tissue properties as well as to the nature of the signal. We have previously applied a similar approach to resting state fMRI data [6]. For the purposes of this competition we have intentionally biased the algorithm in favour of white matter by stepping up the threshold values. Alternatively, we could use a step down method to capture properties of grey matter. It is also possible to target specific tissues using a gating procedure.

A further advantage, which we consider important, is that the workflow can be fully automated and potentially, this can be implemented in such a way that, for example, a computer Web Service running in the cloud would ingest MRI data and return a diagnosis within a short time over the network; issues of trust, privacy, authorisation and security would need to be looked at to ensure compliance with the legislation for medical data protection.

The work we have described takes a data parallel approach to processing many scans at the same time, and for the most part, results are obtained within

¹⁸ http://en.wikipedia.org/wiki/Bash_%28Unix_shell%29

a matter of minutes. However as we have seen, MRI scans in which the set A_{θ} is rather large can take substantial amounts of time to compute, and thus are candidates for a finer granularity of parallelism, which we shall investigate in future work. In addition to increasing the speed of the diagnosis we can also improve accuracy by applying our approach to the parts of the brain whose structural changes are known to be highly correlated with the presence or absence of AD and/or MCI.

7 Acknowledgements

We thank Tom Johnstone and Shan Shen from CINN, for their helpful comments and suggestions. We gratefully acknowledge the support of the University of Reading Research Endowment Trust Fund (research fellowship for GS). The computational aspects of this work were based on a foundation previously laid down during the EPSRC NeuroCloud project under grant EP/I016856/1. Data collection and sharing for this project was partly funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award W81XWH-12-2-0012).

References

- Anirban Banerjee and Joergen Jost. Graph spectra as a systematic tool in computational biology. Discrete Applied Mathematics, 157(10):2425 – 2431, 2009. Networks in Computational Biology.
- Fan R. K. Chung. Spectral Graph Theory. CBMS Regional Conference Series in Mathematics. AMS, 1997.
- 3. A. J. Dobson and A. G. Barnett. An Introduction to Generalized Linear Models. Chapman and Hall/CRC. Taylor and Francis Group, 2008.
- M. Fiedler. Algebraic connectivity of graphs. Czechoslovak Mathematical Journal, 23:298–305, 1973.
- P. Grindrod. Mathematical Underpinnings of Analytics. Oxford University Press, 2014.
- Peter Grindrod, Zhivko V Stoyanov, Garry M Smith, and J Doug Saddy. Primary evolving networks and the comparative analysis of robust and fragile structures. *Journal of Complex Networks*, September 2013.
- A. Jamakovic and S. Uhlig. On the relationship between the algebraic connectivity and graph's robustness to node and link failures. In Next Generation Internet Networks, 3rd EuroNGI Conference on, pages 96–102, May 2007.
- Javier Martn Hernndez, Zongwei Li, and Piet Van Mieghem. Weighted betweenness and algebraic connectivity. *Journal of Complex Networks*, 2014.
- P. McCullagh and J. A. Nelder. *Generalized Linear Models*. New York: Chapman and Hall, 1990.
- J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE. Trans. on* Pattern Analysis and Machine Intelligence, 22:888–905, 2000.

Dementia Diagnosis using MRI Cortical Thickness, Shape, Texture, and Volumetry

Lauge Sørensen¹, Akshay Pai¹, Cecilie Anker², Ioana Balas¹, Martin Lillholm³, Christian Igel¹, Mads Nielsen¹

¹ Department of Computer Science, University of Copenhagen, Denmark ² DTU Compute, Technical University of Denmark, Denmark ³ Biomediq A/S, Denmark

Abstract. This paper proposes a structural magnetic resonance imaging (MRI) biomarker that combines a range of individual biomarkers (cortical thickness measurements, hippocampal shape, hippocampal texture, and volumetric measurements) for the purpose of multi-class classification of Alzheimer's disease, mild cognitive impairment and, normal controls. The combination is achieved by entering the biomarkers as features in a linear discriminant analysis. The fully automated method is trained on a combination of two publicly available datasets and is evaluated on the training set from the CADDementia challenge. Test set scores using two different priors are submitted to the same challenge.

1 Introduction

Structural magnetic resonance imaging (MRI) is an integral part of the diagnostic work flow in many memory clinics. The modality allows for non-invasive in vivo inspection of the degree and the location of brain atrophy, a hallmark of several dementias including Alzheimer's disease (AD), the most frequent type. The importance of structural MRI has been underlined by the inclusion of MRI volumetry as a surrogate biomarker of atrophy in international diagnostic guidelines for AD [13] and its prodromal stage, mild cognitive impairment (MCI) due to AD [2].

Volumetry, and in particular hippocampal volumetry, is in general the most widely studied and used MRI biomarker of AD, and there are already efforts towards standardization of this biomarker [8, 12]. However, it is evident that there are other sources of information, than what is captured by volumetry, to extract from a structural MRI scan. This include cortical thickness [14] as well as less established biomarkers such as the hippocampal shape [1] and the textural patterns within the hippocampal tissue [15, 17]. Both shape and texture have shown to provide volume-independent diagnostic or prognostic information, and to improve prediction of conversion from MCI to AD when combined with volume [1, 15, 17]. Cortical thickness may be more reliable than volume in detecting differences between MCI and AD [14] and combining cortical thickness measurements with volumetric measurements has shown good NC vs. AD discrimination [18].

In this study, we propose to combine a range of volumetric measurements with cortical thickness measurements, hippocampal texture, and shape, in order to obtain a combination biomarker that uses more of the information contained in a structural MRI scan. Such a biomarker has potential of improved diagnosis of MCI and AD compared to, e.g., a pure volumetry-based biomarker. To the best of our knowledge, this is a unique combination of basic MRI biomarkers not tried before. The combination is achieved by entering all biomarkers as features in a linear discriminant analysis (LDA). The proposed method is developed and trained on a combination of MRI scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the MRI imaging arm of the Australian Imaging, Biomarker & Lifestyle Flagship Study of Aging (AIBL). The trained method is then evaluated on the training set from the Computer-Aided Diagnosis of Dementia based on structural MRI data challenge⁴ (CADDementia), a challenge at the 17th International Conference on Medical Image Computing & Computer-Assisted Intervention. Two different test set scores using different priors are further submitted to the CADDementia challenge.

2 Data

The following five datasets are used: the "complete annual year 2 visits" 1.5-T dataset from the collection of standardized datasets recently released by ADNI [19]; a subset of manual hippocampal segmentations from the Harmonized Hippocampal Protocol (HHP) [8] and associated MRI scans; the MRI imaging arm of AIBL [5]; and the CADDementia training and test sets. Table 1 summarizes the characteristics of the datasets.

The ADNI dataset and the AIBL dataset are merged into one combined dataset (termed ADNI+AIBL) that is used for training, and HHP is used in a special purpose hippocampal segmentation method described in Section 3.3. The CADDementia training set is used for evaluation. Finally, the CADDementia test set is classified using the trained combination MRI biomarker, and the obtained scores are submitted to the CADDementia challenge.

All MRI scans were conformed to $1 \times 1 \times 1 \text{ mm}^3$ resolution followed by bias correction. Both operations were performed using FreeSurfer (version 5.1.0, default parameters) [7].

3 Individual MRI Biomarkers

A range of structural MRI biomarkers are used in the proposed combination biomarker, which are the following: volumetry of brain structures and of the ventricles, cortical thickness measurements, hippocampal shape, and hippocampal texture. These biomarkers are detailed in the following subsections, and an overview is provided in Table 2.

