Application of Machine Learning to Arterial Spin Labeling in Mild Cognitive Impairment and Alzheimer's Disease

This manuscript is an **original research paper**   Word count: 3015

Advances in Knowledge:

- Automated machine learning methods can be trained to distinguish between subjects with subjective cognitive decline (SCD), patients with mild cognitive impairment (MCI), and patients with Alzheimer's disease (AD) based on arterial spin labeling (ASL) images with high classification training accuracy (range, 83.8% - 89.0% , $p <$ .01).

- Classifiers based on these trainings can predict the diagnosis of single subjects with high diagnostic accuracy (area under the receiver operating curve range, .89 - .96, $p < .001$).

Implications for Patient Care:

- Automated classification of 3D pseudo-continuous ASL scans that detects AD patients with high accuracy (> 82%) may support image-based diagnosis, especially in centres without experienced (neuro)radiologists.

- Automated classification of 3D pseudo-continuous ASL scans may be used for AD screening purposes without compromising diagnostic accuracy.

Summary statement: Automated classification of perfusion maps enable distinguishing patients with various stages of Alzheimer's disease with high accuracy.

ABSTRACT

Purpose: This current study investigates whether multivariate pattern recognition analysis of arterial spin labeling (ASL) perfusion maps can be used for classification and single-subject prediction of patients with

Alzheimer's disease (AD), mild cognitive impairment (MCI), and subjective cognitive decline (SCD), after using the W-score method to remove confounding effects of gender and age.

Materials and Methods: The local institutional review board approved the study. Subjects provided written informed consent. 3.0-T pseudo-continuous ASL images were acquired from 100 probable AD patients, 60 MCI patients, of which 12 remained stable (MCIs), 12 converted to AD (MCIc), and 36 without follow-up, 100 SCD subjects, and 26 healthy controls. The three main groups (i.e. AD, MCI,

SCD) were divided into a gender and age-matched training-set ($N = 130$) and independent prediction-set ($N = 130$). Standardized perfusion scores adjusted for age and gender (W-scores) were computed per voxel for each subject. Training of a Support Vector Machine (SVM) classifier used diagnostic status and perfusion maps. Discrimination maps were extracted and used for single-subject classification in the prediction-set. Prediction performance was assessed by means of a ROC analysis, generating an area under the curve (AUC) and sensitivity/specificity distribution.

Results: Single-subject diagnosis in the prediction-set using the discrimination maps yielded excellent performance for AD vs. SCD (AUC .96, $p < .01$), good performance for AD vs. MCI (AUC $= 0.89$, $p < .01$), and poor performance for MCI vs. SCD (AUC $= 0.63$, $p = .06$). Application of the AD vs. SCD discrimination map for prediction of MCI subgroups resulted in

good performance for MCIc vs. SCD (AUC = .84, $p < .01$) and fair performance for MCIc vs. MCIs (AUC = .71, $p > .05$).

Conclusion: Using automated methods, age- and gender adjusted ASL perfusion maps can be used to classify and predict diagnoses of AD, MCI-converters, stable MCI patients and SCD subjects with good to excellent accuracy and AUC values.

1. INTRODUCTION

Alzheimer's disease (AD) is the most common type of dementia [1] and the fifth leading cause of death in people aged 65 years and older [2]. Structural magnetic resonance imaging (sMRI), provides high accuracy when diagnosing patients with AD vs. controls, especially in advanced stages of the disease [3,4]. However, early diagnostic accuracy for identifying AD and prognostic value for patients diagnosed with early AD using sMRI remains problematic [5].

In the dynamic biomarker model from Jack and colleagues (2010), functional AD-related brain changes occur before structural changes [6]. Whole-brain hypo-perfusion as measured by dedicated MR scans has been demonstrated in patients with AD compared to controls [7,8], most prominently in the parietal lobe structures such as the posterior cingulate cortex (PCC), precuneus (PC), and inferior parietal lobule (IPL) [7-9]. Patients with mild cognitive impairment (MCI) show similar but less pronounced hypo-perfusion patterns compared to patients with AD [10], with quantitative cerebral blood flow (CBF) values intermediate between those of patients with AD and controls [11]. Hypo-perfusion has also been reported in the occipital and temporal lobes in patients with MCI compared to controls [11].

