



**Automatic classification of AD and bvFTD based on cortical atrophy for single-subject diagnosis**

Journal:	<i>Radiology</i>
Manuscript ID:	Draft
Manuscript Type:	Original Research
Manuscript Categorization Terms:	Adults < 1. SUBJECT MATTER, MR-Imaging < 2. MODALITIES/TECHNIQUES, Head/Neck < 4. AREAS/SYSTEMS, Brain/Brain Stem < 5. STRUCTURES, Dementia < 6. TOPICS, Technology Assessment < 7. METHODOLOGY

SCHOLARONE™  
Manuscripts

1  
2  
3 **Automatic classification of AD and bvFTD based on cortical atrophy for**  
4  
5 **single-subject diagnosis**  
6  
7  
8  
9

10 Manuscript type: original research  
11

12 Advances in Knowledge:  
13

- 14 • Automated classifiers are able to discriminate between scans of patients with different  
15 forms of dementia and controls, based on gray matter (GM) patterns with high  
16 accuracy (75.3%-85.4%).  
17  
18
- 19 • Automated classifiers based on GM patterns can be used for single-subject diagnosis  
20 in an independent dataset with good to excellent diagnostic accuracy (AUC 0.81-  
21 0.95).  
22  
23
- 24 • Automated classifiers based on a generally available structural T1-weighted scans,  
25 make automated single-subject diagnosis more accessible and easy to use in daily  
26 clinical routine.  
27  
28  
29  
30  
31  
32  
33  
34

35 Implications for Patient Care:  
36

- 37 • Machine learning-based categorization methods could improve the diagnostic process  
38 in the daily practice, especially in centers without experienced neuroradiologists.  
39
- 40 • The application of automatic classification may be used for screening purposes in the  
41 future for high-risk groups.  
42  
43  
44  
45

46 Summary statement: Machine learning techniques are able to distinguish disease-specific GM  
47 atrophy between AD and bvFTD in a standard T1-weighted structural MRI scan for single-  
48 subject diagnosis.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Abstract**

**Purpose:** Assessment of the diagnostic accuracy of a support vector machine (SVM) classifier for individual patients based on common T1-weighted gray matter (GM) images without extensive preprocessing, using two independent data sets.

**Materials and Methods:** The local Institutional Review Board approved the study. 84 patients with Alzheimer's disease (AD), 51 patients with behavioral variant frontotemporal dementia (bvFTD), and 94 control subjects were divided into independent training- (n=115) and test-sets (n=114) with identical distributions of diagnosis and scanner type. The training-set was entered in a SVM for disease specific predictions of diagnostic status between two groups based on GM patterns. Weight-values of each voxel for classification were extracted. We conducted discriminant function analyses with leave-one-out cross-validation to determine if the extracted weight-values could be used for single-subject classification in the independent test-set, created receiver operating characteristic (ROC) curves and calculated the area under the curve (AUC). Threshold for statistical significance was  $p < 0.05$ .

**Results:** In the training-set, the accuracy of the SVM to discriminate AD from controls was 85.3%, bvFTD from controls 72.3%, and AD from bvFTD 78.7% ( $p \leq 0.029$ ). For single-subject diagnosis in the test-set, the extracted weight-values discriminated 87.6% of AD and controls, 84.7% of bvFTD and controls, and 82.1% of AD and bvFTD correctly. ROC curves revealed good to excellent AUC (0.81-0.95;  $p \leq 0.001$ ).

**Conclusions:** SVMs can be used in single-subject discrimination and help the clinician in making a diagnosis. SVMs can distinguish disease-specific GM patterns in AD and bvFTD from those of normal aging using common T1-weighted structural MRI scans.

## Introduction

Alzheimer's disease (AD) and behavioral variant frontotemporal dementia (bvFTD) are the most common causes of young onset dementia (1). Clinical diagnostic criteria are available (2, 3), but the frequent overlap of the clinical symptoms associated with AD and bvFTD poses serious problems in the differential diagnosis.

Magnetic resonance imaging (MRI), has been shown to be able to detect disease-specific macroscopic brain changes in an early disease stage. Several studies investigated morphometric gray matter (GM) changes to discriminate AD from bvFTD. These studies typically report group-level differences for various brain structures (4-6). Besides structural MRI, other imaging modalities, as positron emission tomography, functional or diffusion MRI, have been promising in the discrimination between AD and bvFTD. However, some of these techniques are invasive, time-consuming, require the availability of specialized scanners and are difficult to implement in a clinical setting without technical support (7, 8).

In most memory clinics, patients are usually scanned once during dementia screening, with a standardized protocol generally including a T1-weighted 3-dimensional (3DT1) MRI sequence. This sequence is representative of the disease-specific structural changes, capable of providing similar information irrespective of different scanners and is easy to obtain.