⁴ http://caddementia.grand-challenge.org/

		n	Age, years $(\text{mean}\pm\text{std})$	Male (%)	MRI field strength $(1.5-T/3-T)$
ADNI	Total NC MCI AD	504 169 234 101	75.3 ± 6.5 76.0 ± 5.1 74.8 ± 7.0 75.3 ± 7.4	57.9 50.9 66.5 50.5	504/0 169/0 234/0 101/0
ННР	Total NC MCI AD	40 12 11 17	$74.1 \pm 7.4 \\76.9 \pm 6.2 \\70.9 \pm 6.8 \\74.2 \pm 8.6$	$\begin{array}{c} 47.5 \\ 41.7 \\ 54.6 \\ 47.1 \end{array}$	40/0 12/0 11/0 17/0
AIBL	Total NC MCI AD	145 88 29 28	$75.4 \pm 7.4 \\ 75.2 \pm 7.2 \\ 77.5 \pm 7.1 \\ 73.6 \pm 8.1$	$\begin{array}{c} 44.6 \\ 47.7 \\ 51.7 \\ 35.7 \end{array}$	1/144 1/87 0/29 0/28
CADDementia train	Total NC MCI AD	30 12 9 9	65.2 ± 7.0 62.3 ± 6.3 68.0 ± 8.5 66.1 ± 5.2	43.3 25.0 44.4 66.7	0/30 0/12 0/9 0/9
CADDementia test	Total	354	$65.1 {\pm} 7.8$	60.2	0/354

Table 1: Characteristics of the datasets.

3.1 FreeSurfer Volumetry

Sub-cortical and ventricular volumetric measurements were computed using crosssectional FreeSurfer (version 5.1.0, default parameters) [7]. We used measurements from ROIs provided by FreeSurfer (i.e., in the Aseg atlas). Bilateral ROIs were joined. In addition to individual ROIs, we also computed total ventricular volume and whole brain volume, resulting in a total of 7 volumetric FreeSurfer measurements. All volumetric measurements were normalized for head size by dividing by the intra-cranial volume (ICV) also computed during the crosssectional FreeSurfer pipeline.

3.2 FreeSurfer Cortical Thickness

Cortical thickness measurements were computed using cross-sectional FreeSurfer (version 5.1.0, default parameters) [6]. We used measurements from the ROIs in the Desikan-Killiany atlas that were joined into the four lobes and the cingulate cortex⁵. Left and right hemispheres were further joined, resulting in a total of 5 cortical thickness measurements. We did not normalize cortical thickness measurements for head size (i.e., ICV) [18].

⁵ http://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation

3.3 Hippocampal Volume

In addition to FreeSurfers estimate of the hippocampal volume, we also computed the hippocampal volume using a special purpose algorithm. This was motivated by the fact that hippocampal volume is the most widely used MRI biomarker of AD [12] and FreeSurfer is not optimized for this structure specifically. The leftand right hippocampus were segmented separately using a multi-atlas, affine registration, non-local patch-based segmentation (N-L Patch) technique [4, 3]. The atlas comprised 40 segmentations from HHP [8] (12 NC, 11 MCI, 17 AD). All 40 HHP segmentations were used as atlases during pre-selection, but only the 9 most similar contributed to the final segmentation. A subset of 15 HHP segmentations were used to cross-validate parameters (number of atlases used after pre-selection, cubic patch size, and search volume size) using Dice's coefficient. The bilateral volume was computed and divided by FreeSurfers estimate of ICV. N-L Patch has previously demonstrated a better AD diagnostic performance than static FreeSurfer [4].

3.4 Hippocampal Shape

Two hippocampal shape scores (for the left and right hippocampus, respectively) were computed as well. In a spirit similar to [1], a shape descriptor was computed by aligning each hippocampus surface to a template hippocampus using iterative closest point (ICP), followed by a mapping of 30 uniformly distributed landmarks from the template to the hippocampus. The set of hippocampi, each now represented by 30 landmarks, were all aligned using generalized Procrustes alignment [9]. Finally, principal component analysis was applied the set of aligned hippocampus landmarks, and the components explaining 90 % of the variance were retained. This representation was used as features in a naive Bayes classifer. The feature extraction was performed on all data simulteanously, i.e., on the combination of ADNI+AIBL and the CADDementia data. Subsequently, only NC and AD observations from ADNI+AIBL were used for training of the naive Bayes classifier. The trained classifier was finally applied to score the CADDementia data. The FreeSurfer hippocampus segmentation was used to defined the ROI in each MRI scan. The whole procedure was computed for the left and the right hippocampus separately, resulting in two hippocampal shape scores.

3.5 Hippocampal Texture

A hippocampal texture score was computed using a texture descriptor recently proposed for quantification of chronic obstructive pulmonary disease in computed tomography [16] in combination with a support vector machine (SVM) with a radial Gaussian kernel. This specific MRI biomarker has previously shown good results [15, 17]. The texture descriptor comprised marginal filter response histograms of a 3-dimensional, rotation-invariant, multi-scale, Gaussian derivative-based filter bank with the following scales: 0.6, 0.86, 1.2, and 1.7 mm. Compared to [16], the Gaussian filter was excluded in order to be invariant to the lack

of a standardized scale in MRI. The morphologically post-processed bilateral FreeSurfer hippocampus segmentation was used to define the ROI. The texture biomarker was trained on all NC and AD observations from ADNI+AIBL. The training involved estimation of adaptive histogram binning for the different filters, and training of the SVM to separate NC from AD. During SVM training, the width of the radial Gaussian kernel and the regularization parameter was estimated using grid search in a nested cross-validation loop. The SVM was subsequently trained on all training data using the optimal parameter combination.

4 Combination Biomarker

The individual MRI biomarkers $\{x^{(i)}\}_{i=1...N}$ were combined by entering them as features to a regularized linear discriminant analysis (LDA) with λ added to the diagonal of the covariance matrices [10]. Prior to entering the LDA, each individual biomarker x was z-score transformed dependent on the age of the subject according to $z = (x - \mu_{age})/\sigma_{age}$. The age-dependent weighted mean, μ_{age} , and the age-dependent weighted standard deviation, σ_{age} , of the biomarker used in the transformation were estimated from the training set using an adaptive width Gaussian interpolation kernel centered on the respective age. The agedependent z-score transformation was applied within each group, resulting in a tripling of the features $\{z_{\rm NC}^{(i)}, z_{\rm MCI}^{(i)}, z_{\rm AD}^{(i)}\}_{i=1...N}$. The LDA with $\lambda = 0.001$ was trained directly for the three-class problem of discriminating NC, MCI, and AD using ADNI+AIBL, and the Shark C++ library was used for this purpose [11].

Table 2: Overview of individual MRI biomarkers. Hippocampal shape and hippocampal texture uses the FreeSurfer hippocampal segmentation as ROI. FreeSurfer is not trained.

Biomarker	Segmentation method	Training		
Cortical thickness				
Frontal lobe	FreeSurfer	-		
Parietal lobe	FreeSurfer	-		
Temporal lobe	FreeSurfer	-		
Occipital lobe	FreeSurfer	-		
Cingulate cortex	FreeSurfer	-		
Volumetry				
Amygdala	FreeSurfer	-		
Caudate nucleus	FreeSurfer	-		
Hippocampus	FreeSurfer/N-L Patch	$-/\mathrm{HHP}$		
Pallidum	FreeSurfer	-		
Putamen	FreeSurfer	-		
Ventricular	FreeSurfer	-		
Whole brain	FreeSurfer	-		
Hippocampal shape	FreeSurfer	ADNI+AIBL		
Hippocampal texture	FreeSurfer	ADNI+AIBL		

5 Results

The combination biomarker was evaluated by training on ADNI+AIBL and scoring of the CADDementia training dataset. We also report average 10-fold cross-validation performance on ADNI+AIBL with splits stratified on group and dataset. The python script supplied by the CADDementia team is used for this purpose, and the computed performance measures are described on the CADDementia website⁶. The results are summarized in Table 3, and associated receiver operating characteristic (ROC) curves and areas under the ROC curves (AUCs) are shown in Figure 1 and confusion matrices in Table 4.

Table 3: Performance measures. AUC classification true positive fraction Total NC MCI AD NC MCI AD accuracy CADDementia train 73.377.8 $91.7 \ 44.4$ 83.2 86.6 68.3 95.8 62.2 ADNI+AIBL 79.8 53.245.778.4 85.5 68.1 82.7

The proposed method is fully automated. Approximate computation time in order to classify a new MRI scan is presented in Table 5 where we also provide the computation time of individual components of the method.