Arterial spin labeling (ASL) MRI is a non-invasive, rapid and increasingly widely available method for quantifying CBF; ASL represents a potential alternative modality for measuring brain perfusion as compared to positron emission tomography (PET) [12-14] that may facilitate routine clinical application in the work-up of dementia. AD-associated perfusion changes measured by ASL are strongly correlated with glucose metabolism alterations as

measured by PET [9,15,16]. This constellation of findings suggest that ASL is a promising alternative functional biomarker for the early diagnosis of AD.

In addition to the perfusion-related diagnostic parameters mentioned above, the application of multivariate pattern recognition software to sMRI data has yielded high diagnostic accuracy in AD [17]. Compared to visual assessment of hippocampal volumes [18,19], automatic classification has resulted in high accuracy in predicting MCI conversion [20], suggesting that it includes more features of the neurodegenerative pathology [18].

Support vector machines (SVM) represent binary machine learning multivariate methods that can be trained to classify individual scans in a leave-one-out cross-validation framework [17]. Advantages of this multivariate method over an univariate method include increased statistical power, single-subject examination applicability, with the ability to process large amounts of dependent voxel data that more accurately resembling global brain functioning [21]. Specifically in the setting of SVM, the *W*-score method is a statistical tool to reduce the effect of confounds in a binary test [22,23]. As automated image-based classifiers have not been used diagnostically in the setting of ASL, developing such a tool for single-subject classification is clinically relevant and instrumental for screening purposes.

This current study investigates whether multivariate pattern recognition analysis of ASL perfusion maps can be used for classification and single-subject prediction of patients with Alzheimer's disease (AD), mild cognitive impairment (MCI), and subjective cognitive decline (SCD), after using the *W*-score method to remove confounding effects of gender and age.

2. MATERIALS AND METHODS

2.1 *Participants*

The local institutional review board approved this study. All subjects provided written informed consent. This study retrospectively included 311 participants from the Alzheimer Center of the VU University Medical Center Dementia Cohort who underwent an ASL MRI between October 2010 and November 2012 [24]. Exclusion criteria were space occupying processes ($N = 7$), posttraumatic deviations ($N = 6$), large vessel hemorrhages or infarcts ($N = 4$), indications for epilepsy ($N = 4$) or psychiatric disorder(s) ($N = 5$), a-typical clinical representation of AD ($N = 3$) (P.S., 27 years of experience) or failed brain extraction (BET) or ASL acquisition ($N = 22$). Clinical diagnosis was established by consensus in a multidisciplinary team based on a standard dementia screening that included medical history, physical and neurological examinations, screening laboratory tests, neuropsychological testing and brain MRI. Cerebral spinal fluid (CSF) was obtained when possible. Patients with AD met the National Institute on Aging-Alzheimer's Association (NIA-AA) criteria for probable AD [25]. Prior to 2012, MCI diagnosis was based on the criteria defined by Petersen and colleagues [26]; subsequent to 2012, MCI diagnosis was based on the NIA-AA criteria [25]. Patients were considered as SCD subjects if they did not meet AD or MCI criteria. Healthy controls (HC) had normal clinical outcomes and no complaints of cognitive decline [25]. These criteria resulted in the inclusion of a 100 patients with probable AD, 60 patients with MCI, 100 subjects with SCD, and 26 HCs. Level of education was rated on a seven-point scale [27].

Participants within a diagnostic group were randomly assigned to either the training- or prediction-set (each set, N = 130; 50 patients with AD, 30 patients with MCI, 50 subjects with SCD), with balanced distribution of age and sex (Figure 1; Table 1).

Twenty-four of the 60 MCI patients had a follow-up diagnosis 1-4 years (M = 1.96, SD = .75) after the initial diagnosis: 12 had remained stable (MCIs) and 12 had converted to AD (MCIc). Because of the small number of patients, it was not possible to make separate training and prediction sets. Both MCI subgroups were matched to 12 subjects with SCD (Table 2).

2.2 Data Acquisition

Imaging data were collected on a 3.0-T whole-body MR system (Signa HDxt; GE Medical Systems, Milwaukee, USA) using an 8-channel head coil. Structural imaging used a sagittal 3D T1-weighted sequence (IR-FSPGR, repetition time 7.8 ms, echo time 3.0 ms, inversion time 450 ms, flip angle 12°, voxel size 1×0.9×0.9 mm). Pseudo-continuous ASL (PC-ASL) perfusion images (3D fast spin-echo acquisition with background suppression, labelling time 1.5s, postlabel delay 2.0 seconds, echo time 9ms, repetition time 4.8s, spiral readout of eight arms × 512 samples; 36×5.0-mm axial sections, 3.2×3.2-mm in-plane resolution, reconstructed pixel size of 1.7×1.7 mm, acquisition time of 4 minutes) were computed using a single-compartment model [28] after subtraction of labelled images from control images. An approximately proton-densityweighted image was obtained to scale the perfusion image for each subject, using a saturation recovery acquisition with identical parameters.

