



Universiteit
Leiden
The Netherlands

Single Subject Classification of Alzheimer's Disease and Behavioral Variant Frontotemporal Dementia Using Anatomical, Diffusion Tensor, and Resting-State Functional Magnetic Resonance Imaging.

Bouts, M.J.R.J.; Möller, C.; Hafkemeijer, A.; Swieten, J.C. van; Dopper, E.; Flier, W.M. van der; ... ; Rooij, M. de

Citation

Bouts, M. J. R. J., Möller, C., Hafkemeijer, A., Swieten, J. C. van, Dopper, E., Flier, W. M. van der, ... Rooij, M. de. (2018). Single Subject Classification of Alzheimer's Disease and Behavioral Variant Frontotemporal Dementia Using Anatomical, Diffusion Tensor, and Resting-State Functional Magnetic Resonance Imaging. *Journal Of Alzheimer's Disease*, 62(4), 1827-1839. doi:10.3233/JAD-170893

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/73457>

Note: To cite this publication please use the final published version (if applicable).

Single subject classification of Alzheimer's disease and behavioral variant frontotemporal dementia using anatomical, diffusion tensor, and resting-state functional magnetic resonance imaging

Mark J.R.J. Bouts^{a,b,c*}, Christiane Möller^{a,b,c}, Anne Hafkemeijer^{a,b,c}, John C. van Swieten^{d,e}, Elise Dopper^{e,f}, Wiesje M. van der Flier^{f,g}, Hugo Vrenken^{h,i}, Alle Meije Wink^h, Yolande A.L. Pijnenburg^f, Philip Scheltens^f, Frederik Barkhof^{h,j}, Tijn. M. Schouten^{a,b,c}, Frank de Vos^{a,b,c}, Rogier A. Feis^{b,c}, Jeroen van der Grond^b, Mark de Rooij^{a,c}, Serge A.R.B. Rombouts^{a,b,c}

^aInstitute of Psychology, Leiden University, Leiden, the Netherlands;

^bDepartment of Radiology, Leiden University Medical Center, Leiden, the Netherlands;

^cLeiden Institute for Brain and Cognition, Leiden University, Leiden, the Netherlands;

^dDepartment of Clinical Genetics, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, the Netherlands;

^eDepartment of Neurology, Erasmus Medical Center, Rotterdam, the Netherlands;

^fAlzheimer Center & Department of Neurology, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, the Netherlands;

*^gDepartment of Epidemiology & Biostatistics, Neuroscience Campus Amsterdam,
VU University Medical Center, Amsterdam, the Netherlands;*

*^hDepartment of Radiology & Nuclear Medicine, Neuroscience Campus Amsterdam,
VU University Medical Center, Amsterdam, the Netherlands,*

*ⁱDepartment of Physics & Medical Technology, Neuroscience Campus Amsterdam,
VU University Medical Center, Amsterdam, the Netherlands.*

*^jInstitute of Neurology & Healthcare Engineering, University College London,
London, United Kingdom.*

Running title: MRI-based differentiation of bvFTD and AD

*Corresponding author:

Mark Bouts, PhD

Leiden University, Institute of Psychology

PO Box 9555

2300 RB Leiden, The Netherlands

Phone: +31 71 5276911

Email: m.j.r.j.bouts@fsw.leidenuniv.nl

Abstract

Background / Objective:

Overlapping clinical symptoms often complicate differential diagnosis between patients with Alzheimer's disease (AD) and behavioral variant frontotemporal dementia (bvFTD). Magnetic resonance imaging (MRI) reveals disease specific structural and functional differences that aid in differentiating AD from bvFTD patients. However, the benefit of combining structural and functional connectivity measures to – on a subject-basis - differentiate these dementia-types is not yet known.

Methods:

Anatomical, diffusion tensor (DTI), and resting-state functional MRI (rs-fMRI) of 30 patients with early stage AD, 23 with bvFTD, and 35 control subjects were collected and used to calculate measures of structural and functional tissue status. All measures were used separately or selectively combined as predictors for training an elastic net regression classifier. Each classifier's ability to accurately distinguish dementia-types was quantified by calculating the area under the receiver operating characteristic curves (AUC).

Results:

Highest AUC values for AD and bvFTD discrimination were obtained when mean diffusivity, full correlations between rs-fMRI-derived independent components, and fractional anisotropy (FA) were combined (0.811). Similarly, combining gray matter density (GMD), FA, and rs-fMRI correlations resulted in highest AUC of 0.922 for control and bvFTD classifications. This was however not observed for control and AD differentiations. Classifications with GMD (0.940) and a GMD and DTI combination (0.941) resulted in similar AUC values ($P=0.41$).

Conclusion:

Combining functional and structural connectivity measures improve dementia-type differentiations and may contribute to more accurate and substantiated differential diagnosis of AD and bvFTD patients. Imaging protocols for differential diagnosis may benefit from also including DTI and rs-fMRI.

Key words: Alzheimer's disease, behavioral variant frontotemporal dementia, machine learning, classification, differential diagnosis, diffusion tensor imaging, fMRI

Introduction

Alzheimer's disease (AD) and behavioral variant frontotemporal dementia (bvFTD) are the most common causes of young onset dementia [1–3]. Accurate and confident differentiation between these disease types is crucial for the proper management, prognosis, and potential treatment of patients with dementia [4–6]. Yet, despite distinct clinical diagnostic criteria [1,7], heterogeneity and overlap of clinical manifestations often complicate differential diagnosis [8].

Complementary to clinically derived correlates, magnetic resonance imaging (MRI) has shown to reveal important disease-specific brain changes that may corroborate differential diagnosis on an individual basis [9]. Differences in the extent and location of gray matter (GM) and white matter (WM) atrophy have been recognized as important markers that distinguish each dementia-type [10–14]. Additionally, varying levels of diffusion tensor imaging (DTI)-derived measures elucidate regional differences in WM integrity impairment that are specific to AD or bvFTD patients [15–17]. It even has been suggested that these integrity impairments may precede or facilitate cortical degeneration [18]. However, despite these detectable group differences, MRI-based single subject classifications remain challenging. Per subject analyses are affected more by between subject variations than group-based analyses and it is still unclear which MRI measure contributes most to MRI-based dementia-type classifications [18–20]. Particularly in the earlier disease stages, accurate dementia-type classifications based on structural neuroimaging alone may be hampered by atrophy and tract specific deficit patterns that overlap or are hardly distinguishable [17,21,22].