Fig. 1: Per-class ROC curves and AUCs for the proposed combination biomaker.

Table 4: Confusion matrices. Rows are predicted and columns are true class. (a) CADDementia train (b) ADNI+AIBL

				()			
	NC	MCI	AD		NC	MCI	AD
NC	11	3	0	NC	205	74	13
MCI	1	5	2	MCI	48	140	57
AD	0	1	7	AD	4	49	59

⁶ http://caddementia.grand-challenge.org/evaluation/

Table 5: Per subject	computation	time divided	into comp	onents.
Total (+FreeSurfer) Fre	Surfer N_L P	atch Shape	Texture	Combine

iotai (†iiceballei)	ricosurior	н п п атон	Shape	rontaro	Combine
~ 1 (19) h	$\sim 18 \ h^1$	$\sim 40 \min$	$\sim 0.5 \text{ min}$	$\sim \! 15 \min$	$\sim 1 \text{ sec}$

¹ Median processing time due to high variability with some extreme outliers.

As seen from the confusion matrices, there was a tendency of classifying subjects as too healthy (MCI as NC, AD as MCI). We therefore also produced a second score using the following priors optimized for the CADDementia training set: P(NC) = 1/8, P(MCI) = 3/8, P(AD) = 1/2. These priors resulted in a classification accuracy of 80 % on the CADDementia training set and in a more balanced confusion matrix. The results of using these priors were submitted to the challange as *LDA-optimized-priors* whereas the previous results were submitted as *LDA-equal-priors*.

6 Discussion and Conclusion

In this paper, we proposed to combine a range of structural MRI biomarkers for the purpose of multi-class classification of AD, MCI and NC. The individual biomarkers used in the combination were cortical thickness measurements, hippocampal shape and texture, and volumetric measurements. Combining such diverse biomarkers may potentially improve diagnostic performance from structural MRI. This is appealing because the modality is less invasive than state-ofthe-art biomarkers based on lumbar puncture and positron emission tomography imaging that are directly measuring pathological hallmarks such as amyloid load. However, despite this potential improvement and despite the inclusion of biomarkers such as hippocampal texture that is sensitive to earlier stages of the disease process, there were problems discriminating MCI from AD and NC (see Table 4). Advancing performance further would probably need combination with other non-structural MRI biomarkers such as the aforementioned.

Irrespective of this, a dementia diagnosis is in practice based on several sources of information, such as neuropsychological assessment, physical examination, blood sampling, and visual inspection of some form of anatomical medical imaging (e.g., computed tomography or structural MRI), and structural MRI biomarkers should be used in conjunction with all this information. The proposed structural MRI combination biomarker is a promising direction for obtaining improved diagnostic information from MRI to be used in clinical assessment.

References

 Achterberg, H.C., et al.: Hippocampal shape is predictive for the development of dementia in a normal, elderly population. Hum Brain Mapp 35(5), 2359–2371 (2014)

7

 Albert, M.S., et al.: The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 7(3), 270–279 (2011)

- 3. Anker, C.: Segmentation of Subcortical structures in T1 weighted MRI as a component of a Brain Atrophy Computation Pipeline. Master's thesis, Technical University of Denmark, Department of Applied Mathematics and Computer Science (2014)
- Anker, C., et al.: Automated hippocampal segmentation using new standardized manual segmentations from the Harmonized Hippocampal Protocol. Alzheimers Dement, suppl. (2014), accepted.
- Ellis, K.A., et al.: The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. Int Psychogeriatr 21(4), 672–687 (2009)
- Fischl, B., Dale, A.M.: Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci USA 97(20), 11050–11055 (2000)
- Fischl, B., et al.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33(3), 341–355 (2002)
- 8. Frisoni, G.B., Jack, C.R.: Harmonization of magnetic resonance-based manual hippocampal segmentation: a mandatory step for wide clinical use. Alzheimers Dement 7(2), 171–174 (2011)
- 9. Gower, J.: Generalized procrustes analysis. Psychometrika 40(1), 33–51 (1975)
- Guo, Y., Hastie, T., Tibshirani, R.: Regularized linear discriminant analysis and its application in microarrays. Biostatistics 8(1), 86–100 (2007)
- Igel, C., Heidrich-Meisner, V., Glasmachers, T.: Shark. Journal of Machine Learning Research 9, 993–996 (2008)
- Jack, Jr, C.R., et al.: Steps to standardization and validation of hippocampal volumetry as a biomarker in clinical trials and diagnostic criterion for Alzheimer's disease. Alzheimers Dement 7(4), 474–485.e4 (2011)
- McKhann, G.M., et al.: The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 7(3), 263–269 (2011)
- Singh, V., et al.: Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer's disease. Brain 129(Pt 11), 2885–2893 (2006)
- Sørensen, L., et al.: Hippocampal texture provides volume-independent information for Alzheimer's disease diagnosis. Alzheimers Dement 8(4), suppl. (2012)
- 16. Sørensen, L., et al.: Texture-based analysis of COPD: A data-driven approach. IEEE Trans Med Imaging 31(1), 70–78 (2012)
- 17. Sørensen, L., et al.: Hippocampal texture predicts conversion from MCI to AD. Alzheimers Dement 9(4), suppl. (2013)
- Westman, E., et al.: Regional magnetic resonance imaging measures for multivariate analysis in alzheimer's disease and mild cognitive impairment. Brain Topogr 26(1), 9–23 (2013)
- Wyman, B.T., et al.: Standardization of analysis sets for reporting results from ADNI MRI data. Alzheimers Dement 9(3), 332–337 (2013)

8

MIND-BA: Fully automated method for Computer-Aided Diagnosis of Dementia based on structural MRI data.

S. Tangaro¹, P. Inglese¹, R. Maglietta², A. Tateo¹

 ¹ Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Italy
 ² Istituto di Studi sui Sistemi Intelligenti per l'Automazione, CNR, Bari, Italy for the Alzheimers Disease Neuroimaging Initiative **

Abstract. Neurodegenerative diseases are frequently associated with structural changes in the brain. Magnetic Resonance Imaging (MRI) scans can show these variations and therefore be used as a supportive feature for a number of neurodegenerative diseases. The hippocampus has been known to be a biomarker for Alzheimers disease and other neurological and psychiatric diseases. However, it requires accurate, robust and reproducible delineation of hippocampal structures. This work utilises a datasets consisting of MR images shared by EADC-ADNI working group . Hippocampus volume is a feature used in this analysis. For the other features we used publicly available brain segmentation package FreeSurfer v.5.1 (FS) (freesurfer.nmr.mgh.harvard.edu) [17] to process the structural brain MRI scans and compute morphological measurements. The FreeSurfer pipeline is fully automatic and provides 184 features per MRI scan in total. Volumes of cortical and sub-cortical structures such as the caudate and average thickness measurements within cortical regions, such as the precuneus. We use the FS features but for hippocampus volume we use the segmentation proposed in [16]. For the diagnosis classification we passed all the features to a C-Support Vector Classifier (C-SVC) with a linear kernel on a 5-fold cross validation. The goal is evaluating the performance of an algorithms for multi-class classification: AD, MCI and controls. Methods that are developed for binary classification can be used for three-way classification by using either a one-vs-one (ovo) or one-vs-all (ova) strategy. In this approach, three classifiers are trained for the three binary problems using the ovo methodology and thereafter their outputs are combined into three predictions.