2.3 Preprocessing of MR imaging data

T1-weighted and PC-ASL images were corrected for gradient non-linearities in three directions. Further analyses were performed in FSL [29]. Preprocessing of T1-weighted images included removal of non-brain tissue, normalisation to MNI space and tissue segmentation with partial volume estimation. ASL images were linearly registered to the grey matter density maps and mapped to MNI standard space, followed by Gaussian smoothing with 6mm FWHM and resampling at 3mm isotropic (A.W.M., 15 years of experience).

2.4 *W-score maps*

Because women show a higher CBF than men [7] and CBF gradually decreases with age [30], these confounds were removed prior to binary classification using the *W*-score method [22,23], using the script github.com/amwink/bias/blob/master/scripts/bash/compute_w.sh. It computes voxel-wise effects of each confound on brain perfusion in a general linear model (GLM) analysis of the reference ASL images (Figure 2, top). The voxel-wise intercept (β_0), gender- and agerelated regression coefficients ($\beta1$ and $\beta2$) and residuals (ε) are used to compute a confoundcorrected, normalised statistic as: [(measured perfusion) – (predicted perfusion)] / (standard deviation of residuals). The measured perfusion is the preprocessed ASL intensity and the predicted perfusion is the sum of voxel-wise effects weighted by the subject's age and gender parameters, respectively (Figure 2, bottom). Like *Z*-scores, negative *W*-scores indicate lower perfusion and positive scores indicate higher perfusion than expected from the reference, given the subject's age and gender.

2.5 *SVM: Multivariate Pattern Recognition in Training-Set*

The pattern recognition for neuroimaging toolbox (PRoNTo) [21], implemented in MATLAB (Mathworks, Natick, USA), provides multivariate pattern analyses for neuro-images. A linear support vector machine (SVM) produces a multi-dimensional hyperplane that optimally separates data in labelled groups (supervised learning). For 2D vectors this would be a straight line [31], but in our case the hyperplane has the dimensionality of the number of included voxels. Discrimination maps, representing the normal to this hyperplane [21], store the relative weight of each voxel to the classification.

The SVM was trained in a leave-one-out cross-validation framework to discriminate between patients with AD, patients with MCI, and subjects with SCD. We assessed the classifier's diagnostic value by comparing the *W*-score maps of both patient groups to the SCD subjects. The classifiers sensitivity to disease progression was assessed by differentiating between the *W*-score maps of patients with AD and MCI. Finally, an exploratory classification training using the *W*score maps of patients with MCIc and MCIs was done to investigate whether the classifier showed prognostic value.

Classification accuracy reflects the predictive power of the algorithm and is therefore of direct diagnostic relevance. Thus, classifier performance was assessed by computing the accuracy, sensitivity, specificity, and a receiver operating curve (ROC), from which the area under the curve (AUC) was calculated. Permutation testing was used to derive a *p*-value of the accuracy (100 permutations) [21].

Training accuracies were first computed based on the whole-brain. Subsequently, masks of Alzheimer-specific regions of interest (ROI) were used to maximize training accuracies. ROIs

were based on the literature [1,7,8,10,11,16] and thresholded group mean perfusion maps. They included the parietal lobe, hippocampus, occipital lobe, and combinations thereof. Masks were created using the MNI structural and Harvard-Oxford subcortical structural atlases [29] (L.E.C., F.H. and A.W.M., 1,1 and 15 years of experience, respectively).

2.6 *SVM: Prediction in new Subjects*

Replication of results in an independent sample supports both internal validity and the generalisability of the classifier [32]. Discrimination maps were used to predict the labels for the prediction-set perfusion maps. Discrimination maps display the discriminative power of voxels in the predefined ROIs, but should not be interpreted as statistical tests. Instead, they provide a spatial representation of the decision boundary normal, i.e., the weight of each voxel in discriminating between groups. Positive values (red and yellow) indicate voxels with a predictive value for the more severe condition and negative values (dark and light blue) indicate voxels with a predictive value for the less severe condition. Single-subject perfusion maps were multiplied by the discrimination map, adjusted by the training bias. The individual integral product scores defined the class, which could be predicted by a simple threshold using ROC analysis (L.E.C., F.H. and A.W.M.).