Functional connectivity measures may inform on dementia state [23–25] or dementia progression [23–25] even well before clinical or structural differences can be detected [25,26]. Differences in functional connectivity strengths as measured with resting-state functional MRI (rs-fMRI) have been heralded as early and useful markers to differentiate AD and bvFTD patients [27,28]. However, the contribution of rs-fMRI measures as a (complementary) measure to differentiate dementia-types on an individual basis is still largely unclear [27,29,30].

It may be argued that one single MRI-derived measure may not sufficiently capture the complex pathophysiological processes that underlie dementia development. Multiparametric MRI-based statistical algorithms integrate various MRI measures to compute a single, quantitative probabilistic index and have shown to be more accurate in differentiating cognitively healthy from demented patients than single MRI measure-based algorithms alone [18,19,29,31,32]. We speculate that MRI-based classification algorithms that differentiate dementia-types benefit from integrating structural and functional connectivity measures. In this study we therefore aimed to determine the diagnostic accuracy of MRI-based classification algorithms to, on a subject basis, differentiate between AD and bvFTD patients, when combining measures derived from anatomical MRI, DTI, and rs-fMRI.

Methods

Participants

This two-center study involved a retrospective analysis of data previously published [18,28] and was conducted in accordance with regional research regulations and conformed to the Declaration of Helsinki. Local medical ethics committees of both centers approved the study and all patients gave written informed consent for their clinical and biological data to be used for research purposes.

For this study, we selected 37 patients with probable AD and 28 patients with bvFTD, who visited either the Alzheimer Center of the VU University Medical center (VUmc) (AD: n=22, bvFTD: n=19) or the Alzheimer Center of the Erasmus University Medical Center Rotterdam (EMC) (AD: n=15, bvFTD: n=9). In addition, we included 40 cognitively normal controls that were recruited by local newspaper advertisements (VUmc: n=22; EMC: n=18).

Patients underwent a standardized one-day assessment including medical and informant-based history, medical history (dementia, psychiatry, cardiovascular) of first-degree relatives, physical and neurological examination, blood tests, neuropsychological assessment, and brain MRI. 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 [7] and according to the clinical diagnostic criteria of FTD for bvFTD [1]. To minimize center effects, all diagnoses were re-evaluated in a panel that included clinicians from both centers. Controls 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 that included

medical history, physical examination, neuropsychological assessment, and brain MRI comparable to the patient work-up.

For both cohorts inclusion criteria were: (1) age between 40 and 80 years and (2) availability of a T₁-weighted 3-dimensional MRI (3DT_{1w}) scan, a diffusion tensor imaging (DTI) dataset, and a rs-fMRI T₂*-weighted scan. Exclusion criteria were: (1) large image artifacts (n=7); (2) failure of imaging analyzing software to process MR scans (n=10); and (3) gross brain pathology other than atrophy, including severe white matter hyperintensities and/or lacunar infarctions in deep gray matter structures.

MRI acquisition and review

Patients and controls of the VUmc cohort were scanned at the VUmc Amsterdam using a 3T MRI scanner (Signa HDxt, GE Healthcare, Milwaukee, WI, USA) with an 8-channel head coil with foam padding to restrict head motion. Patients and controls of the EMC cohort were scanned at the Leiden University Medical Center (LUMC) using a 3T MRI scanner (Achieva, Philips Medical Systems, Best, The Netherlands) with an 8-channel SENSE head coil. MRI sequence parameter settings are detailed in Supplementary Table 1. In brief, the imaging protocol included a whole-brain near-isotropic 3D T₁-weighted (3DT_{1w}) sequence for cortical and subcortical tissue-type segmentation, a diffusion tensor imaging (DTI) sequence for assessments of white matter integrity, and a resting state functional MRI T₂*-weighted MRI for the calculation of functional connectivity measures. Participants were instructed to lie still with their eyes closed and not to fall asleep during rs-fMRI. Additionally, a 3D fluid attenuated inversion recovery, dual-echo T₂-weighted, and susceptibility weighted imaging datasets were acquired to allow for review of brain pathology other than atrophy by an experienced radiologist.

MRI preprocessing

Preprocessing of 3DT_{1w} images involved non-uniformity correction [33] and segmentation of parenchymal tissue signal from surrounding tissue [34]. Images were then spatially aligned to the MNI152 2x2x2 mm T₁ template (Montreal Neurological Institute, Canada) using a non-linear registration procedure with a warp resolution of 10 mm [35]. Voxel-wise densities GM, WM, and cerebrospinal fluid (CSF) were determined with the initial steps of the voxel-based morphometry pipeline of the Statistical Parametric Mapping toolbox (SPM12; Functional Imaging Laboratory, University College London, London, UK)[36] implemented in MATLAB R2015b (MathWorks, Natick, MA). Except for manual placement of the image's origin approximately on the anterior commissure and applying the light cleanup option to remove any remaining non-brain tissue, default settings were used for tissue-type segmentations. Deep gray matter (DGM) structures including the bilateral thalamus, caudate nucleus, putamen, globus pallidus, nucleus accumbens, amygdala, and left and right hippocampus were separately identified using a dedicated registration and segmentation procedure with default settings [37].

For DTI, preprocessing included motion and eddy-current induced distortion correction and gradient vector direction correction [38]. The corrected DTIs were subsequently used to voxel-wise calculate measures of fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (A_{\parallel} ; largest eigenvalue), and radial diffusivity (A_{\perp} ; average of the two eigenvalues λ_2 and λ_3) using weighted least square fitting [39]. The tract-based spatial statistics (TBSS) pipeline was used to create a study-specific TBSS-skeleton, by non-linearly aligning all FA maps to the

FMRI58_FA template [40]. The derived skeleton was thresholded at 0.2 to ensure values originated from WM tissue (see below).