Automatic individual patient classifications based on a structural 3DT1 scan at one time point could support the clinician's diagnosis.

A support vector machine (SVM) is a machine learning technique that can categorize individual brain images by differentiation of images from two groups. These automated classifiers can be objective, quantitative and easy to implement and potentially satisfy the requirements of a diagnostic tool (9).

The available literature on SVM shares common limitations: Discriminating AD from FTD using the whole spectrum of FTD and not only the behavioral variant (7, 10-12) will be only

1  
2  
3 representative for the language variants as their asymmetric atrophy will determine the  
4  
5 classification accuracy (13). Using only cross-validation in the discrimination may result in  
6  
7 biased estimates, especially when the sample size is small (11, 14, 15). The use of a region-of-  
8  
9 interest (ROI) approach or different imaging modalities for the discrimination (16, 17) restrain  
10  
11 the implementation of the SVM approach in the daily practice, as extraction of ROIs is time-  
12  
13 consuming. Another problem of the ROI based approach is the limited generalizability of the  
14  
15 trained automatic classifier using single-center data when applied to new data sets acquired on  
16  
17 different scanners.  
18

19  
20 Therefore, we explore the diagnostic accuracy of a SVM for individual patients based on a  
21  
22 generally available T1-weighted GM image without extensive preprocessing. To increase  
23  
24 generalizability we used MRI scans from different scanners and two independent data sets in  
25  
26 a cross-sectional design.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Materials and Methods

### *Patients*

In this study, we included 84 patients with AD, 51 patients with bvFTD, and 53 patients with subjective memory complaints who visited either the Alzheimer Center of the VU University Medical center or the Alzheimer Center of the Erasmus University Medical Center Rotterdam. All patients underwent a standardized one-day assessment including medical history, informant-based history, physical and neurological examination, blood tests, neuropsychological assessment, and MRI of the brain. Diagnoses were made in a multidisciplinary consensus meeting according to the core clinical criteria of the National Institute on Aging and the Alzheimer's Association workgroup for probable AD (3, 18) and according to the International FTD Consortium criteria for bvFTD (2) based on the results of the one-day assessment as described above. Patients were diagnosed as having subjective memory complaints when they presented with memory complaints, but cognitive functioning was normal and criteria for MCI (19), dementia or any other neurological or psychiatric disorder known to cause cognitive decline were not met. To minimize center effects, all diagnoses were re-evaluated in a panel including clinicians from both centers. In addition, we included 41 cognitively normal controls, who were recruited by advertisements in local newspapers. Before inclusion in the present study, they were screened for memory complaints, family history of dementia, drugs- or alcohol abuse, major psychiatric disorder, and neurological or cerebrovascular diseases. They underwent an assessment including medical history, physical examination, neuropsychological assessment, and MRI of the brain comparable to the work-up of patients. Cognitively normal controls and patients with subjective memory complaints served both as controls to obtain a representative control group for the general population. Patients and controls were randomly split into equally sized

1  
2  
3 independent training- (n=115) and test-sets (n=114) with identical distribution of diagnosis  
4  
5 and scanner type.

6  
7 Disease duration was calculated based on the time difference between date of diagnosis and  
8  
9 the year patients caregivers noticed the first symptoms. The local medical ethics committee of  
10  
11 both centers approved the study. All patients and controls gave written informed consent for  
12  
13 their clinical data to be used for research purposes.  
14  
15

### 16 17 18 *MR image acquisition and review* 19

20  
21 Imaging at the VUmc was carried out on two 3T scanners (Signa HDxt, GE Healthcare,  
22  
23 Milwaukee, WI, USA and TF PET/MR, Philips Medical Systems, Cleveland, Ohio, USA),  
24  
25 using an 8-channel head coil with foam padding. Patients and controls from the Erasmus  
26  
27 University Medical Center Rotterdam were all scanned at the Leiden University Medical  
28  
29 Center (LUMC) on a 3T scanner (Achieva, Philips Medical Systems, Best, the Netherlands)  
30  
31 using an 8-channel head coil. The scan protocol included a whole-brain near-isotropic 3DT1-  
32  
33 weighted sequence. At the VUmc this was a fast spoiled gradient echo sequence (FSPGR;  
34  
35 repetition time TR 7.8 ms, echo time TE 3 ms, inversion time TI 450 ms, flip angle 12°, 180  
36  
37 sagittal slices, voxel size 0.98x0.98x1 mm, total scan time 4.57 minutes) or a turbo field echo  
38  
39 sequence (T1TFE; TR 7 ms, TE 3 ms, flip angle 12°, 180 sagittal slices, voxel size 1x1x1  
40  
41 mm, total scan time 6.14 minutes). At the LUMC this was a turbo field echo sequence  
42  
43 (T1TFE; TR 9.8 ms, TE 4.6 ms, flip angle 8°, 140 transversal slices, voxel size 0.88x0.88x1.2  
44  
45 mm, total scan time 4.57 minutes).  
46  
47  
48