2.7 *Statistical Analysis*

Statistical analyses of the prediction outcomes were performed in SPSS (version 20; SPSS, Chicago, I11, USA). Specificity, sensitivity, and AUC of the predictions were computed by a ROC analysis. Differences in continuous measures between groups were assessed using one-

way analyses of variance (ANOVA), with post-hoc Bonferroni correction for multiple

comparisons. A χ^2-test was conducted to assess frequency distributions of gender (L.E.C.).

3. RESULTS

3.1 *Participant Characteristics*

No significant gender ($p = 0.74$) or age ($p = 0.68$) differences were observed between the 6

training- and prediction-sets. Within diagnostic groups, no significant differences in MMSE

score were observed between the training- and prediction-sets ($p = 1.00$) (Table 1).

No significant gender (p = 1.00) and age (p = 0.96) differences were found between the

MCI subgroups and 12 matched SCD subjects. SCD subjects had a higher MMSE score than

both MCI subgroups ($p < 0.01$), while the MCI subgroups did not differ in MMSE score ($p =$

0.47) (Table 2).

3.2 *Training of the classifiers*

An overview of training results are shown in Table 3. For AD vs. SCD, training accuracy using

the whole-brain ASL *W*-score map was 87.0% (84.0% sensitivity and 90.0% specificity), with

an

AUC of 0.94 ($p = .01$). A further improvement to 89.0% (84.0% sensitivity, 94.0% specificity,

AUC = 0.93, $p = .01$) was obtained when training was restricted to the parietal lobe and

hippocampus ROI. AD vs. MCI whole-brain analysis resulted in a training accuracy of 78.8%

(84.0% sensitivity and 70.0% specificity), with an AUC of 0.84 ($p = .01$). Restricting the

analysis to the parietal and occipital lobe ROI improved the accuracy to 83.8% (83.3%

sensitivity, 84.0% specificity, AUC = 0.88, $p = .01$). Finally, MCI vs. SCD whole-brain

analysis produced an accuracy of 57.5% (40.0% sensitivity and 68.0% specificity), with an

AUC of 0.49 (p = .42), which was not improved in ROI-based analysis. Resulting discrimination maps from the trainings with the highest accuracies are shown in Figure 3.

3.3 *Predictions: assessment of generalisability*

Results are summarized in Table 4. The use of discrimination weights in patients with AD vs. subjects with SCD enabled correct prediction in 90.0% of individuals (94.0% sensitivity and 86.0% specificity). The ROC curve revealed excellent performance (AUC = 0.96, [95% CI = .92-1], p < .001). The use of discrimination weights in patients with AD vs. patients with MCI enabled correct prediction in 82.0% of individuals (84.0% sensitivity and 80.0% specificity). The ROC curve revealed good performance (AUC = 0.89, [95% CI = .81-.97], p < .001). The use of discrimination weights in patients with MCI vs. subjects with SCD enabled correct prediction in only 60.0% of individuals (60.0% sensitivity and 60.0% specificity). The ROC curve revealed poor performance (AUC = 0.63, [95% CI = .50-.76], p = .06) (Figure 4).

3.4 *Exploratory analyses: classifying MCI subgroups*

For MCIc vs. SCD, whole-brain training produced an accuracy of 83.8% (66.7% sensitivity and 100% specificity), with an AUC of 0.90 (p = .01). A further improvement to 87.5% (75.0% sensitivity, 100% specificity, AUC = 0.92, p = .01) was obtained when training was restricted to the PCC and hippocampus ROI. For MCIc vs. MCIs, whole-brain training produced an accuracy of 70.8% (66.7% sensitivity and 75.0% specificity), with an AUC of 0.77 (p = .05). A further improvement to 83.3% (83.3% sensitivity, 83.3% specificity, AUC = 0.77, p = .01) was obtained when training was restricted to the hippocampus ROI (Table 3). The resulting discrimination maps are shown in Figure 5.

Due to the small cohort, no matching data was available to create an independent MCIconverted or MCI-stable prediction-set. However, use of the AD vs. SCD training discrimination weights for MCIc vs. SCD resulted in correct prediction in 79.0% (83.0% sensitivity, 75.0% specificity) of individuals. The ROC curve revealed good performance (AUC = .84, [95% CI = .68-1], $p < .01$). The use of the same discrimination weights in MCIc vs. MCIs resulted in correct prediction in 71.0% (67.0% sensitivity and 75.0% specificity) of individuals. The ROC curve revealed fair performance (AUC = .71, [95% CI = .49-.93], $p = .08$) (Table 4 and Figure 6).