^{**} Data used in preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

1 Introduction

In the last few years the enormous development of neuroimaging has deeply altered the research and clinical prospects in the field of neurodegenerative disorders. This has been particularly relevant for Alzheimerss disease (AD), the most common dementia in the world, affecting currently over 36 millions of people (World Alzheimer Report 2011), a number destined to grow due to the increasing aging population. AD is characterised by the formation and deposition of abnormal proteins in the brain, with subsequent functional disruption, neuronal suffering and cell death (neurodegeneration), the latter ultimately translated into a loss of brain volume (atrophy). The cognitive decline is related to the degree of brain atrophy, which accelerates with the progression of the disease, as detected on MRI at a rate of 2% per vear for the whole brain (versus 0.2 - 0.5%in normal aging), and at a rate of 5% per year for the hippocampus, a complex structure located in the medial temporal lobe with a primary role in memory and learning, thus making hippocampal atrophy the most important imaging biomarker of the condition [2]. It is not surprising, then, that the accurate measurement of hippocampal volume, is of crucial importance and has become the focus of an increasingly large body of work. Until recently the segmentation of the hippocampus, ie its identification and separation from surrounding brain structures, had been performed mainly manually or with semi-automated techniques, followed by manual editing. This is obviously time-consuming and subject to investigator variability, so a number of automated segmentation methods have been developed. These have relied so far mainly on image intensity-based methods, often adopting multi-atlas registration approaches, in order to minimize errors due to individual anatomical variation. More recently, though, a number of methods that exploit shape information have been developed, based on preliminary work carried out in the nineties with the Active Shape Models (ASM) [1] and the Active Appearance Models (AAM) [3]). ASM address the issue of identifying objects of a known shape in a digital image when the shape is characterized by a certain degree of variability, as in the case of anatomical structures. AAM combine grey-level information with shape information provided by a training set, but this may fail to capture the intrinsic variability of biological structures, a limit attempted to overcome by the use of the wavelet transform and the principal component analysis (PCA) [4]. Alternative methods have used deformable representations or deformable M-reps [5]. Also, algorithms that associate geometric information (obtained by expert priors or learning procedures in a Bayesian framework) to powerful statistical tools, such as region competition algorithms (Zhu and Yuille, 1996) have been combined with homotopic deformations in automated hippocampal segmentation methods [6]. Recent work has employed probabilistic tree frames for brain segmentation [7], at times adopting specific models such as Markovian random fields or graphical cuts [8]. The use of machine learning techniques enables the processing of high-dimensional feature vectors without time-consuming computations thanks to optimization procedures. Alternative approaches involve, for example, labeling strategies combined

with other methods such as multiple segmentations [9], longitudinal 4-D methods with graph cuts [10] and label fusion with template libraries [11].

2 Materials

This work utilises two datasets named DB - 1 and DB - 2. DB - 1 consists of 98 MR images and their corresponding expert manual labels. The dataset is shared by EADC-ADNI working group using a standard harmonized protocol (www.hippocampal-protocol.net). The most inclusive definition of the Harmonized protocol [12] may limit the inconsistencies due to the use of arbitrary lines and tissue exclusion of the currently available manual segmentation protocols. The second dataset used -DB - 2, is from ADNI screening images and cosists of 160 MR images. The two databases used are described with demographics given in table 1.

Data	Size	Age	M/F	Subjects
DB-1	98	60-90	56/44	29 NC - 32 MCI - 37 AD
DB-2	160	28 - 96	76/84	68 NC -63 MCI - 29 AD

Table 1. Description of database used. Group size, range age (years) and sex of the two clinical datasets, containing normal control (NC) subjects, Alzheimer's Disease (AD) and mild cognitive impairment (MCI) patients.

As data for the evaluation framework, CADDementia project composes a multi-center data set consisting of 384 scans. The participating centers are: Erasmus MC (EMC), Rotterdam, the Netherlands; VU University Medical Center (VUmc), Amsterdam, the Netherlands; University of Porto / Hospital de So Joo (UP), Porto, Portugal This data set contains structural MRI (T1w) scans of subjects with the diagnosis of probable Alzheimers disease (AD), mild cognitive impairment (MCI) and participants without a dementia syndrome (controls). In addition to the MR scans, demographic information (age, gender) and information on which data are from the same institute is included. A large set is needed for comparison of the different methods. In addition, a large set increases the scientific value of our framework, as the data better represents a clinical population.

Most of the data is used for evaluation of the methods: 354 MRIs for the test set. Additionally, a small training dataset is provided, which consists of 30 scans distributed over the diagnostic groups was added to our training set.

3 Methods

Statistical classification is an active area of pattern recognition and computer vision research in which scalar- or vector-valued observations are automatically assigned to specific groups, often based on a training set of previously labeled examples. In medical imaging, different types of classification tasks are performed. e.g., classifying image voxels as belonging to a certain anatomical structure, or classifying an individual scanned into one of several diagnostic groups (disease versus normal, semantic dementia versus Alzheimers disease, for example). We use a fully automated method for voxel classification in a brain MRI scan as belonging to the hippocampus versus not. In this method a procedure of training dataset selection based on active learning machine is used during learning of voxel classification according to their Haar-like features, variables that describe complex images based on a statistical analysis of adjacent groups of voxels. The system consists of three processing levels: (a) linear registration of all brains to a standard template and automated method to capture the global shape of the hippocampus. (b) Feature extraction: all voxels included in the previously selected volume were characterized by 315 features computed from local information. (c) Voxel classification: a Random Forests algorithm was used to classify voxels as belonging or not belonging to the hippocampus. The procedure has been detailed in [13, 14, 15, 16].

Hippocampus volume is a feature used in this analysis. For the extraction of the other features we used publicly available brain segmentation package FreeSurfer v.5.1 (FS) (freesurfer.nmr.mgh.harvard.edu) [17], which aim is to process the structural brain MRI scans and compute morphological measurements. The FreeSurfer pipeline is fully automatic and provides 184 features per MRI scan in total. Volumes of cortical and sub-cortical structures such as the caudate and average thickness measurements within cortical regions, such as the precuneus. We use the FS features but for hippocampus volume we use the segmentation proposed in [16].

For the diagnosis classification we passed all the features to a C-Support Vector Classifier (C-SVC) with a linear kernel on a 5-fold cross validation. Specifically, CTRL vs MCI, CTRL vs AD, and MCI vs AD classifications were computed, after that the three probabilities were combined using the formula [20],

$$P(C_i|x) = \frac{\prod\limits_{j \neq i} p(C_i|\phi_{ij})}{\sum\limits_k \prod\limits_{j \neq k} p(C_k|\phi_{kj})}$$
(1)

where ϕ_{ij} represents the output of the ovo classifiers. Since C-SVC does return only predicted class labels, we used the Platt scaling [23] to generate posteriors from the output of binary classifiers.

The five-per-pair binary classifiers were then used to predict the diagnosis of the blind test set, by averaging the posteriors over all the cross validation rounds. Final classification is obtained with a max-wins-all rule on the three class posteriors.

3.1 Computational infrastructure

The method is developed in ITK and MATLAB framework for hippocampus segmentation and we use FS to brain feature extraction. The computational

resources required is about 13 hour per image. Therefore the availability of distributed computing software environments and adequate infrastructures was of fundamental importance.

In this study, the LONI pipeline processing environment [21, 22] was used: a user-friendly and efficient software for complex data analyses, available at http://pipeline.loni.ucla.edu.

The present study was carried out using the local computer farm BC2S ³: a distributed computing infrastructure consisting of about 5000 CPU and allowing up to 1,8 PB storage. A further study for grid deployment was also performed, with the aim of creating a pipeline tool suitable for large clinical trials. It was carried out on the European Grid Infrastructure (EGI) which consists of about 300 geographically distributed sites around the world. In particular all the results presented in this study were obtained on the BC2S using the 484 MR images at our disposal, the run-time reduction with the grid implementation allowed to produce results in a reasonable time with respect to the application execution as a sequential process on limited resources. The advantages of the grid execution are evident since we obtained the 90% of the analysis of 484 images after less than 16 hours.

4 Results and Conclusion

We compute our estimates of the performance metrics (ACC, AUC) for 5-fold cross validation extracting random the test set. Results showed that our method performs very good in discriminating the three classes (CTRL, MCI, and AD), in line with those of literature, with an overall accuracy of 0.8090 on the entire dataset (D_{All}), whereas for the 30 subjects downloaded from MICCAI (D_{30}), we obtained an accuracy of 0.7333. In table 2 and table 3 we report the confusion matrices for both the datasets.

	Pred. CTRL	Pred. MCI	Pred. AD
CTRL	95	9	4
MCI	15	80	10
AD	5	12	58

 Table 2. Confusion matrix for the classification of the entire dataset.