49  
50 In addition, the MRI protocols included a 3D Fluid Attenuated Inversion Recovery (FLAIR)  
51  
52 sequence, dual-echo T2-weighted sequence, and susceptibility weighted imaging (SWI) which  
53  
54 were reviewed for brain pathology other than atrophy by an experienced radiologist.  
55  
56  
57  
58  
59  
60

*MR processing*

DICOM images of the 3DT1- weighted sequence from the Signa HDxt were corrected for gradient nonlinearity distortions. All scans were converted to Nifti format. The linear transformation matrix to MNI space was calculated using FSL-FLIRT (20) and used to place the image coordinate origin (0,0,0) on the anterior commissure by using the Nifti s-form. The structural 3DT1 images were then analyzed using the voxel-based morphometry toolbox (VBM8; version 435; University of Jena, Department of Psychiatry) in Statistical Parametric Mapping (SPM8; Functional Imaging Laboratory, University College London, London, UK) implemented in MATLAB 7.12 (MathWorks, Natick, MA). Data of the training- and test set were preprocessed separately with VBM8 to avoid any bias (21). The first module of the VBM8 Toolbox (“Estimate and Write”) segments the 3DT1 volumes into GM, white matter (WM) and cerebrospinal fluid (CSF), applies a registration to MNI space (affine) and subsequently a non-linear deformation. The non-linear deformation parameters are calculated via the high dimensional DARTEL algorithm and the MNI 152 template. Remaining non-brain tissue was removed by the ‘light clean-up’ option. Tissue classes were normalized in alignment with the template with the ‘non-linear only’ option which allows comparing the absolute amount of tissue corrected for individual brain size. The correction is applied directly to the data, which makes a head-size correction to the statistical model redundant. In the second module, images were smoothed using an eight mm full width at half maximum (FWHM) isotropic Gaussian kernel. All images were visually inspected after every processing step.

*Support vector machine classification*

For the pattern recognition analysis we used the “Pattern Recognition for Neuroimaging Toolbox” (PRoNTo) (21), implemented in MATLAB 7.12 following the standard descriptions

1  
2  
3 of the manual (<http://www.mnl.cs.ucl.ac.uk/pronto>). The normalized, modulated, smoothed  
4  
5 GM images of all subjects in the training dataset were used as inputs. We used a custom  
6  
7 mask, which was made by the mean of all smoothed GM segmentations and binarized at a  
8  
9 threshold of 0.2. We used a binary SVM to classify (1) AD from controls, (2) bvFTD from  
10  
11 controls, and (3) AD from bvFTD, all with leave-one out cross-validation. To estimate how  
12  
13 much each voxel contributes to the classification task, we calculated the voxel-wise  
14  
15 ‘discrimination maps’(21). The weights are the model parameters learned by the SVM  
16  
17 projected back to the input space. The weight-value can be positive or negative, where a  
18  
19 positive value represents a higher weighted average for class one, a negative value represents  
20  
21 a higher weighted average for class two. These maps are shown for each classifier in Figure 1.  
22  
23 For illustrative purposes the weight maps are thresholded at 30% of the maximum positive  
24  
25 and negative weight values, in line with Mourao-Miranda (22). The performance of the  
26  
27 classifier in the training-set was expressed in balanced accuracy (class-specific accuracy),  
28  
29 sensitivity and specificity. Permutation testing was used to derive a p-value to determine  
30  
31 whether the balanced accuracy exceeded chance levels (50%). Class labels were permuted  
32  
33 1000 times.

34  
35  
36  
37  
38 To test whether the learned weight values from the training-set could classify unseen subjects,  
39  
40 we extracted the weights of every voxel of the weight map over all folds and multiplied them  
41  
42 with the normalized, modulated, smoothed GM images of each subject from the independent  
43  
44 test-set. This integrated product of the weight map and the smoothed GM image per subject  
45  
46 was averaged and transferred to SPSS for further analyses. This was done separately for the  
47  
48 classification between AD-controls, bvFTD-controls, and AD-bvFTD.  
49  
50

### 51 52 53 54 *Statistical analyses* 55 56 57 58 59 60

1  
2  
3 SPSS version 20.0 for Windows was used for statistical analysis. Differences between groups  
4  
5 were assessed using univariate analysis of variance (ANOVA), Kruskal-Wallis tests and  $\chi^2$   
6  
7 tests, where appropriate.  
8