4. DISCUSSION

In this current study, automated classification of perfusion maps enable distinguishing patients with various stages of Alzheimer's disease with high accuracy. Additionally, the discrimination weights can be used for single-subject diagnostic prediction in an independent data-set, with good to excellent accuracy and AUC values.

Our training and prediction accuracies were similar compared to traditional assessment strategies used by radiologists [18,19] and previous studies applying pattern recognition software to wholebrain sMRI data [17,18,20,35]. Moreover, the classifier presented in this paper performed with good sensitivity and specificity. This suggests that perfusion differences are relevant to the diagnosis of AD.

In concordance with previous ASL studies [7-11,16], highest accuracies for SVM training were observed using AD-specific ROIs instead of whole-brain. For AD vs. SCD, the combined parietal lobe and hippocampus ROI yielded the highest accuracy. However, we did not find increased accuracies using only the PC or PCC as a ROI, indicating that although these areas show the most pronounced hypo-perfusion [7-9,16], important information for optimal differentiation is lost when not taking the entire parietal lobe into account. A possibly explanation is the involvement of the default mode network (DMN), which has been shown to be less active in AD patients and includes several parietal areas [4].

We achieved maximal training accuracy for AD vs. MCI when applying both the parietal and occipital lobe ROI. Although the use of the parietal lobe is in line with previous ASL results [10,11], application of the occipital lobe is contradictive, as occipital hypo-perfusion has been observed in patients with MCI compared to controls, but not compared to patients with AD

[11,13]. The fact that the majority of our AD sample consisted of early-onset AD patients, who present with a more pronounced overall and occipital hypo-perfusion and hypo-metabolism compared to late-onset AD patients [33,34], could explain this discrepancy.

Automated classification of patients with MCI vs. subjects with SCD did not yield high accuracy, which was unexpected based on previous work [8-10]. The lack of homogeneity within the MCI group probably let to this low accuracy, as it prohibited the SVM from finding a pattern. Indeed, training with the MCI subgroups improved classification, with accuracies slightly higher compared to previous studies using sMRI [20,35]. This observation demonstrates the issue of heterogeneity between patients for classifiers [35]. However, this problem can be avoided by using large training data-sets. Since training with the MCI subgroups consisted of small samples ($N = 12$), replication of our results with a larger sample is necessary to support the prognostic value of the automated method for MCI-converters.

SVM classification training uses a leave-one-out cross-validation framework and thus reuses the same data for learning and classification, probably producing biased results. Unlike most preceding studies in this field, our study used independent data-sets for training and prediction to assess generalisability. In addition, our study extends previous results with sMRI by using larger sample sizes, reducing problems of disease heterogeneity and achieving increased classification accuracy. However, the number of data necessary for true clinical application of SVM in daily practice is notably higher [35]. Thus, a large multicentre study is of great clinical interest. Our results are limited by the fact that a relatively high proportion of our AD sample consists of early-onset patients. Further research using samples with older subjects will allow us to study the applicability for late-onset patients. Secondly, we used SCD subjects instead of healthy controls for our classification. However, SCD subjects are encountered in memory

clinics and hospitals and are therefore a good representation of this method's operational environment.

Our results support how automated classification can facilitate and possibly improve diagnosis, specifically in centres without experienced (neuro)radiologists. In addition, automated classification may be applicable for screening purposes, considering the high prevalence of AD [1].

5. CONCLUSION

Using automated methods, age- and gender adjusted ASL perfusion maps can be used to classify and predict diagnoses of AD, MCI-converters, stable MCI patients and SCD subjects with good accuracy and AUC values.

REFERENCES

1. Alzheimer's Association. Alzheimer's Association Report: disease facts and figures. Alzheimer's & Dementia 2014;10:e47-e92.

2. Murphy SL, Xu JQ, Kochanek KD. Deaths: final data for 2010. National Vital Statistics Reports 2013;61:4.

3. Morinaga A, Ono K, Ikeda T, et al. A comparison of the diagnostic sensitivity of MRI, CBF-SPECT, FDG-PET and cerebrospinal fluid biomarkers for detecting Alzheimer's disease in a memory clinic. Dementia and geriatric cognitive disorders 2010; 30(4): 285292.

4. Filippi M, Agosta F, Barkhof F, et al. EFNS task force: the use of neuroimaging in the diagnosis of dementia 2012;19:1487-1511.