For rs-fMRI, preprocessing included motion correction [41], brain extraction, spatial smoothing using a Gaussian kernel with a full width half maximum of 3 mm, grand mean intensity normalization, motion artifact removal, and high-pass temporal filtering (cutoff frequency = 0.01Hz). Motion artifacts were removed from the data using the ICA-based automatic removal of motion artifacts (ICA-AROMA (vs0.3-beta)) procedure [42]. Subsequently, the rs-fMRI volumes were linearly aligned to the corresponding 3DT_{1w} [43]. Spatial alignments to the MNI152 template were achieved by concatenating the registration parameters of the previous step with the nonlinear parameters from the 3DT_{1w} to the MNI152 template.

All registration and segmentation steps were critically reviewed and errors were corrected accordingly.

Feature extraction

Two anatomical atlases were used to parcellate the entire brain. These atlases were used to extract, in each subject's native space, cortical and subcortical GM and WM features from the structural 3DT_{1w}-images. The 48 cortical regions of the Harvard-Oxford (HO) probabilistic anatomical brain atlas were split into left and right hemisphere regions, resulting in 96 distinct cortical regions that covered the complete cortical GM [44]. For WM regional analysis, we selected 20 WM tract regions from the probabilistic Johns-Hopkins-University (JHU) white-matter tractography atlas [45]. Voxel probability values less than 25% were excluded. The remaining values

were used to weight the WM or GM densities. Weighted GM densities were calculated by weighting the GM segmentations by the probability of a voxel being part of that specific HO brain atlas-derived region. Weighted WM densities were calculated by weighting the WM segmentations by the probability of a voxel being part of that specific JHU brain atlas-derived tract. This way we emphasized the regions' likelihood of being GM or WM without introducing bias resulting from conservatively selecting brain regions. For DGM regions, the dedicated segmentations of DGM, hippocampus, and amygdala were used to calculate the regional volumes normalized by intracranial volume to compensate for individual differences in brain volume [46]. This resulted in a feature vector of 110 GMD values (96 cortical GMD values and 14 DGM volumes), and a feature vector of 20 average WMD values per subject.

DTI-based features were calculated by, on a voxel-wise basis, projecting each subject's FA, MD, AxD, or RD values on to the TBSS group skeleton. Analogous to the anatomical WMD features, the 20 WM tracts of the probabilistic JHU white-matter tractography atlas were used to calculate a weighted mean FA, MD, AxD, and RD per tract per subject. This resulted in 4*20 feature vectors of mean FA, MD, AxD, and RD values per subject.

The functional connectivity features were calculated by combining all processed rs-fMRI datasets in a temporally concatenated independent component analysis (ICA) [47], with dimensionality fixed at 70 components and an ICA threshold of 0.99 [48]. This meant that each voxel included in the ICA map was 99 times more likely to be part of that component than to be caused by Gaussian background noise. For each

subject, we calculated the mean time course for each component, weighted by the ICA weight map and GM probability of that component's region. These mean time courses were subsequently used to determine the functional connectivity of a component with the 69 other components. Functional connectivity was either expressed as full correlations (corr) or as sparse L1-regularized partial correlations (pcor) between the components' time courses. Partial correlations were calculated using the graphical lasso algorithm [49]. Both functional connectivity measures resulted in two feature vectors of each $(70*69)/2=2415$ (partial) correlations per subject.

Differential classification

These feature vectors were subsequently used to train an elastic net regression algorithm with default settings to provide a statistically sound solution for the imbalance between a large number of features and a small number of subjects. An elastic net regression model estimates a sparse regression model that selects a subset of all the features provided as input by imposing feature selection and feature weight penalties during regression, effectively selecting only those features relevant for classification [50–52]. A cross validation procedure was used to determine the optimal set of penalty parameters and generalized classification performance of the elastic net classifiers. Cross-validation reduces classification bias by iteratively subdividing the data in separate test and training sets. In this study we used two nested, 10-fold cross validation loops. The first, outer, loop was used to determine the overall classification performance, while the second, inner, loop subdivided the training set further to determine the lowest binomial deviance that corresponded with the best operational parameters for the penalty terms (including the optimal number of

features) without overestimating classification performance [53,54]. This process was repeated 10 times to ascertain that each subject was part of the test set exactly once. To ensure that estimated feature regression coefficients were conditional on subject age and gender and to adjust for scanner effects, age, gender, and center were included into the model without any penalty. For AD and bvFTD differentiations, models were furthermore made conditional on disease severity by also including disease duration (in months). The entire classification procedure was repeated 50 times to reduce variance resulting from random partitioning in training and test folds, and to report the range of observed outcomes under different train and test conditions.

We trained three types of classification models following the above-described procedures. One model aimed to differentiate AD from bvFTD patients. To contextualize the classification performance values of this model, we also trained two additional models. A second model aimed to differentiate control subjects from bvFTD patients, and a third model to differentiate control subjects from AD patients. For each subject, these models produced a predicted value between 0 and 1. The higher the predicted value the more likely this subject belongs to the bvFTD- group or AD group in case of the control versus AD model.

Classification performance

Classification performances were quantified using a threshold-independent measure based on receiver-operating characteristic (ROC) analyses. After each 10-fold nested cross-validation, the predictions were used to calculate ROC curves by continuously increasing a threshold between 0 and 1 and for each threshold classifying a subject either as control, AD, or bvFTD patient, depending on the model being evaluated.

These classifications were then compared with the subject's actual differential diagnosis. The area under this ROC curve (AUC) was calculated as a measure of classification performance insensitive to the distribution of each patient group [55]. Additionally, we calculated the optimal operating point on each curve to calculate the model's accuracy, sensitivity, and specificity given equal class distribution and equal penalty for false positive and false negative predictions. Reported AUC, accuracy, sensitivity, and specificity values are averages obtained from repeating the cross-validation 50 times.