Moreover, in Fig.1, and Fig.2 are reported the ROC curves for the three classes, for both the datasets. Also AUC scores are reported.

³ http://www.recas-pon.ba.infn.it



Fig. 1. Roc curves on the entire dataset.

	Pred. CTRL	Pred. MCI	Pred. AD
CTRL	10	1	1
MCI	3	5	1
AD	0	2	7

Table 3. Confusion matrix for the classification of the dataset D_{30} .



Fig. 2. Roc curves on D_{30} .

The proposed fully automated approach may be suitable for large-scale research studies, in the first instance on Alzheimers disease, where the hippocampal volume and morphological changes are important biomarkers, potentially also on other brain disorders in which atrophy and structural brain changes plays a relevant pathogenetic role.

Acknowledgments

Data used in preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

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: Alzheimers Association: Alzheimers Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. 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 Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

This research was supported by Istituto Nazionale di Fisica Nucleare (INFN), Italy. This research was also supported by grants from Università degli Studi di Bari, Italy.

All authors disclose any actual or potential conflicts of interest, including any financial, personal, or other relationships with other people or organizations that could inappropriately influence their work. All experiments were performed with the informed consent of each participant or caregiver in line with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Local institutional ethics committees approved the study.

References

- T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active Shape Models-Their Training and Application. Computer Vision and Image Understunding Vol. 61, No. 1, January, pp. 38-59 (1995)
- [2] C. R. Jack, M. M. Shiung, J. L. Gunter, P. C. Obrien, S. D. Weigand, D. S. Knopman, B. F. Boeve, R. J. Ivnik, G. E. Smith, R. H. Cha, et al., Comparison of different mri brain atrophy rate measures with clinical disease progression in ad, Neurology 62 (4) (2004) 591600.
- [3] T. F. Cootes, G. J. Edwards, C. J. Taylor, Active appearance models, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (6), 681685, (2001).
- [4] C. Davatzikos, S. M. Resnick, X. Wu, P. Parmpi, C. M. Clark, Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI, NeuroImage 41 (4) 12201227, (2008).
- [5] S. M. Pizer, P. T. Fletcher, S. Joshi, A. Thall, J. Z. Chen, Y. Fridman, D. S. Fritsch, A. G. Gash, J. M. Glotzer, M. R. Jiroutek, C. Lu, K. E. Muller, G. Tracton, P. Yushkevich, E. L. Chaney, Deformable M-Reps for 3D Medical Image Segmentation, International Journal of Computer Vision 55 (2) 85106, (2003).
- [6] M. Chupin, E. Grardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehricy, H. Benali, L. Garnero, O. Colliot, The Alzheimers Disease Neuroimaging Initiative, Fully automatic hippocampus segmentation and classification in Alzheimers disease and mild cognitive impairment applied on data from ADNI, Hippocampus 19 (6), 579587 (2009).
- [7] Z. Tu, K. L. Narr, P. Dollar, I. Dinov, P. M. Thompson, A. W. Toga, Brain anatomical structure segmentation by Hyhrid discriminative /generative models, IEEE Trans. Med. Imaging 27 (4), 495508, (2008).
- [8] Y. Zhang, M. Brady, S. Smith, Segmentation of brain MR images through a hidden Markov random field model and the expectationmaximization algorithm, IEEE Trans. Med. Imaging 20 (1), 4557, (2001).
- [9] 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 (1) (2006) 115126.
- [10] R. Wolz, P. Aljabar, J. V. Hajnal, A. Hammers, D. Rueckert, The Alzheimers Disease Neuroimaging Initiative, LEAP: learning embeddings for atlas propagation, NeuroImage 49 (2) (2010) 13161325.
- [11] D. L. Collins, J. C. Pruessner, Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion, NeuroImage 52 (4) (2010) 13551366.

- [12] M. Boccardi, M. Bocchetta, L. Apostolova, J. Barnes, G. Bartzokis, G. Corbetta, C. DeCarli, L. DeToledo-Morrell, M. Firbank, R. Ganzola, L. Gerritsen, W. Henneman, R. Killiany, N. Malykhin, P. Pasqualetti, J. Pruessner, A. Redolfi, N. Robitaille, H. Soininen, D. Tolomeo, L. Wang, H. Watson, S. Duchesne, C. Jack, G. B. Frisoni, DelphiDelphi Consensus on Landmarks for the Manual Segmentation of the Hippocampus on MRI: Preliminary Results from the EADC-ADNI Harmonized Protocol Working Group, Neurology 78 Suppl. 1, 171174, (2012).
- [13] Amoroso, N., et al.: Automated shape analysis landmarks detection for medical image processing. Computational Modelling of Objects Represented in Images III: Fundamentals, Methods and Applications, 139 (2012).
- [14] Tangaro, S., et al.: Active Learning Machines for Automatic Segmentation of Hippocampus in MRI. ICDM2013 proceedings, (2013).
- [15] Maglietta, R., et al.: Random Forest Classification for Hippocampal Segmentation in 3D MR Images. In Machine Learning and Applications (ICMLA), 2013 12th International Conference on IEEE 1, 264–267 (2013, December).
- [16] Tangaro, S., et al.: Automated voxel-by-voxel tissue classification for hippocampal segmentation: Methods and Validation. Physica Medica, in press.
- [17] Fischl, B.: FreeSurfer. Neuroimage 62, 774-781 (2012).
- [18] Dale, A.M., Fischl, B., Sereno, M.I.: Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage 9, 179-194 (1999).
- [19] Fischl, B., Sereno, M.I., Dale, A.M.,: Cortical surface-based analysis: II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195-207 (1999).
- [20] Hamamura, T.; Mizutani, H.; Irie, B., "A multiclass classification method based on multiple pairwise classifiers," Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on , vol., no., pp.809,813, 3-6 Aug. 2003 doi: 10.1109/ICDAR.2003.1227774
- [21] I. D. Dinov, K. Lozev, P. Petrosyan, Z. Liu, P. Eggert, J. Pierce, A. Zamanyan, S. Chakrapani, J. D. Van Horn, D. S. Parker, R. Magsipoc, K. Leung, B. Gutman, R. P. Woods, A. W. Toga, Neuroimaging Study Designs, Computational Analyses and Data Provenance Using the LONI Pipeline, PLoS ONE 5 (9).
- [22] D. E. Rex, J. Q. Ma, A. W. Toga, The LONI Pipeline Processing Environment, NeuroImage 19 (3) (2003) 10331048.
- [23] John C. Platt, "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods", ADVANCES IN LARGE MARGIN CLASSIFIERS, 1999, 61–74, MIT Press

BrainPrint in the Computer-Aided Diagnosis of Alzheimer's Disease

C. Wachinger^{1,2}, K. Batmanghelich¹, P. Golland¹, M. Reuter^{1,2}

¹Computer Science and Artificial Intelligence Lab, MIT ²Massachusetts General Hospital, Harvard Medical School

Abstract. We investigate the potential of shape information in assisting the computer-aided diagnosis of Alzheimer's disease and its prodromal stage of mild cognitive impairment. We employ BrainPrint to obtain an extensive characterization of the shape of brain structures. BrainPrint captures shape information of an ensemble of cortical and subcortical structures by solving the 2D and 3D Laplace-Beltrami operator on triangular and tetrahedral meshes. From the shape descriptor, we derive features for the classification by computing lateral shape differences and the projection on the principal component. Volume and thickness measurements from FreeSurfer complement the shape features in our model. We use the generalized linear model with a multinomial link function for the classification. Next to manual model selection, we employ the elasticnet regularizer and stepwise model selection with the Akaike information criterion. Training is performed on data provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI) and testing on the data provided by the challenge. The approach runs fully automatically.

1 Introduction

This paper describes our method submitted to the *challenge on Computer-Aided Diagnosis of Dementia based on structural MRI data* held in conjunction with MICCAI 2014 in Boston. The task of the challenge is to differentiate between patients with Alzheimer's disease (AD), mild cognitive impairment (MCI), and healthy controls (CN) given T1-weighted MRI data. Alzheimer's disease leads to structural changes in the brain that are visible in T1 images. The challenge provides multi-center data from hospitals in the Netherlands, including 30 subjects for validation and 354 subjects for testing. The objective of the study is to provide a large-scale objective validation of methods for computer-aided diagnosis of dementia and therefore promoting their application in clinical practice.