5. Viola KL, Sbarboro J, Sureka R, et al. Towards non-invasive diagnostic imaging of earlystage Alzheimer's disease. Nature nanotechnology 2015;10(1): 91-98.

6. Jack CR Jr, Knopman DS, Jagust WJ, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurology 2010;9:119-128.

7. Binnewijzend MAA, Kuijer JPA, van der Flier, WM, et al. Distinct perfusion patterns in Alzheimer's disease, frontotemporal dementia and dementia with lewy bodies. Radiology 2014;24:2326-2333.

8. Binnewijzend MAA, Kuijer JPA, Benedictus MR, et al. Cerebral blood flow measured with 3D pseudocontinuous arterial spin-labelling MR imaging in Alzheimer disease and mild cognitive impairment: A marker for disease severity. Radiology 2013;267:221-230.

9. Wolk DA, Detre JA. Arterial spin labelling MRI: an emerging biomarker for Alzheimer's disease and other neurodegenerative conditions. Current Opinion in Neurology 2012; 25(4):421-428.

10. Dai W, Lopez OL, Carmichael OT, Becker JT, Kuller LH, Gach HM. Mild cognitive impairment and Alzheimer's disease: Patterns of altered cerebral blood flow at MR imaging. Radiology 2009;250(3);856-866.

11. Binnewijzend MAA, Benedictus MR, Kuijer JPA, et al. Cerebral perfusion in the predementia stages of Alzheimer's disease. European Radiology 2015;DOI 10.1007/s00330-015-3834-9.

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

13. Detre JA, Leigh JS, Williams DS, Koretsky AP. Perfusion imaging. Magnetic Resonance in Medicine 1992;23:37-45.

14. Alsop DC, Detre JA, Golay X, et al. Recommended implementation of arterial spinlabelled perfusion MRI for clinical applications: A consensus of the ISMRM Perfusion Study Group and the Europe Consortium for ASL in dementia. Magnetic Resonance in Medicine 2015;73:102-116.

15. Chen Y, Wolk DA, Redding JS et al. Voxel-level comparison of arterial spin-labelled perfusion MRI and FDT-PET in Alzheimer disease. Neurology 2011;**77**:1977-1985.

16. Verfaillie S, Adriaanse SM, Binnewijzend M, et al. Resting state cerebral perfusion and metabolism in subjective memory complaints: Alzheimer's disease and frontotemporal dementia – two sides of the same coin? Alzheimer's & Dementia 2014;10(4):543.

17. Klöppel S, Stonnington CM, Chu C, Draganski, et al. Automatic classification of MR scans in Alzheimer's disease. Brain 2008, 131(3):681-689.

18. Vemuri P et al. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. NeuroImage 2008;39:1186–1197.

19. DeCarli C. et al. Qualitative estimates of medial temporal atrophy as a predictor of progression from mild cognitive impairment to dementia. Arch. Neurology 2007;64:108–115.

20. Teipel S.J. et al. Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. NeuroImage 2007;38:13-24.

21. Schrouff J, Rosa MJ, Rondina JM. PRoNTo: Pattern Recognition for Neuroimaging Toolbox. Neuroinformatics 2013;11:319-337.

22. O'Brien PC, Dyck PJ. Procedures for setting normal values. Neurology 1995, 45(1), 17-23.

23. La Joie R, Perrotin A, Barré L, et al. Region-specific hierarchy between atrophy, hypometabolism, and β-amyloid (Aβ) load in Alzheimer's disease dementia. Neurobiology of Disease 2012;32(46):16265-16273.

24. Flier van der WM, Pijnenburg YA, Prins N, et al. Optimizing patient care and research: the Amsterdam Dementia Cohort 2014;41(1):313-327.

25. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. 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 1984;34(7):939-944.

26. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: Clinical characterization and outcome. Arch Neurol 1999;56: 303-308.

27. Verhage F. Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zevenenzeventigjaar 1964;Proefschrift. Assen:VanGorcum.

28. Buxton RB, Frank LR, Wong EC, Siewert B, Warach S, Edelman RR. A general kinetic model for quantitative perfusion imaging with arterial spin labelling. Magnetic Resonance in Medicine 1998;40(3):383-396.

29. Smith SM, Jenkinson M, Woolrich MW. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 2004;23(S1):208-219.

30. Pase MP, Grima NA, Stough C, Scholey A, Pipingas A. Association of pulsatile and mean cerebral blood flow velocity with age and neuropsychological performance. Elsevier Physiology and Behavior 2014;130:23-27.