Multiparametric classifications

To determine whether combining multiple MRI measures improved classification, we first assessed classification performance for each single measure separately by alternately providing all features of that specific measure (i.e. 110 GMD, 20 WMD, 20 FA, 20 MD, 20 AxD, 20 RD, 2514 corr, or 2514 pcor features) as input for cross-validation. Subsequently, we step-wise concatenated measures each time adding all features of a new measure to the best performing measure combination (i.e. highest AUC) of the previous step till all measures and thus all features were included into the model. The model subsequently determined the importance of each feature by assigning the proper weight.

All classification analyses and evaluations were implemented in R (R core 2010, GLMnet package) [51].

Statistical analysis

Statistical analyses of between-group differences were performed using SPSS (IBM SPSS Versions 22.0, IBM, Armonk, NY). Demographic group differences between age, MMSE, and disease duration were assessed using analysis of variance (ANOVA). Sex and center distributions were assessed with χ^2 tests. Permutation testing was used to determine whether single or multiparametric models performed above-chance level (one-tailed, N=5000). The maximum statistic method was used for family-wise error correction [56]. A bootstrap percentile method was used to compare ROC curves of single measures (two-tailed, N=5000), and between single measures and the best performing multiparametric combination (one-tailed, N=5000) [57,58]. False discovery rate correction within each patient-group was used to correct for multiple comparisons. $P < 0.05$ was considered statistically different.

Results

Demographics

For this study, 88 subjects met the inclusion criteria. Thirty AD subjects were diagnosed with probable AD, 23 patients with bvFTD. Thirty-five control subjects were included for classification performance evaluation of the classification models involving control subjects. MMSE values of control subjects were higher than those of patients. MMSE values of AD patients were comparable to those of bvFTD patients (Table 1).

Classification performance: AD versus bvFTD

For AD and bvFTD differentiations, combining multiple MRI measures within a multiparametric classification model resulted in higher mean AUC values than those with models that were based on single MRI-based measures only (Figure 1; Table 2).

The level of AUC values was however conditional on the combination being considered. After step-wise adding multiple MRI-derived measures to the best performing single MRI measure model (i.e. MD: 0.708 [0.625-0.775] (mean AUC [min-max]); Table 2 underlined), highest AUC, sensitivity, and specificity values were obtained for a classification model that included MD, full correlations, and FA (0.811 [0.755 – 0.862]; Table 2: bold; Table 3: bold). When comparing the single measure-based classifications, AUC values of the single measure-based classifications were not significantly different. However, AUC values obtained with the multiparametric model were significantly higher compared with those obtained with single MRI measure-based models (Table 3), except for MD-based classifications (P=0.06).

Classification performance: control versus dementia

MRI measures that resulted in highest classification performance rates for control versus dementia classifications (i.e. control versus bvFTD or control versus AD) were different from those obtained with AD versus bvFTD classifications. For control and bvFTD classifications highest AUC, sensitivity, and specificity values were found for a multiparametric model that included measurements of FA, GMD, and full correlations (0.922 [0.877-0.954]; Table 4: bold; Supplementary Table 2: bold). AUC values of this multiparametric model were significantly higher than those obtained with single MRI measure-based models, except for FA- (0.862 [0.810-0.903], P=0.08) and GMD-based (0.858 [0.827-0.896], P=0.11) classifications (Table 4). For classifications between control subjects and AD patients on the other hand, classification performance values of a combination of GMD and DTI measures (0.941 [0.910-0.966]) were not better than classifications based on GM measurements only

(0.940 [0.913-0.962] P=0.41) (Table 5, Supplementary Table 3). While the combination outperformed the other single measure-based classifications, GMD-based classifications were also better than all other single MRI measure classification except for RD-based classifications (0.832 [0.784-0.865], P=0.10).

Features selected

MRI-derived measures outweighed the contribution of gender, age, and center to the classification. AD and bvFTD differentiations using gender, age, and center distributions only did not outperform random chance (AUC=0.596, P=0.12). AUC values of the best performing AD versus bvFTD model (i.e. combination of MD, full correlations, and FA) were similar to classifications where age (P=0.89), gender (P=0.70), or center (P=0.14) was excluded as a covariate. Identical AUC values were also obtained when these covariates were excluded from the best performing control versus bvFTD model (i.e. combination of FA, GMD, and full correlations: age P=0.10, gender P=0.65, center P=0.50) or control versus AD model (i.e. combination of GMD, AxD, FA, RD, MD: age P=0.60, gender P=0.64, center P=0.79).

Nested cross-validation resulted in models that selected a subset of all the MRI features provided as input for classification. The best performing AD versus bvFTD model selected 5% of all the MD, full correlations, and FA features provided for classification. This model focused on differences in uncinate fasciculus, forceps minor, cingulum bundle, cortical spinal tracts, and altered functional connectivity with the dorsal default mode-network. In contrast, the control versus dementia models also included GM regions for classification. The best performing control versus bvFTD model selected 3% of all provided features and considered anterior thalamic

radiation, cortical spinal tract, inferior longitudinal fasciculus, and hippocampal regions as key regions for classification. The best performing control versus AD model selected 37% of all the features provided and primarily focused on volume differences in DGM structures including the hippocampus, nucleus accumbens, pallidum, and thalamus.

Discussion

This study investigated the diagnostic accuracy of MRI-based classification algorithms to - on a subject basis – differentiate between AD and bvFTD patients using anatomical MRI, DTI, and rs-fMRI. Our study showed that, classification algorithms that combine measures of DTI and rs-fMRI are more effective in discriminating AD from bvFTD patients than algorithms that use DTI, GMD, or rs-fMRI measures separately. Furthermore, when compared with classification models that aim to differentiate patients from controls we found that the level of improvement was conditional on MRI measure and dementia-type differentiation being considered.