9  
10 To determine the performance of the SVM for single-subject classification in an independent  
11  
12 test-set, we conducted three discriminant function analyses between two groups with leave-  
13  
14 one-out cross validation. As predictor we entered the averaged integrated product of the  
15  
16 weight map and the smoothed GM image per subject from the corresponding classification of  
17  
18 the SVM (e.g. for the discriminant analysis between AD and controls, the average integrated  
19  
20 product from weight map “AD vs. controls” was used for prediction). Additionally, we  
21  
22 created a receiver operation characteristic (ROC) and calculated the area under the curve  
23  
24 (AUC). Statistical significance was set at  $p < 0.05$ .  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Results

### *Demographics*

Demographic data for training- and test-set are summarized in Table 1. There were no differences in age, sex, scanner type, disease duration, MMSE score, and diagnosis. To get more insight in the different datasets and if patient groups were comparable, we examined the demographic data of training set and the test set based on diagnosis: There were differences in sex, scanner type and disease duration. In both datasets, AD patients were older than controls and had the lowest MMSE scores (Table 2).

### *Support vector classification*

#### *Training-set*

Performance of the binary SVM is summarized in Figure 2. The classifier discriminated AD patients from controls with 85.3% balanced accuracy ( $p=0.001$ ). Sensitivity for classification of AD patients was 83.3% ( $p=0.001$ ) and specificity 87.2% ( $p=0.001$ ).

A correct distinction between bvFTD patients and controls, was made in 72.3% (balanced accuracy,  $p=0.001$ ). Sensitivity for classification of bvFTD patients was 61.5% ( $p=0.001$ ) and specificity 83% ( $p=0.029$ ).

The classifier discriminated AD from bvFTD with 78.7% balanced accuracy ( $p=0.001$ ). Sensitivity for classification of AD patients was 88.1% ( $p=0.001$ ) and specificity was 69.2% ( $p=0.001$ ).

#### *Generalizability of the classifiers for single-subject diagnosis*

Results are summarized in Figure 3. In the independent test-set the extracted weights discriminated 87.6% of AD patients and controls correctly, with correct classification of 36 AD patients (85.7%) and 42 controls (89.4%). The ROC curve revealed an excellent AUC for

1  
2  
3 the extracted weights (0.95;  $p < 0.001$ ). For the discrimination between bvFTD patients and  
4  
5 controls, the extracted weights predicted 84.7% of all cases correctly, with correct  
6  
7 classification of 15 bvFTD patients (60%) and 46 controls (97.9%). The ROC curve revealed  
8  
9 a good AUC for the extracted weights (0.87;  $p < 0.001$ ). For the discrimination between AD  
10  
11 and bvFTD patients, the extracted SVM weights predicted 82.1% of all cases correctly, with  
12  
13 correct classification of 39 AD (92.9%) and 16 bvFTD patients (64%). The ROC curve  
14  
15 revealed a good AUC for the extracted weights (0.81;  $p < 0.001$ ).  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Discussion

In this study we showed that it is possible to discriminate between scans of patients with different forms of dementia and controls, based on GM patterns with an automated classifier with high accuracy (75.3%-85.4%). Crucially, we have also demonstrated that automated classifiers can be used in single-subject diagnosis, as its diagnostic accuracy in an independent dataset of patients and controls was good to excellent (AUC 0.81-0.95).

The accuracy levels in our study are comparable with other studies using SVM in differentiation of AD and FTLD (9, 11, 16, 23). The slightly higher accuracy values found in two other studies can be explained by the use of all subtypes of FTLD. It is possible that the specific atrophy patterns of the language variant of the FTLD spectrum increased the diagnostic power (9, 11). As it is clinically relevant to make a clear distinction between the language variant and bvFTD, a combination of these forms based on atrophy patterns is undesirable. Besides that, AD and FTD patients in those previous studies had lower scores on the MMSE than in our study, which are indicative of a later disease stage and presumably more disease related GM atrophy, facilitating discrimination. In our study, only bvFTD was used and patient groups had a comparable age and disease duration, rendering these confounding effects less likely. Furthermore, as the SVM identified brain regions, which contribute highly to the classification, are in agreement with results of literature on GM atrophy in AD and bvFTD, we are confident that our results are valid.

To our knowledge, there is only one other study that used a separate training- and test-set to evaluate the predictive power of a SVM (16). The accuracy levels of our study are comparable and demonstrate the robustness of the performance of the automated classifier in independent datasets. Our study extends these results by testing larger sample sizes and fully

1  
2  
3 independent training- and test-sets. We also used whole brain information instead of a ROI  
4  
5 approach, making our approach easier to implement in daily clinical routine.  
6  
7

8  
9 Although several studies examining cross-sectional and longitudinal effects in volumes of  
10  
11 brain regions have shown significant group differences between AD, bvFTD, and controls, the  
12  
13 ability to detect structural patterns that enable accurate single-subject predictions will  
14  
15 ultimately assist the diagnostic process in the daily practice. Our study focused on making  
16  
17 automated single-subject diagnosis more accessible, taking into account multiple factors of  
18  
19 the daily clinical practice, such as scans from different MR scanners and a control group  
20  
21 consisting of healthy elderly controls and people with subjective memory complaints. The  
22  
23 method we described, clearly has potential in achieving more accurate dementia diagnosis in  
24  
25 clinical practice. Admittedly, the processing and preparation of a training dataset is time  
26  
27 consuming, but once a training dataset is available, the spatial normalization and classification  
28  
29 of any new scan can be performed instantaneously.  
30  
31  
32  
33  
34  
35