31. Vapnik, V. The Nature of Statistical Learning Theory. Springer Verlag 1995, New York.

32. Kukull WA & Ganguli M. Genereralizability: The tress, the forrest, and the low-hanging fruit. Neurology 2012;78(23):1886-1891.

33. Kim EJ, Cho SS, Jeong Y, et al. Glucose metabolism in early onset versus late onset Alzheimer's disease: an SPM analysis of 120 patients. Brain 2005; 128: 1790–1801.

34. Kemp PM, Holmes C, Hoffmann SM, et al. Alzheimer's disease: differences in technetium-99m HMPAO SPECT scan findings between early onset and late onset dementia. Journal of Neurology, Neurosurgery, and Psychiatry 2003;74:715–19.

35. Klöppel S, Abdulkadir A, Jack Jr. CR, Koutsouleris N, Mourao-Miranda J, Vemuri P. Diagnostic neuroimaging across diseases. NeuroImage 2012;61:457-463.

TABLES

Table 1

Demographics and CSF Findings

Parameter*	Healthy Controls (reference group)	SCD Subjects	MCI Patients	AD Patients
No. of subjects	26	100	60	100
Age (years)	62.47 (7.33)	61.69 (6.56)	62.93 (6.48)	63.13 (5.66)
Women	62.05 (7.67)	61.54 (6.98)	63.19 (8.11)	62.64 (5.15)
Men	62.76 (7.32)	61.81 (6.30)	62.76 (5.25)	63.67 (6.17)
No. of women	11 (42.3%)	43 (43%)	24 (40%)	52 (52%)
MMSE[#]	29.00 (1.39)	28.07 (1.53)	26.72 (1.84)	20.27 (4.53)
Level of education	6.08 (0.48)	5.32 (1.23)	5.23 (1.40)	4.91 (1.24)
β-amyloid-42	N/A	899.42 (260.32)	770.58 (324.70)	513.22 (139.46)
Total Tau	N/A	323.54 (264.96)	452.35 (287.50)	705.10 (376.42)
Phosphorylated Tau	N/A	49.86 (25.20)	59.73 (30.61)	81.04 (32.08)

PRoNTo[§] Phase Groups	Training	Prediction	Training	Prediction	Training	Prediction
No. of subjects	50	50	30	30	50	50
Age (years)	61.73 (6.68)	61.66 (6.51)	62.67 (6.48)	63.19 (6.58)	63.25 (5.66)	63.02 (5.71)
Women	60.80 (6.11)	62.24 (7.79)	60.25 (8.87)	66.12 (6.31)	62.89 (5.01)	62.38 (5.38)
Men	62.40 (7.09)	61.21 (5.40)	64.28 (3.71)	61.24 (6.17)	63.64 (6.37)	63.70 (6.10)
No. of women	21 (42%)	22 (44%)	12 (40%)	12 (40%)	26 (52%)	26 (52%)
MMSE[#]	28.31 (1.29)	27.84 (1.71)	26.63 (1.90)	26.83 (1.80)	20.12 (4.90)	20.42 (4.18)
Level of education	5.38 (1.29)	5.26 (1.18)	5.17 (1.37)	5.30 (1.44)	4.66 (1.32)	5.16 (1.11)

*Data are presented as means +/- standard deviations

[#]MMSE = Mini-Mental State Examination

[§]Pattern Recognition for Neuroimaging Toolbox

Table 2.

Demographics and CSF Findings for MCI subgroups and matched SCD subjects

Parameter*	SCD Subjects	Stable MCI	MCI-converters
No. of subjects	12	12	12
Age (years)	64.99 (6.14)	65.79 (7.67)	65.62 (6.98)
Women	64.72 (5.47)	66.52 (9.03)	65.63 (7.11)
Men	65.37 (7.66)	64.77 (6.08)	65.60 (7.64)
No. of women	7 (58%)	7 (58%)	7 (58%)
MMSE[#]	28.75 (.97)	26.36 (1.91)	25.42 (1.78)
Level of education	6.08 (.67)	4.75 (1.67)	5.50 (1.35)
β-amyloid-42	805.90 (354.42)	805.00 (275.75)	446.33 (92.56)
Total Tau	340.10 (195.13)	483.14 (319.57)	560.67 (176.15)
Phosphorylated Tau	49.10 (16.65)	55.14 (30.00)	78.44 (19.97)
*Data are presented as means +/- standard deviations			
[#]MMSE = Mini-Mental State Examination			

Table 3.