We propose to augment the commonly used volume and thickness measurements for AD classification with shape information. Shape information can contribute valuable information to the characterization of brain structures, which is only coarsely represented by its volume. For quantifying shape, we use the recently introduced *BrainPrint* [22], which is a holistic representation of the brain anatomy, containing the shape information of an ensemble of cortical and subcortical structures. Previous studies that employed shape information for the

2 C. Wachinger, K. Batmanghelich, P. Golland, M. Reuter

classification focused on single structures, e.g. the hippocampus [8, 6, 3, 21], while *BrainPrint* provides an extensive characterization of the brain anatomy. Such a holistic characterization seems promising in diagnosing Alzheimer's disease, which is associated with global atrophy across the brain. *BrainPrint* naturally extends the ROI-based analysis in FreeSurfer that provides volume and thickness measurements with shape information.

BrainPrint quantifies the shape information by calculating the spectrum of the Laplace-Beltrami operator on both triangular meshes that represent boundary surfaces, e.g. the white matter surface, and tetrahedral meshes for volumetric representations. The meshes are constructed from the segmentation provided by FreeSurfer. From the detailed, high-dimensional characterization with Brain-Print, we compute lateral shape differences and the projection on the principal component to obtain shape features that can easily be integrated in the classification. These shape features complement volume and thickness measurements in our analysis, which are appropriately normalized to account for head size and gender. The classification is performed with the generalized linear model and a multinomial link function. Next to manual model selection, we employ the elastic-net regularizer and stepwise model selection with the Akaike information criterion. Training is performed on data provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI) and testing on the data provided by the challenge.

1.1 Related Work

A recent review of machine learning approaches in Alzheimer's disease [5], refers to only two articles that employ shape features [6,3], illustrating that shape is not commonly used. A recent comparison of methods for AD classification [4] included one shape-based approach [8]. The focus of these shape-based approaches is the hippocampus. More precisely, [8] approximate the hippocampal shape by a series of spherical harmonics. [6] use permutation tests to extract surface locations that are significantly different among patients with AD and controls. [3] incorporate shape information by deriving thickness measurements of the hippocampus from a medical representation. [21] use statistical shape models to detect hippocampal shape changes. Other structures of interest for shape analysis were the cortex and ventricles. [11] use multi-resolution shape features with non-Euclidean wavelets for the analysis of cortical thickness. [12] analyze the fractal dimension of the cortical ribbon. [10] model surface changes of the ventricles in a longitudinal setup with a medial representation.

The shape characterization in *BrainPrint* [22] builds upon shapeDNA [20], which contains the spectral information of objects. Spectral signatures have previously been investigated for AD classification [1], with a focus on right hippocampus, right thalamus and right putamen. In contrast, we characterize cortical structures and the wide range of subcortical structures. Moreover, a 3D object can be represented by its volume (*e.g.* tetrahedra mesh) or its boundary (*e.g.* triangle meshes), where the spectra of 3D solid objects and their 2D boundary surfaces contain complementary information [20]. Most prior work computes

BrainPrint in the Computer-Aided Diagnosis of Alzheimer's Disease

the shapeDNA for triangular surface meshes [1, 2, 16], while we also work with spectra of tetrahedral volume tessellations [19]. Given that the Laplace spectra are isometry invariant, the 2D boundary representation alone may yield a weaker descriptor, due to the large set of potential (near-) isometric deformations. E.g. a closed 2D surface with a protrusion pointing inwards yields the same descriptor as one with the protrusion pointing outwards, while the spectra of the enclosed 3D solids differ. Furthermore, it has been shown in [17] that spectra of the 2D boundary surface are capable of distinguishing two isospectral 3D solids (GWWprisms). For these resons we combine the information from both the 3D solid and 2D boundary shape representations in the *BrainPrint*.

2 Shape Features for Classification

2.1 Shape Descriptor BrainPrint

The shape description is based on the automated segmentation of anatomical brain structures with FreeSurfer [7]. In this work we use the shapeDNA [20] as shape descriptor, which performed among the best in a comparison of methods for non-rigid 3D shape retrieval [14]. The ShapeDNA is computed from the intrinsic geometry of an object by calculating the Laplace-Beltrami spectrum. It provides a compact shape representation that facilitates the further analysis. Considering the Laplace-Beltrami operator Δ , we obtain the spectrum by solving the Laplacian eigenvalue problem (Helmholtz equation)

$$\Delta f = -\lambda f \tag{1}$$

using the finite element method. The solution consists of eigenvalue $\lambda_i \in \mathbb{R}$ and eigenfunction f_i pairs (sorted via their eigenvalues: $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots$). The first l non-zero eigenvalues form the shapeDNA: $\lambda = (\lambda_1, \ldots, \lambda_l)$. Next to this direct definition of shapeDNA, we also consider the normalization of the eigenvalues to make the representation independent of the objects' scale

$$\lambda' = \operatorname{vol}^{\frac{2}{D}}\lambda,\tag{2}$$

where vol is the Riemannian volume of the D-dimensional manifold (i.e. the area for 2D surfaces) [20].

The eigenvalues are isometry invariant with respect to the Riemannian manifold, meaning that length-preserving deformations will not change the spectrum. Isometry invariance includes rigid body motion and therefore permits the comparison of subjects by directly comparing the shapeDNA, without the need for alignment. A second property is that the spectrum continuously changes with topology-preserving deformations of the boundary of the object. Fig. 1 illustrates the eigenfunctions calculated on the surface of the cerebral cortex. The eigenfunctions show natural vibrations of the shape when oscillating at a frequency specified by the square root of the eigenvalue.

We compute the spectra for all cortical and subcortical structures on the 2D boundary surfaces (triangle meshes) and additionally for cortical structures

C. Wachinger, K. Batmanghelich, P. Golland, M. Reuter

4



Fig. 1: Left cerebral cortex and first eigenfunctions of the Laplace-Beltrami operator calculated on the surface (yellow positive, red negative, and green zero).

(white and pial surfaces in both hemispheres) also on the full 3D solid (tetrahedra meshes), forming the *BrainPrint* $\Lambda = (\lambda_1, \ldots, \lambda_\eta)$. Triangle meshes of the cortical surfaces were obtained automatically for each hemisphere using FreeSurfer. Surface meshes of subcortical structures were constructed via marching cubes from the FreeSurfer subcortical segmentation. To construct tetrahedral meshes, first handles were removed from the surface meshes, which were then uniformly resampled to 60K vertices. The gmsh package [9] was used to compute the tetrahedra volume meshes. For the computation of the spectra we used the linear finite element method [20] with Neumann boundary condition (zero normal derivative) for tetrahedra meshes.

2.2 Shape Features from BrainPrint

BrainPrint contains shape information from $\eta = 44$ brain structures, each one described by l = 50 eigenvalues. This results in a characterization of a subject's brain shape by over 2000 variables, which can easily cause overfitting of the model. To decrease the number of variables and increase robustness, we therefore (i) include a 1D asymmetry measure (the distance of BrainPrint across hemispheres), and (ii) employ principal component analysis. When working with BrainPrint, the potential switching of eigenfunctions across shape deformations as discussed in [18] are not problematic as we compare the sorted sequences of eigenvalues (shapeDNA). In order for eigenfunctions to switch, their eigenvalues have to be very close initially, so a switch has only a limited effect on the BrainPrint. Another aspect of shapeDNA is that the eigenvalues form an increasing sequence with the variance increasing as well. Depending on the distance measure, this can cause higher eigenvalues to dominate the similarity measure between two shapes, although these components do not necessarily contain the most important geometric information. Additionally, the increased impact of higher eigenvalues puts a particular emphasis on number of selected eigenvalues l. To account for these issues we normalize the *BrainPrint* and employ appropriate distance computations as described next.