GM- and DTI-derived measures have been heralded as important corroborating measures for differentiating AD and bvFTD patients [15,18,19,31]. Classification methods that combined these measures invariably showed improvement of classification accuracy over models that used these measures separately [15,18,31]. We also found that combining measures improves classification accuracy. However, we found highest classification rates when DTI and rs-fMRI measures were combined. GM-based values contributed - either as a single (for AD) or complementary measure (for bvFTD) - to the differentiation of dementia versus control subjects, but not to the discrimination of AD and bvFTD patients. Differential

diagnosis between dementias may be challenged by overlapping GM atrophy patterns particularly in subcortical regions [21,22]. Both best performing dementia versus control models considered volume reductions in subcortical regions like the hippocampus as decisive regions for classification and may therefore have had little contribution in the differentiation between the dementia-types. Distinct microstructural changes in WM may be potential markers for bvFTD. Compared with AD patients, bvFTD patients show distinct differences in diffusion especially in key regions like uncinate fasciculus, genu of the corpus callosum, and cingulum bundle [16,17,59]; differences that may even precede cortical degeneration [26]. In line with these findings, we found that the best performing bvFTD classification models specifically focused on these regions of DTI-derived WM pathology. We also found that AD versus bvFTD classifications improved when rs-fMRI measures were added. Altered connectivity measurements mainly with the dorsal default mode network regions were considered important and in the end increased classification performance rates. This corresponds with other work that showed that connectivity deviations of the default mode network alongside salience / executive control network contrast between AD and bvFTD patients [12,27,28,60]. Nevertheless, the role of rs-fMRI as a separate measure for disease-type classifications has not yet been established. It has found limited application for single subject disease-type classifications [27,29,30,32]. In AD versus control classifications its performance rates are consistently among the lowest compared with other MRI measures [29,30,32] and we also found that rs-fMRI as a single measure was neither effective in differentiating dementia types nor between patients and controls subjects.

We first trained classifications models with varying combinations of MRI measures to

differentiate AD from bvFTD patients. Subsequently, we trained models that aimed to differentiate AD or bvFTD patients from cognitively healthy controls in order to assess classification model accuracy under conditions where large MRI-based differences are expected [10,14,15,17,18,28]. Combining MRI-derived structural and functional measures improves bvFTD classifications. The level of improvement is however dependent on the dementia-types and MRI measures being considered. Substantial improvement was for instance not observed for AD versus control classifications while others did [29,32] or did not show benefit [30] from GM, DTI, and rs-fMRI measures. Furthermore, in line with previous work, we observed that concatenating all measures did not result in the most accurate classification model and as more measures were added the level of improvement became less apparent [29,31,61]. The marginal improvements observed for some measures or combinations may indicate a negligible contribution to the model. These may result from variability due to non-stationarities like correlation across features and samples within our data. [62] These measures are most likely of clinically insufficient added relevance, and may not justify the additional processing time.

Overall, classification performance values for dementia-type differentiations were consistently lower than dementia versus control subject differentiations. Dementia type classifications remain challenging. Compared with control versus patient comparisons, differences between patients are more subtle and, particularly in the earlier stages of the disease, may be hindered by overlapping patterns of GM atrophy [21,22,63], WM integrity impairment [17], or comparable patterns of functional connectivity loss in specific functional network regions [12,28]. It remains questionable whether any of these algorithms can fully capture the complexity of

structural and functional dynamics of neurodegenerative processes underlying dementia. We speculate that other algorithms that utilize more sophisticated feature combination approaches, like sparse group lasso models [61], or hierarchical or longitudinal algorithms that aim to differentiate patients from a general population in order to subsequently differentiate between dementia-types may further exploit and weigh the additional information from multiple measures [64]. Incorporating other or additional imaging-derived biomarkers as cerebral blood flow [65], amplitude of low frequency fluctuations [32], GM derived connectomics [19], or diffusion tractography derived graph-based analytics [61] may further contribute to MRI-based dementia-type classification estimates without increasing diagnostic complexity.

In our analysis, we took several steps to reduce classification bias and augment generalizability of our results. First, while specific functional network connections were found to deviate between AD and bvFTD patients [12,27,28,60], we hypothesized that restricting our analysis to specific network regions may introduce bias and may unnecessarily exclude other regions that show deficits at different stages of the disease [23,24,52,66]. The number of network regions that need to be distinguished to optimally differentiate between dementia-types still needs clarification. ICA dimensionality continues to be an actively debated topic and is a trade-off between detail in functionally connected regions and feature space dimensionality. Second, we employed a repeated nested cross validation approach in which two levels of data segregation ensured unbiased operational parameter optimization and classification performance estimation [53]. Thirdly, for classification, we used regularized regression to establish homogeneous and stable dementia type estimates and to accommodate proper selection of relevant features

despite high dimensionality and collinearity of our data [52]. Our study was limited by the clinical diagnoses used to validate the differentiations. While our multidisciplinary, multicenter team carefully diagnosed each patient according to the newest criteria for AD [7] or bvFTD [1], uncertainty in the diagnosis remains. Overlapping clinical symptoms may complicate dementia-type differentiations [8] and postmortem pathological data to confirm the diagnosis were unavailable. Furthermore, we included patients with possible and probable FTD diagnosis to maximize our patient cohort. This may have increased the complexity of our classifications. We included relatively young patients who were diagnosed in a relatively early stage of the disease, therefore likely to have less apparent structural or functional deficits. Nevertheless, we were able to differentiate bvFTD patients from controls with high accuracy. Further validations with larger, multicenter cohorts are necessary to contextualize and compare our findings. The trained models depend on both random and non-random class differences in the training sample and especially in light of our limited population sizes, we cannot reliably differentiate between real and random class differences in the trained models [62]. Consequently, we refrained from biological interpretation of the model's parameters, speculating on the exact measure order, and cautiously interpreted the features selected for classification.

In conclusion, in this study we investigated whether multiparametric MRI improves MRI-based differential diagnosis of dementia-types with overlapping clinical symptoms. We found that combining multiple MRI-derived measures of structural and functional connectivity improve MRI-based differentiations between bvFTD and AD patients. Our results imply that current MRI protocols for differential diagnosis of AD or bvFTD may benefit from adding functional and diffusion connectivity

measures complementary to the anatomical (GM-based) measurements already being acquired. Yet, the MRI measures and dementia-types being differentiated should be carefully considered to attain most optimal result. These results furthermore highlight the potential of these multiparametric imaging-based classification algorithms to aid in and possibly improve diagnosis, particularly in situations where experienced neuroradiologists or other supporting diagnostic measures are available to a limited extent. Further analysis should reveal how these results generalize to larger cohorts, to cases where patients' symptoms are less conspicuous, or involve subjects in a pre-symptomatic disease stage. Nevertheless, these observations may serve as a first guidance towards integrating quantitative imaging-based measures that may contribute to more confident per subject differential diagnoses and subsequent tailored patient care.