Accuracies and AUC values ASL *W*-score maps – Training-set

Region of Interest	AD vs. SCD		AD vs. MCI		MCI vs. SCD		MCIc vs. SCD		MCIc vs. MCIs	
	Acc.	AUC	Acc.	AUC	Acc.	AUC	Acc.	AUC	Acc.	AUC
Whole Brain	87.0%**	0.94	78.8%**	0.84	**57.5%**	**0.49**	83.3%**	0.90	70.8%*	0.77
Parietal Lobe	87.0%**	0.92	72.5%**	0.84	55.0%	0.47	75.0%*	0.86	62.5%	0.67
PC$^{\$}$	74.0%**	0.81	68.8%**	0.81	45.0%	0.42	79.2%**	0.81	62.5%	0.68
PCC$^{\#}$	73.0%**	0.84	68.8%**	0.73	56.2%	0.56	83.3%**	0.87	62.5%	0.75
Hippocampus	68.0%**	0.79	61.3%	0.64	53.8%	0.51	79.2%**	0.83	**83.3%**	**0.77**
Occipital Lobe	71.0%**	0.81	75.0%**	0.79	55.0%	0.47	75.0%*	0.86	70.8%*	0.73
Parietal lobe + Hippocampus	**89.0%**	**0.93**	73.8%**	0.85	52.5%	0.49	79.2%*	0.88	66.7%	0.70
Parietal + Occipital Lobe	84.0%**	0.92	**83.8%**	**0.88**	50.0%	0.48	70.8%	0.83	66.7%	0.70
PCC + Hippocampus	74.0%**	0.83	72.5%**	0.78	50.0%	0.49	**87.5%**	**0.92**	75.0%*	0.81

$^{\$}$Precuneus
$^{\#}$ Posterior Cingulate Cortex
*$p \leq .05$
**$p \leq .01$

Table 4.

Accuracies and AUC values ASL *W*-score maps – Prediction-set

	AD vs. SCD	AD vs. MCI	MCI vs. SCD	MCIc vs. SCD	MCIc vs. MCIs
		82.0%			
Sensitivity	94.0%	84.0%	60.0%	83.0%	67.0%
Specificity	86.0%	80.0%	60.0%	75.0%	75.0%
AUC$^{\$}$	.96***	.89***	.63	.84**	.71

$^{\$}$Area Under the Curve

*$p \leq .05$

**$p \leq .01$ ***$p \leq .001$

Figures

Figure 1. Schematic overview of the grouping of participants for training and prediction.

Figure 2. Illustration of single-subject W-score computation. PC-ASL scans of the healthy controls (a) are used to compute a general linear model (b) and to compute the gender and age related confounds, resulting in the intercept group map (β0), the gender and age-related regression group maps (β1 and β2) and the single-subject maps (c) of residuals (ε), which is used to compute the SD of the residuals (d). The PC-ASL scans of the AD patients, MCI patients and SCD subjects (e) are as inputs in the W-score formula (f) and W-score maps are computed for each subject (g). The maps above show values higher than 1.65 (red) and lower than -1.65 (blue).

Figure 3. Discrimination maps training analysis main diagnostic groups. a) AD vs. SCD: Parietal lobe and hippocampus, AUC = .93 and accuracy = 89.0%. b) AD vs. MCI: Parietal and occipital lobe, AUC = .88 and accuracy 83.8%. c) MCI vs. SCD: Whole brain, AUC = .49 and accuracy 57.5%. MNI coordinates (x = 26, y = -20, z = 0).

Figure 4. Results prediction analysis main diagnostic groups. a) AD vs. SCD: AUC = .96 and accuracy = 90.0%. b) AD vs. MCI: AUC = .89 and accuracy 82.0%. c) MCI vs. SCD: AUC = .63 and accuracy = 60.0%.

Figure 5. Discrimination maps training analysis subgroups. a) MCIc vs. SCD: Posterior cingulate cortex and hippocampus, AUC = .92 and accuracy = 87.5%. b) MCIc vs. MCIs: Hippocampus, AUC = .77 and accuracy = 83.3%. MNI coordinates (x = 26, y = -20, z = 0).

Figure 6. Results prediction analysis MCI subgroups. a) MCIc vs. SCD: AUC = .84 and accuracy = 79.0%. b) MCIc vs. MCIs: AUC = .71 and accuracy 71.0%.