36 A possible limitation of this study is that we did not have post-mortem data available, so the  
37  
38 possibility of misdiagnosis cannot be excluded. Nevertheless, we used an extensive  
39  
40 standardized work-up and all AD patients fulfilled clinical criteria of probable AD, 42  
41  
42 patients fulfilled the criteria for probable bvFTD and nine for possible bvFTD. All diagnoses  
43  
44 were re-evaluated in a panel including clinicians from both centers to minimize sample  
45  
46 effects. We only used binary classifiers, which means that a test case not belonging to one of  
47  
48 the two groups will be incorrectly assigned to one of these. Multi-class classifiers software for  
49  
50 MRI are currently not widely available. Future releases of available machine learning models  
51  
52 will facilitate multi-class classifications, and thereby improve the diagnostic usefulness.  
53  
54  
55 However, the current finding indicates that a binary classifier can assist the diagnostic process  
56  
57  
58  
59  
60

1  
2  
3 by predicting different dementia subtypes based on localized changes in GM density.  
4  
5 Discrimination with pattern recognition models is based on whole brain information, rather  
6  
7 than on individual regions or voxels. Therefore, all voxels contribute to the classification and  
8  
9 no conclusions should be drawn about a particular subset of voxels in isolation. However,  
10  
11 some brain areas may be more informative about class membership than others. It could be  
12  
13 argued that valuable classification power is lost due to the whole brain approach and well-  
14  
15 placed ROIs would improve categorization. However, a disadvantage of a ROI based  
16  
17 approach is that it might not be as generalizable as a classifier that takes into account whole  
18  
19 brain information. Nevertheless, the aim of our study was to achieve optimal classification  
20  
21 based on whole brain information to build an optimal classifier in a way that it could be easily  
22  
23 used in daily practice.  
24  
25  
26  
27  
28

29  
30 Our results indicate that **machine learning techniques can aid the clinical diagnosis. The**  
31  
32 **analytical technique presented here is able to distinguish disease-specific GM atrophy**  
33  
34 **between AD and bvFTD in a standard T1-weighted structural MRI scan for single-**  
35  
36 **subjects.** A goal of machine learning based automated MR image analysis is higher  
37  
38 sensitivity and specificity of ante-mortem diagnosis than is currently possible. A study by  
39  
40 Klöppel (23) showed that computer-based diagnosis is equal to or better than that achieved by  
41  
42 radiologists. Together with our results, it is conceivable that in the future machine learning-  
43  
44 based categorization methods could improve diagnosis, especially in centers without  
45  
46 experienced neuroradiologists. An important next step will be the application of automatic  
47  
48 classification for screening purposes. If machine learning discrimination is sensitive enough  
49  
50 to classify subtle differences, early screening of high-risk groups could be easily  
51  
52 implemented.  
53  
54  
55  
56  
57  
58  
59  
60

## Reference List

1. Ratnavalli E, Brayne C, Dawson K, Hodges JR. The prevalence of frontotemporal dementia. *Neurology* 2002; 58:1615-1621.
2. Rascovsky K, Hodges JR, Knopman D, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* 2011; 134:2456-2477.
3. McKhann GM, Knopman DS, Chertkow H, 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* 2011; 7:263-269.
4. Rabinovici GD, Seeley WW, Kim EJ, et al. Distinct MRI atrophy patterns in autopsy-proven Alzheimer's disease and frontotemporal lobar degeneration. *Am J Alzheimers Dis Other Dement* 2007; 22:474-488.
5. Whitwell JL, Jack CR, Jr. Comparisons between Alzheimer disease, frontotemporal lobar degeneration, and normal aging with brain mapping. *Top Magn Reson Imaging* 2005; 16:409-425.
6. Du AT, Schuff N, Kramer JH, et al. Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia. *Brain* 2007; 130:1159-1166.
7. Dukart J, Mueller K, Horstmann A, et al. Combined evaluation of FDG-PET and MRI improves detection and differentiation of dementia. *PLoS One* 2011; 6:e18111.
8. Bron EE, Steketee RM, Houston GC, et al. Diagnostic classification of arterial spin labeling and structural MRI in presenile early stage dementia. *Hum Brain Mapp* 2014.
9. Kloppel S, Stonnington CM, Chu C, et al. Automatic classification of MR scans in Alzheimer's disease. *Brain* 2008; 131:681-689.
10. Kloppel S, Stonnington CM, Barnes J, et al. Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. *Brain* 2008; 131:2969-2974.
11. Davatzikos C, Resnick SM, Wu X, Parnpi P, Clark CM. Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. *Neuroimage* 2008; 41:1220-1227.
12. Vemuri P, Simon G, Kantarci K, et al. Antemortem differential diagnosis of dementia pathology using structural MRI: Differential-STAND. *Neuroimage* 2011; 55:522-531.
13. Whitwell JL, Josephs KA, Rossor MN, et al. Magnetic resonance imaging signatures of tissue pathology in frontotemporal dementia. *Arch Neurol* 2005; 62:1402-1408.