Acknowledgements

Gradient non-linearity corrections were kindly provided by GE medical systems, Milwaukee.

Disclosure statement

This study was supported by VICI grant no. 016.130.677 of the Netherlands Organization for Scientific Research (I). The Vumc Alzheimer Center is supported by Alzheimer Nederland and Stichting VUMC Fonds. Research of the Vumc Alzheimer Center is part of the neurodegeneration research program of the Neuroscience Campus Amsterdam.

Dr. Mark Bouts reports no disclosures.

Dr. Christiane Möller is appointed on a grant from the national project 'Brain and Cognition' ("Functionele Markers voor Cognitieve Stoornissen" (# 056-13-014)) by the Netherlands Initiative Brain and Cognition (NIHC).

Dr. Anne Hafkemeijer is appointed on a grant from the national project 'Brain and Cognition' ("Functionele Markers voor Cognitieve Stoornissen" (# 056-13-010)) by the Netherlands Initiative Brain and Cognition (NIHC).

Prof. Dr. John van Swieten reports no disclosures.

Dr. Elise Dopper reports no disclosures.

Prof. Dr. Wiesje van der Flier is recipient of the Alzheimer Nederland grant (Influence of age on the endophenotype of AD on MRI, # 2010-002).

Dr. Hugo Vrenken has received research support from MerckSerono, Novartis, and Pfizer, and speaker honoraria from Novartis; all funds were paid to his institution.

Dr. Alle Meije Wink reports no disclosures.

Dr. Yolande Pijnenburg reports no disclosures.

Prof. Dr. Philip Scheltens serves/has served on the advisory boards of: Genentech, Novartis, Roche, Danone, Nutricia, Baxter and Lundbeck. He has been a speaker at symposia organized by Lundbeck, Merz, Danone, Novartis, Roche and Genentech. He serves on the editorial board of Alzheimer's Research & Therapy and Alzheimer's Disease and Associated Disorders, is a member of the scientific advisory board of the EU Joint Programming Initiative and the French National Plan Alzheimer. For all his activities he receives no personal compensation.

Prof. Dr. Frederik Barkhof serves/has served on the advisory boards of: Bayer-Schering Pharma, Sanofi-Aventis, Biogen Idec, UCB, Merck-Serono, Novartis and Roche. He received funding from the Dutch MS Society and has been a speaker at symposia organized by the Serono Symposia Foundation. He was supported by the

NIHR biomedical research centre at UCLH. For all his activities he receives no personal compensation.

Mr. Tijn Schouten reports no disclosures.

Mr. Frank de Vos reports no disclosures.

Mr Rogier Feis reports no disclosures.

Dr. Jeroen van der Grond reports no disclosures.

Prof. Dr. Mark de Rooij reports no disclosures.

Prof. Dr. Serge Rombouts is supported by The Netherlands Organization for Scientific Research (NWO), Vici project nr. 016.130.677.

References

- [1] Rascovsky K, Hodges JR, Knopman D, Mendez MF, Kramer JH, Neuhaus J, Van Swieten JC, Seelaar H, Dopper EGP, Onyike CU, Hillis AE, Josephs KA, Boeve BF, Kertesz A, Seeley WW, Rankin KP, Johnson JK, Gorno-Tempini ML, Rosen H, Prioleau-Latham CE, Lee A, Kipps CM, Lillo P, Piguet O, Rohrer JD, Rossor MN, Warren JD, Fox NC, Galasko D, Salmon DP, Black SE, Mesulam M, Weintraub S, Dickerson BC, Diehl-Schmid J, Pasquier F, Deramecourt V, Lebert F, Pijnenburg Y, Chow TW, Manes F, Grafman J, Cappa SF, Freedman M, Grossman M, Miller BL (2011) Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* **134**, 2456–2477.
- [2] Ratnavalli E, Brayne C, Dawson K, Hodges JR (2002) The prevalence of frontotemporal dementia. *Neurology* **58**, 1615–1621.
- [3] Harvey RJ, Skelton-Robinson M, Rossor MN (2003) The prevalence and

- causes of dementia in people under the age of 65 years. *J. Neurol. Neurosurg. Psychiatry* **74**, 1206–1209.
- [4] Mohs RC, Doody RS, Morris JC, Ieni JR, Rogers SL, Perdomo CA, Pratt RD (2001) A 1-year, placebo-controlled preservation of function survival study of donepezil in AD patients. *Neurology* **57**, 481–8.
- [5] Mendez MF, Shapira JS, McMurtray A, Licht E (2007) Preliminary findings: behavioral worsening on donepezil in patients with frontotemporal dementia. *Am. J. Geriatr. Psychiatry* **15**, 84–87.
- [6] Pressman PS, Miller BL (2014) Diagnosis and management of behavioral variant frontotemporal dementia. *Biol. Psychiatry* **75**, 574–581.
- [7] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH (2011) 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 Dement.* **7**, 263–269.
- [8] Mathias JL, Morphet K (2010) Neurobehavioral differences between Alzheimer’s disease and frontotemporal dementia: A meta-analysis. *J. Clin. Exp. Neuropsychol.* **32**, 682–698.
- [9] Risacher SL, Saykin AJ (2013) Neuroimaging biomarkers of neurodegenerative diseases and dementia. *Semin. Neurol.* **33**, 386–416.
- [10] Rabinovici GD, Seeley WW, Kim EJ, Gorno-Tempini ML, Rascovsky K, Pagliaro T a, Allison SC, Halabi C, Kramer JH, Johnson JK, Weiner MW, Forman MS, Trojanowski JQ, Dearmond SJ, Miller BL, Rosen HJ (2008) Distinct MRI atrophy patterns in autopsy-proven Alzheimer’s disease and