Lateral differences: As a first shape feature, we compute lateral (left/right) shape differences of brain structures. More precisely, we compute differences for white and gray matter surfaces with triangular and tetrahedral meshes, as well

BrainPrint in the Computer-Aided Diagnosis of Alzheimer's Disease

as triangular meshes for cerebellum white matter and gray matter, striatum, lateral ventricles, hippocampus and amygdala. We calculate the Mahalanobis distance between a brain structure s

$$d_s = \|\boldsymbol{\lambda}_s^{\text{left}} - \boldsymbol{\lambda}_s^{\text{right}}\|_{\Sigma_s},\tag{3}$$

5

with Σ_s the covariance matrix across all subjects for structure s.

Alternative distance functions that have been proposed for shapeDNA are the Euclidean distance (or any p-norm), some Hausdorff distances, the Euclidean distance on re-weighted eigenvalues $\hat{\lambda}_i = \lambda_i/i$ [20, 17], and the weighted spectral distance [13]. The weighted distances (latter two approaches) are motivated by the need to reduce the impact of higher eigenvalues on the distance. The linear re-weighting is based upon the observation that the eigenvalues demonstrate a linear growth pattern (Weyl's law) and therefore yields an approximately equal contribution of each eigenvalue. The weighted spectral distance is similar to a division by the squared eigenvalue number and therefore functions like a low-pass filter. The proposed Mahalanobis distance accounts for the covariance pattern in the data and supports an equal contribution of all eigenvalues in the sequence.

Principal Component Analysis: A second set of features that we derive from *BrainPrint* is the calculation of the principal component analysis for each of the 44 brain structures. Projecting the shapeDNA on the principal component retains most of the variance in the dataset, while reducing the dimensionality. Problematic in this regard is once again that higher eigenvalues show most variance, so that they will dominate the identification of the principal component. We have experimented with (i) linear re-weighting, $\hat{\lambda}_i = \lambda_i/i$, and (ii) the normalization of each eigenvalue to unit variance across the dataset. Evaluation of both approaches yielded similar results, so that we employ the simpler linear re-weighting.

3 Alzheimer Classification with BrainPrint

We use the generalized linear model (GLM) for the classification with a multinomial link function. We employ linear models in this study because we expect non-linear methods in combination with the large number of features to be more susceptible to overfitting.

3.1 Features

Each subject is characterized by 163 features. These split up in 70 thickness and 39 volume measurements provided by FreeSurfer, together with 10 lateral shape differences, and 44 PCA shape variations. According to the recent analysis of the normalization of variables for AD classification [23], we normalize volumetric measures by ICV but do not normalize cortical thickness measures. Linear regression with respect to age is performed for each feature to remove the confounding effect of age in the analysis. After the normalization, we exclude age and gender from the variables.

C. Wachinger, K. Batmanghelich, P. Golland, M. Reuter



Fig. 2: Multinomial deviance of elastic-net computed with cross-validation for different parameters λ (bottom) and the corresponding number of features (top). The plot shows the mean deviance together with upper and lower standard deviation.

3.2 Model Selection

6

We use three different approaches for model selection: manual, stepwise selection, and elastic-net. For the manual model, we choose volume and thickness measurements that have previously been reported important for AD classification and add shape features. For volume, we use hippocampus and amygdala. For thickness, we use entorhinal cortex, middle temporal lobe, parahippocampal gyrus, and banks of the superior temporal sulcus. For shape, we use the lateral differences of hippocampus, amygdala, and lateral ventricles. Experiments show a clear improvement in classification accuracy with the additional shape features.

As second approach, we employ stepwise model selection in the GLM with the Akaike information criterion (AIC). The AIC considers model complexity in terms of the number of variables and the likelihood of the data given the data. The model with the best trade-off between high likelihood and few variables is selected by the AIC. We performed the model selection once on the ADNI training data, yielding a model with 25 thickness, 15 volume, and 13 shape features, referred to as "stepwise I". In a second run, we added the challenge training data to the model selection, leading to a model with 30 thickness, 11 volume, and 19 shape features, referred to as "stepwise II".

In the final approach for model selection, we use elastic-net regularized general linear models provided by the R package glmnet. The regularizer in elasticnet combines lasso and ridge-regression penalties, modulated by the parameter α

$$P_{\alpha}(\beta) = (1 - \alpha)\frac{1}{2} \|\beta\|_{2}^{2} + \alpha \|\beta\|_{1}, \qquad (4)$$

with coefficients β . Our experiments indicated best results for the lasso penalty ($\alpha = 1$). An additional parameter λ , balancing the data fit and penalty term, was determined with cross-validation, see Fig. 2. The lowest multinomial deviance of the model corresponds roughly to $\lambda = \exp(-4.5)$.

BrainPr	rint	in 1	$_{\mathrm{the}}$	C	omputer-A	Aic	led	I	Diagnosi	s of	İΑ	lz	heimer	$^{\rm o}$	D	isea	se
---------	------	------	-------------------	---	-----------	-----	-----	---	----------	------	----	----	--------	------------	---	------	----

	Subjects	Diagnosis (CN/MCI/AD)	Gen Male	ider Female	Age quantiles $(1st/2nd/3rd)$
ADNI	751	(213/364/174)	437(58%)	314(42%)	(71.1/75.3/79.8)
Challenge-Validation	30	(12/9/9)	17(57%)	13(43%)	(59.3/65.0/68.0)
Challenge-Test	354		213(60%)	141(40%)	(59.0/64.0/71.0)

Table 1: Statistics of datasets used for training and testing.

4 Results

We processed the non-uniformity corrected brain scans distributed by the challenge with FreeSurfer v5.3. We processed the baseline scans from ADNI-1 [15] with FreeSurfer v5.1. Table 1 lists statistics about diagnosis and demographics of both datasets. Notable is the age difference between the challenge and ADNI data. The first quantile of ADNI is higher than the third quantile of the challenge data meaning that 75% of ADNI cases are older than 75% of challenge cases. This mismatch may have a detrimental effect on the classification. All results reported in this paper are obtained by training on the ADNI data and predicting on the challenge-validation data. The results on the challenge-test data will be announced by organizers later. For the final training with the submission of the results to the challenge, we also include the validation data to the training set because of the potential benefit of including samples from the target dataset to training.

The processing of the challenge data with FreeSurfer failed for 3 subjects. These subjects are labeled as AD without further inspection, since we assume a relation between processing difficulties and severe atrophy or increased subject motion in AD patients. The *BrainPrint* calculation caused errors on triangular meshes for 1 subject and on tetrahedral meshes on 5 subjects. The errors are related to FreeSurfer producing a mesh that is a non-manifold (the mesh has edges sharing more than two triangles) and problems in the volumetric mesh creation with gmsh. We impute missing values for these 6 subjects by the population mean in the statistical analysis. The mean runtime of FreeSurfer was 16.8h on the challenge data. The calculation time for *BrainPrint* is about 0.6h per subject, where most time is spent on mesh processing. The statistical analysis runs in less than a minute. All processing steps run fully automatically.

Our approach depends on a number of design choices and variables. One is the distance function for the calculation of lateral shape differences. The results across the model selection schemes for different distance functions are not unanimous but the Mahalanobis distance achieved competitive results. We choose a linear re-weighting of the eigenvalues in PCA. Another important parameter is the number of eigenvalues l forming the shapeDNA. Fig. 3 illustrates the classification accuracy on the challenge-validation data as a function of the number of eigenvalues. We compare the manual model selection, the two stepwise approaches, and the elastic-net approach. The plot also shows the impact of

135

C. Wachinger, K. Batmanghelich, P. Golland, M. Reuter



Fig. 3: Classification accuracy on challenge-validation data. The plots contrast the impact of the volume normalization in Eq.(2). Results are shown for different model selection schemes and are plotted as function of the number eigenvalues l.

the volume normalization of shapeDNA with Eq.(2) on the classification performance.

Based on the results, we select l = 40 eigenvalues for the further analysis. Table 2 shows additional evaluation measures for this case. The scripts to calculate these measures were provided by the challenge organizers. For the final prediction on the challenge-test data, we also added the challenge-validation data to the training set. We measure the consistency of the predicted classes on the challenge-test data for the changed training set and list it in the table. We observe a tendency for less consistent results with more sophisticated methods. Since the number of results submitted per group was limited to 5, we select the methods with best performances, highlighted in bold face in table 2. We also included the manual model, although it performs slightly worse than the elastic-net.