14. Plant C, Teipel SJ, Oswald A, et al. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. *Neuroimage* 2010; 50:162-174.
15. Zhang D, Wang Y, Zhou L, Yuan H, Shen D. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 2011; 55:856-867.
16. Raamana PR, Rosen H, Miller B, Weiner MW, Wang L, Beg MF. Three-Class Differential Diagnosis among Alzheimer Disease, Frontotemporal Dementia, and Controls. *Front Neurol* 2014; 5:71.
17. Duchesne S, Caroli A, Geroldi C, Barillot C, Frisoni GB, Collins DL. MRI-based automated computer classification of probable AD versus normal controls. *IEEE Trans Med Imaging* 2008; 27:509-520.
18. 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; 34:939-944.
19. Petersen RC, Stevens JC, Ganguli M, Tangalos EG, Cummings JL, DeKosky ST. Practice parameter: early detection of dementia: mild cognitive impairment (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 2001; 56:1133-1142.
20. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. *Med Image Anal* 2001; 5:143-156.
21. Schrouff J, Rosa MJ, Rondina JM, et al. PRoNTTo: pattern recognition for neuroimaging toolbox. *Neuroinformatics* 2013; 11:319-337.
22. Mourao-Miranda J, Reinders AA, Rocha-Rego V, et al. Individualized prediction of illness course at the first psychotic episode: a support vector machine MRI study. *Psychol Med* 2012; 42:1037-1047.
23. Kloppel S, Stonnington CM, Barnes J, et al. Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. *Brain* 2008; 131:2969-2974.

Table 1. Demographics

	Training set	Test set
N	115	114
Scanners		
GE	84 (73%)	83 (73%)
PETMR	9 (8%)	9 (8%)
Philips	22 (19%)	22 (19%)
Diagnosis		
AD	42 (37%)	42 (37%)
bvFTD	26 (22%)	25 (22%)
Controls	47 (41%) (20 HC, 27 SMC)	47 (41%) (21 HC, 26 SMC)
Age	62.7 ± 7.5	62.7 ± 6.7
Sex, f	36 (31%)	36 (31%)
MMSE	25.2 ± 4.4	25.2 ± 4.3
Disease duration, months	42.6 ± 32.4	43.2 ± 31.0

Values presented as mean ± standard deviation or n (%). Differences between groups for demographics were assessed using ANOVA. Kruskal-Wallis tests and  $\chi^2$  tests, where appropriate. Key: SMC: Subjective memory complaints; MMSE: Mini-Mental State Examination

Table 2. Demographics within training- and test-set based on diagnosis.