- frontotemporal lobar degeneration. *Am. J. Alzheimers. Dis. Other Demen.* **22**, 474–488.
- [11] Whitwell JL, Jack CR (2005) Comparisons between Alzheimer disease, frontotemporal lobar degeneration, and normal aging with brain mapping. *Top. Magn. Reson. Imaging* **16**, 409–425.
- [12] Filippi M, Agosta F, Scola E, Canu E, Magnani G, Marcone A, Valsasina P, Caso F, Copetti M, Comi G, Cappa SF, Falini A (2013) Functional network connectivity in the behavioral variant of frontotemporal dementia. *Cortex* **49**, 2389–2401.
- [13] Chao LL, Schuff N, Clevenger EM, Mueller SG, Rosen HJ, Gorno-Tempini ML, Kramer JH, Miller BL, Weiner MW (2007) Patterns of white matter atrophy in frontotemporal lobar degeneration. *Arch Neurol* **64**, 1619–1624.
- [14] Frings L, Yew B, Flanagan E, Lam BYK, Hüll M, Huppertz HJ, Hodges JR, Hornberger M (2014) Longitudinal grey and white matter changes in frontotemporal dementia and Alzheimer’s disease. *PLoS One* **9**, 1–8.
- [15] McMillan CT, Brun C, Siddiqui S, Churgin M, Libon D, Yushkevich P, Zhang H, Boller A, Gee J, Grossman M (2012) White matter imaging contributes to the multimodal diagnosis of frontotemporal lobar degeneration. *Neurology* **78**, 1761–1768.
- [16] Zhang Y, Schuff N, Du A-T, Rosen HJ, Kramer JH, Gorno-Tempini ML, Miller BL, Weiner MW (2009) White matter damage in frontotemporal dementia and Alzheimer’s disease measured by diffusion MRI. *Brain* **132**, 2579–92.
- [17] Mahoney CJ, Ridgway GR, Malone IB, Downey LE, Beck J, Kinnunen KM, Schmitz N, Golden HL, Rohrer JD, Schott JM, Rossor MN, Ourselin S, Mead

- S, Fox NC, Warren JD (2014) Profiles of white matter tract pathology in frontotemporal dementia. *Hum. Brain Mapp.* **35**, 4163–4179.
- [18] Möller C, Hafkemeijer A, Pijnenburg YAL, Rombouts SAR, Van Der Grond J, Dopfer E, van Swieten J, Versteeg A, Pouwels PJW, Barkhof F, Scheltens P, Vrenken H, Van Der Flier WM (2015) Joint assessment of white matter integrity, cortical and subcortical atrophy to distinguish AD from behavioral variant FTD: A two-center study. *NeuroImage Clin.* **9**, 418–429.
- [19] McMillan CT, Avants BB, Cook P, Ungar L, Trojanowski JQ, Grossman M (2014) The power of neuroimaging biomarkers for screening frontotemporal dementia. *Hum. Brain Mapp.* **35**, 4827–4840.
- [20] Arbabshirani MR, Plis S, Sui J, Calhoun VD (2016) Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage* **145**, 137–165.
- [21] Barnes J, Whitwell JL, Frost C, Josephs KA, Rossor M, Fox NC (2006) Measurements of the amygdala and hippocampus in pathologically confirmed Alzheimer disease and frontotemporal lobar degeneration. *Arch. Neurol.* **63**, 1434–9.
- [22] van de Pol L, Hensel A, van der Flier W, Visser P-J, YAL P, Barkhof F, Josef Gertz H, Scheltens P (2006) Hippocampal atrophy on MRI in frontotemporal lobar degeneration and Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatry* **77**, 439–442.
- [23] Hafkemeijer A, Möller C, Dopfer EGP, Jiskoot LC, van den Berg-Huysmans AA, van Swieten JC, van der Flier WM, Vrenken H, Pijnenburg YAL, Barkhof F, Scheltens P, van der Grond J, Rombouts SAR (2016) A Longitudinal Study on Resting State Functional Connectivity in Behavioral Variant

- Frontotemporal Dementia and Alzheimer's Disease. *J. Alzheimer's Dis.* **55**, 1–17.
- [24] Damoiseaux JS, Prater KE, Miller BL, Greicius MD (2012) Functional connectivity tracks clinical deterioration in Alzheimer's disease. *Neurobiol. Aging* **33**, 828.e19-828.e30.
- [25] Whitwell JL, Josephs KA, Avula R, Tosakulwong N, Weigand SD, Senjem ML, Vemuri P, Jones DT, Gunter JL, Baker M, Wszolek ZK, Knopman DS, Rademakers R, Petersen RC, Boeve BF, Jack CR (2011) Altered functional connectivity in asymptomatic MAPT subjects A comparison to bvFTD. *Neurology* **77**, 866–874.
- [26] Dopper EGP, Rombouts SARB, Jiskoot LC, Heijer T Den, De Graaf JRA, De Koning I, Hammerschlag AR, Seelaar H, Seeley WW, Veer IM, Van Buchem MA, Rizzu P, Van Swieten JC (2014) Structural and functional brain connectivity in presymptomatic familial frontotemporal dementia. *Neurology* **83**, e19-26.
- [27] Zhou J, Greicius MD, Gennatas ED, Growdon ME, Jang JY, Rabinovici GD, Kramer JH, Weiner M, Miller BL, Seeley WW (2010) Divergent network connectivity changes in behavioural variant frontotemporal dementia and Alzheimer's disease. *Brain* **133**, 1352–1367.
- [28] Hafkemeijer A, Möller C, Dopper EGP, Jiskoot LC, Schouten TM, van Swieten JC, van der Flier WM, Vrenken H, Pijnenburg YAL, Barkhof F, Scheltens P, van der Grond J, Rombouts SARB (2015) Resting state functional connectivity differences between behavioral variant frontotemporal dementia and Alzheimer's disease. *Front. Hum. Neurosci.* **9**, 474.
- [29] Schouten TM, Loitfelder M, de Vos F, Seiler S, van der Grond J, Lechner A,