5 Conclusion

8

We investigated the potential of *BrainPrint* for classifying between AD, MCI, and CN in the MICCAI dementia challenge. We derived shape features from *BrainPrint* and used them together with volume and thickness measures in generalized linear models for classification. Three different approaches for model selection were presented. We trained our methods on the ADNI dataset. Our results indicate the good performance in classification when adding the ensemble of shape information from *BrainPrint*.

Acknowledgements: This work was supported in part by the Humboldt foundation, the Martinos Center for Biomedical Imaging (P41-RR014075, P41-EB015896), the National Alliance for Medical Image Computing (U54-EB005149)

BrainPrint in the Computer-Aided Diagnosis of Alzheimer's Disease

Model	Norm	Accuracy	TPF-CN	TPF-MCI	TPF-AD	AUC	CON
Manual		67 (43-80)	83 (50-100)	67 (20-89)	44 (10-80)	78 (63-90)	97
Stepwise I	No	77 (53-87)	83 (50-100)	67(20-89)	78 (33-100)	88 (73-96)	95
Stepwise II	. 110	80 (60-90)	92(55-100)	56(20-86)	89 (50-100)	84(69-95)	91
Elastic-net		73 (53-83)	83 (50-100)	56 (20-86)	78 (33-100)	84 (69-93)	92
Manual		63 (40-77)	75 (42-93)	67 (20-89)	44 (10-80)	79 (63-91)	97
Stepwise I	Vac	77 (53-87)	83 (50-100)	78 (33-100)	67(17-89)	89 (76-96)	98
Stepwise II	res	63(40-77)	92 (55-100)	33 (0-70)	56(11-83)	77 (65-88)	91
Elastic-net		73 (53-83)	83 (50-100)	44 (13-78)	89 (50-100)	86 (73-95)	93

Table 2: Evaluation of classification results for different models calculated on the challenge-validation data. Confidence intervals are shown in parenthesis. All values are in %. (Norm = volume normalization, TPF = true positive fraction, AUC = area under the curve, CON = consistency of prediction). Methods submitted to the challenge are highlighted with bold face.

and the NeuroImaging Analysis Center (P41-EB015902). We thank the Alzheimer's Disease Neuroimaging Initiative (ADNI) for image data.

References

- Bates, J., Pafundi, D., Kanel, P., Liu, X., Mio, W.: Spectral signatures of point clouds and applications to detection of alzheimer's disease through neuroimaging. In: IEEE International Symposium on Biomedical Imaging. pp. 1851–1854 (2011)
- Bernardis, E., Konukoglu, E., Ou, Y., Metaxas, D., Desjardins, B., Pohl, K.: Temporal shape analysis via the spectral signature. In: Medical Image Computing and Computer-Assisted Intervention, pp. 49–56 (2012)
- Costafreda, S.G., Dinov, I.D., Tu, Z., Shi, Y., Liu, C.Y., Kloszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Vellas, B., et al.: Automated hippocampal shape analysis predicts the onset of dementia in mild cognitive impairment. Neuroimage 56(1), 212–219 (2011)
- Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O.: Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the adni database. Neuroimage 56(2), 766–781 (2011)
- 5. Falahati, F., Westman, E., Simmons, A.: Multivariate data analysis and machine learning in Alzheimer's disease with a focus on structural magnetic resonance imaging. Journal of Alzheimer's Disease (2014)
- Ferrarini, L., Frisoni, G.B., Pievani, M., Reiber, J.H., Ganzola, R., Milles, J.: Morphological hippocampal markers for automated detection of alzheimer's disease and mild cognitive impairment converters in magnetic resonance images. Journal of Alzheimer's Disease 17(3), 643–659 (2009)
- Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., et al.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33(3), 341–355 (2002)

- 10 C. Wachinger, K. Batmanghelich, P. Golland, M. Reuter
- Gerardin, E., Chételat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H.S., Niethammer, M., Dubois, B., Lehéricy, S., Garnero, L., et al.: Multidimensional classification of hippocampal shape features discriminates alzheimer's disease and mild cognitive impairment from normal aging. Neuroimage 47(4), 1476–1486 (2009)
- Geuzaine, C., Remacle, J.F.: Gmsh: A 3-d finite element mesh generator with builtin pre-and post-processing facilities. International Journal for Numerical Methods in Engineering 79(11), 1309–1331 (2009)
- Gutman, B.A., Hua, X., Rajagopalan, P., Chou, Y.Y., Wang, Y., Yanovsky, I., Toga, A.W., Jack Jr, C.R., Weiner, M.W., Thompson, P.M.: Maximizing power to track Alzheimer's disease and MCI progression by LDA-based weighting of longitudinal ventricular surface features. Neuroimage 70, 386–401 (2013)
- Kim, W.H., Singh, V., Chung, M.K., Hinrichs, C., Pachauri, D., Okonkwo, O.C., Johnson, S.C.: Multi-resolutional shape features via non-Euclidean wavelets: Applications to statistical analysis of cortical thickness. NeuroImage 93, 107 – 123 (2014)
- King, R.D., Brown, B., Hwang, M., Jeon, T., George, A.T.: Fractal dimension analysis of the cortical ribbon in mild Alzheimer's disease. Neuroimage 53(2), 471– 479 (2010)
- Konukoglu, E., Glocker, B., Criminisi, A., Pohl, K.M.: WESD–Weighted spectral distance for measuring shape dissimilarity. Pattern Analysis and Machine Intelligence, IEEE Transactions on 35(9), 2284–2297 (2013)
- Lian, Z., Godil, A., Bustos, B., Daoudi, M., Hermans, J., Kawamura, S., Kurita, Y., Lavoué, G., Van Nguyen, H., Ohbuchi, R., et al.: A comparison of methods for non-rigid 3D shape retrieval. Pattern Recognition 46, 449–461 (2012)
- Mueller, S.G., Weiner, M.W., Thal, L.J., et al: The alzheimer's disease neuroimaging initiative. Neuroimaging Clinics of North America 15(4), 869–877 (2005)
- Niethammer, M., Reuter, M., Wolter, F.E., Bouix, S., Peinecke, N., Koo, M.S., Shenton, M.: Global medical shape analysis using the Laplace-Beltrami spectrum. In: Ayache, N., Ourselin, S., Maeder, A.J. (eds.) MICCAI 2007. LNCS, vol. 4791, pp. 850–857. Springer, Heidelberg (2007)
- 17. Reuter, M.: Laplace Spectra for Shape Recognition. Books on Demand GmbH (2006)
- Reuter, M.: Hierarchical shape segmentation and registration via topological features of Laplace-Beltrami eigenfunctions. International Journal of Computer Vision 89(2), 287–308 (2010)
- Reuter, M., Niethammer, M., Wolter, F.E., Bouix, S., Shenton, M.: Global medical shape analysis using the volumetric Laplace spectrum. In: International Conference on Cyberworlds, NASA-GEM Workshop. pp. 417–426 (2007)
- Reuter, M., Wolter, F.E., Peinecke, N.: Laplace-Beltrami spectra as "Shape-DNA" of surfaces and solids. Computer-Aided Design 38(4), 342–366 (2006)
- Shen, K.k., Fripp, J., Mériaudeau, F., Chételat, G., Salvado, O., Bourgeat, P.: Detecting global and local hippocampal shape changes in Alzheimer" disease using statistical shape models. Neuroimage 59(3), 2155–2166 (2012)
- 22. Wachinger, C., Golland, P., Reuter, M.: BrainPrint: Identifying subjects by their brain. In: Medical Image Computing and Computer Assisted Interventions (2014)
- Westman, E., Aguilar, C., Muehlboeck, J.S., Simmons, A.: Regional magnetic resonance imaging measures for multivariate analysis in Alzheimer's disease and mild cognitive impairment. Brain topography 26(1), 9–23 (2013)