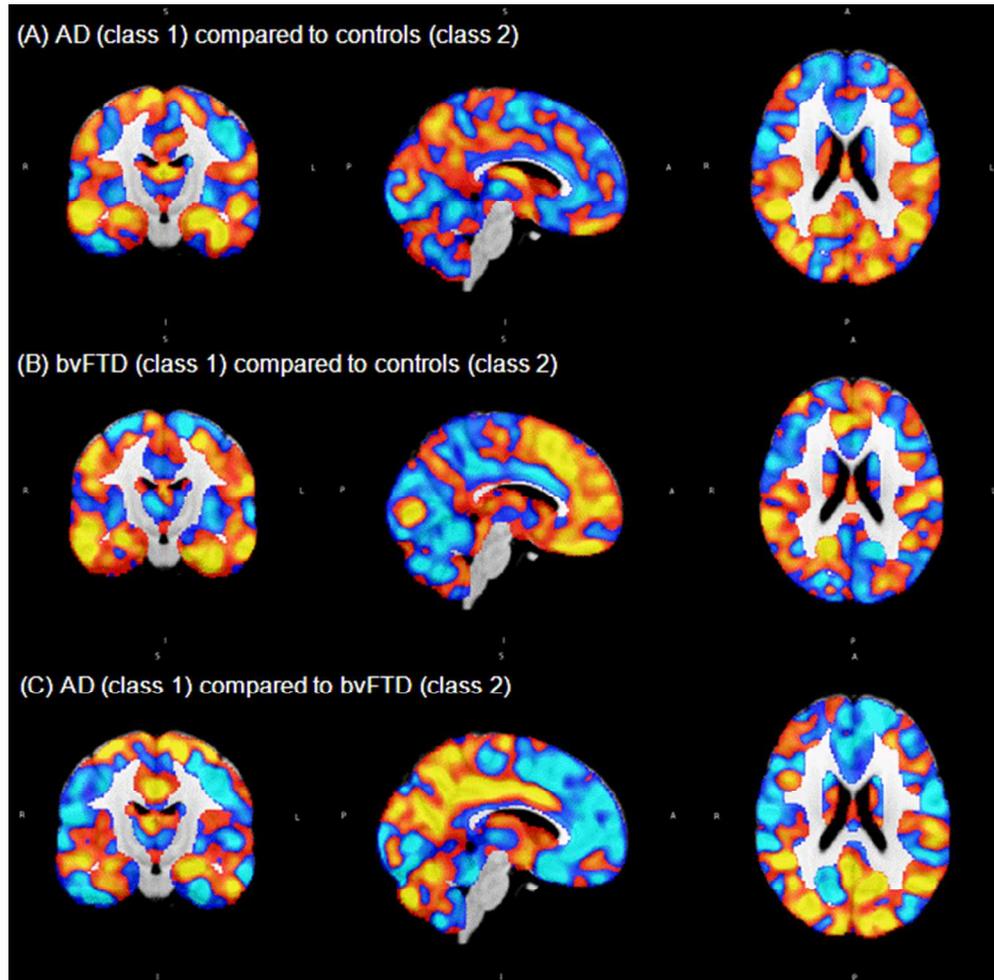
	Training set			Test set		
	AD	bvFTD	Controls	AD	bvFTD	Controls
N	42	26	47	42	25	47
Scanner						
GE	31 (74%)	18 (69%)	35 (75%)	31(74%)	17 (68%)	35 (75%)
PETMR	3 (7%)	3 (12%)	3 (6%)	3 (7%)	3 (12%)	3 (6%)
Philips	8 (19%)	5 (19%)	9 (19%)	8 (19%)	5 (20%)	9 (19%)
Age	64.9 ± 7.1 <sup>a</sup>	62.1 ± 7.8	61 ± 7.3	65.3 ± 7.1 <sup>a</sup>	61.6 ± 6.4	60.9 ± 5.7
Sex, f	13 (31%)	9 (35%)	14 (30%)	16 (38%)	4 (16%)	16 (34%)
MMSE	21.9 ± 4.5 <sup>a</sup>	24.8 ± 3.4 <sup>a,b</sup>	28.5 ± 1.6	22 ± 3.7 <sup>a</sup>	24.6 ± 3.8 <sup>a,b</sup>	28.5 ± 2
Disease duration, months	36.6 ± 22.1	44.6 ± 40.3	-	44.4 ± 30.1	41.0 ± 30.8	-

Values presented as mean ± standard deviation or n (%). Differences between groups for demographics were assessed using ANOVA. Kruskal-Wallis tests and  $\chi^2$  tests, where appropriate. Key: MMSE: Mini-Mental State Examination. <sup>a</sup> different from controls (p<0.05), <sup>b</sup> different from AD (p<0.05)

1  
2  
3 Figure 1. Weight maps for each classifier at a threshold of 30% of the maximum positive and  
4 negative weight values, superimposed onto a standard brain template from FSL  
5 (MNI152\_T1\_1mm) showing areas of the brain most vital for discriminating the two groups.  
6  
7 Red-Yellow: negative values indicative for class 1. Blue-Light blue: positive values indicative  
8 for class 2. (A) AD vs. controls, (B) bvFTD vs. controls, (C) AD vs. bvFTD.  
9  
10  
11  
12  
13  
14  
15

16 Figure 2. Performance of support vector machine classification of training set data. (A) AD  
17 vs. controls, (B) bvFTD vs. controls, (C) AD vs. bvFTD.  
18  
19  
20  
21  
22

23 Figure 3. Discriminating performance in test-set of averaged integrated product of weight map  
24 from the training-set and smoothed GM images from the test-set. Scatterplots showing  
25 discrimination between two groups. The ROC curve illustrates the performance of the binary  
26 classifier.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Figure 1. Weight maps for each classifier at a threshold of 30% of the maximum positive and negative weight values, superimposed onto a standard brain template from FSL (MNI152\_T1\_1mm) showing areas of the brain most vital for discriminating the two groups. Red-Yellow: negative values indicative for class 1. Blue-Light blue: positive values indicative for class 2. (A) AD vs. controls, (B) bvFTD vs. controls, (C) AD vs. bvFTD.

176x173mm (96 x 96 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

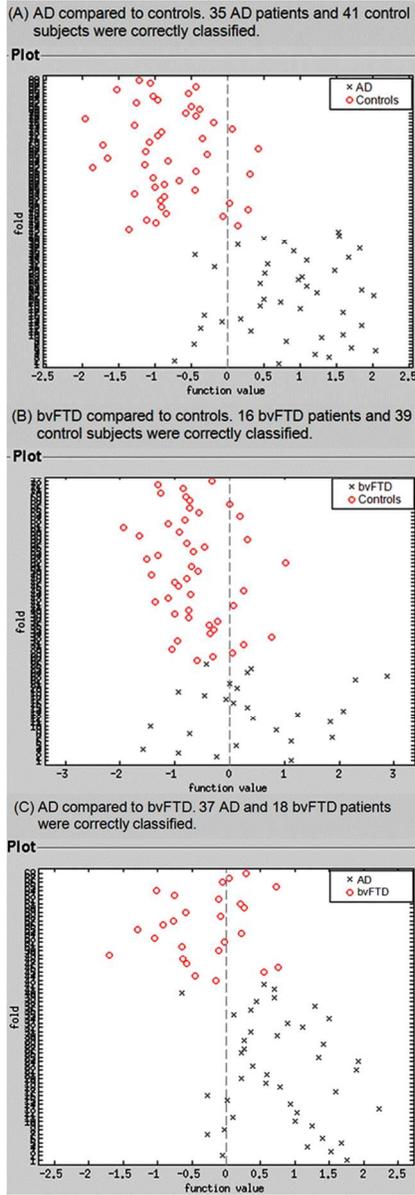


Figure 2. Performance of support vector machine classification of training set data. (A) AD vs. controls, (B) bvFTD vs. controls, (C) AD vs. bvFTD. 132x385mm (96 x 96 DPI)

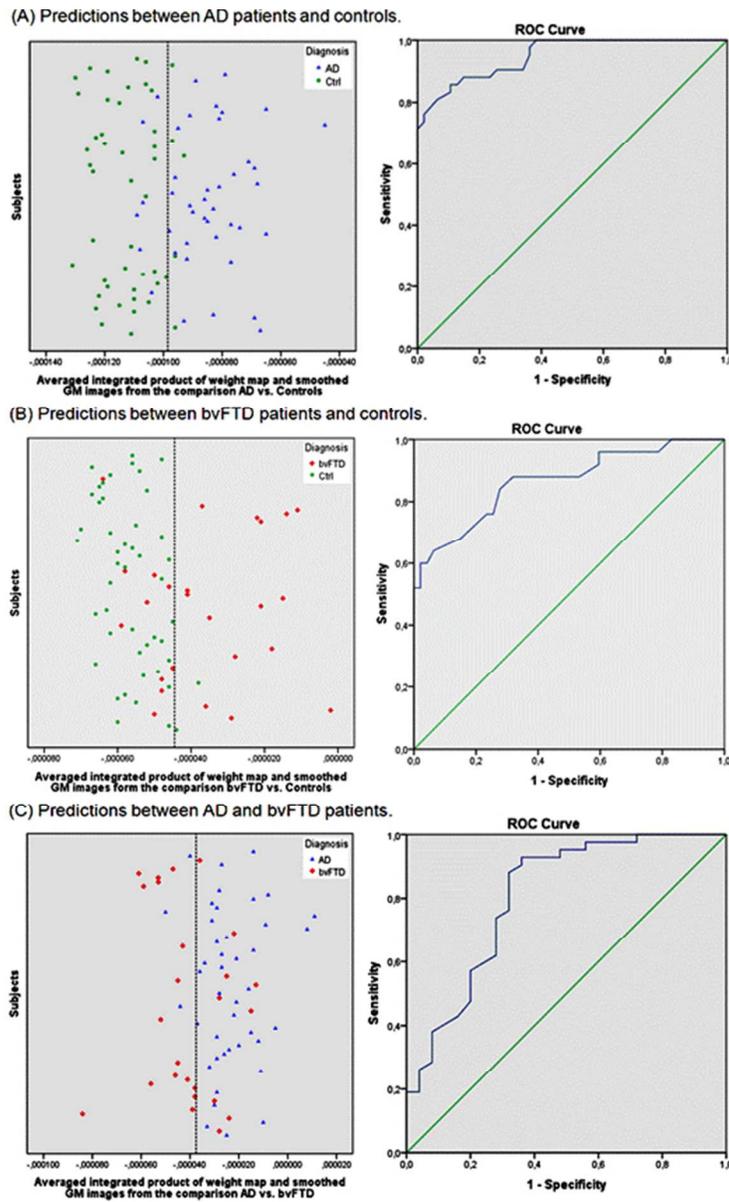


Figure 3. Discriminating performance in test-set of averaged integrated product of weight map from the training-set and smoothed GM images from the test-set. Scatterplots showing discrimination between two groups. The ROC curve illustrates the performance of the binary classifier.  
138x226mm (96 x 96 DPI)