- Hafkemeijer A, Möller C, Schmidt R, de Rooij M, Rombouts SARB (2016) Combining anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate Alzheimer's disease. *NeuroImage Clin.* **11**, 46–51.
- [30] Dyrba M, Grothe M, Kirste T, Teipel SJ (2015) Multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple kernel SVM. *Hum. Brain Mapp.* **36**, 2118–2131.
- [31] Canu E, Agosta F, Mandic-Stojmenovic G, Stojković T, Stefanova E, Inuggi A, Imperiale F, Copetti M, Kostic VS, Filippi M (2017) Multiparametric MRI to distinguish early onset Alzheimer's disease and behavioural variant of frontotemporal dementia. *NeuroImage Clin.* **15**, 428–438.
- [32] Dai Z, Yan C, Wang Z, Wang J, Xia M, Li K, He Y (2012) Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3). *Neuroimage* **59**, 2187–2195.
- [33] Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC (2010) N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320.
- [34] Smith SM (2002) Fast robust automated brain extraction. *Hum. Brain Mapp.* **17**, 143–155.
- [35] Andersson JLR, Jenkinson M, Smith S (2007) *Non-linear registration, aka spatial normalisation. FMRIB Technial Report TR07JA2.*
- [36] Friston KJ, Ashburner J, Kiebel S, Nichols T, Penny WD (2007) *Statistical parametric mapping : the analysis of funtional brain images*, Elsevier/Academic Press.
- [37] Patenaude B, Smith SM, Kennedy DN, Jenkinson M (2011) A Bayesian model

- of shape and appearance for subcortical brain segmentation. *Neuroimage* **56**, 907–922.
- [38] Leemans A, Jones DK (2009) The B-matrix must be rotated when correcting for subject motion in DTI data. *Magn. Reson. Med.* **61**, 1336–1349.
- [39] Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM (2004) Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23 Suppl 1**, S208-219.
- [40] Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM, Behrens TEJ (2006) Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* **31**, 1487–1505.
- [41] Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841.
- [42] Pruim RHR, Mennes M, van Rooij D, Llera A, Buitelaar JK, Beckmann CF (2015) ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage* **112**, 267–277.
- [43] Greve DN, Fischl B (2009) Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* **48**, 63–72.
- [44] Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–

- 980.
- [45] Hua K, Zhang J, Wakana S, Jiang H, Li X, Reich DS, Calabresi PA, Pekar JJ, van Zijl PCM, Mori S (2008) Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. *Neuroimage* **39**, 336–347.
- [46] de Vos F, Schouten TM, Hafkemeijer A, Dopfer EGP, van Swieten JC, de Rooij M, van der Grond J, Rombouts SARB (2016) Combining multiple anatomical MRI measures improves Alzheimer’s disease classification. *Hum. Brain Mapp.* **37**, 1920–1929.
- [47] Beckmann CF, Smith SM (2004) Probabilistic Independent Component Analysis for Functional Magnetic Resonance Imaging. *IEEE Trans. Med. Imaging* **23**, 137–152.
- [48] Smith SM, Vidaurre D, Beckmann CF, Glasser MF, Jenkinson M, Miller KL, Nichols TE, Robinson EC, Salimi-Khorshidi G, Woolrich MW, Barch DM, Ugurbil K, Van Essen DC (2013) Functional connectomics from resting-state fMRI. *Trends Cogn. Sci.* **17**, 666–682.
- [49] Friedman J, Hastie T, Tibshirani R (2007) Sparse inverse covariance estimation with the lasso. *Biostatistics* **9**, 432–441.
- [50] Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B (Statistical Methodol.)* **67**, 301–320.
- [51] Friedman J, Hastie T, Tibshirani R (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33**, 1–22.
- [52] Teipel SJ, Grothe MJ, Metzger CD, Grimmer T, Sorg C, Ewers M, Franzmeier N, Meisenzahl E, Klöppel S, Borchardt V, Walter M, Dyrba M (2017) Robust Detection of Impaired Resting State Functional Connectivity Networks in

- Alzheimer's Disease Using Elastic Net Regularized Regression. *Front. Aging Neurosci.* **8**, 318.
- [53] Varma S, Simon R (2006) Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* **7**, 91.
- [54] Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* **12**, 535–540.
- [55] Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861–874.
- [56] Winkler AM, Ridgway GR, Douaud G, Nichols TE, Smith SM (2016) Faster permutation inference in brain imaging. *Neuroimage* **141**, 502–516.
- [57] Hanley JA, McNeil BJ (1983) A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* **148**, 839–43.
- [58] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J, Müller M (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 1–8.
- [59] Daianu M, Mendez MF, Baboyan VG, Jin Y, Melrose RJ, Jimenez EE, Thompson PM (2016) An advanced white matter tract analysis in frontotemporal dementia and early-onset Alzheimer's disease. *Brain Imaging Behav.* **10**, 1038–1053.
- [60] Farb NAS, Grady CL, Strother S, Tang-Wai DF, Masellis M, Black S, Freedman M, Pollock BG, Campbell KL, Hasher L, Chow TW (2013) Abnormal network connectivity in frontotemporal dementia: Evidence for prefrontal isolation. *Cortex* **49**, 1856–1873.

- [61] Schouten TM, Koini M, Vos F de, Seiler S, Rooij M de, Lechner A, Schmidt R, Heuvel M van den, Grond J van der, Rombouts SARB (2017) Individual Classification of Alzheimer's Disease with Diffusion Magnetic Resonance Imaging. *Neuroimage* **152**, 476–481.
- [62] Varoquaux G (2017) Cross-validation failure: Small sample sizes lead to large error bars. *Neuroimage* 1–10.
- [63] Johnson JK, Head E, Kim R, Starr A, Cotman CW (1999) Clinical and pathological evidence for a frontal variant of Alzheimer disease. *Arch. Neurol.* **56**, 1233–1239.
- [64] Marquand AF, Wolfers T, Mennes M, Buitelaar J, Beckmann CF (2016) Beyond Lumping and Splitting: A Review of Computational Approaches for Stratifying Psychiatric Disorders. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **1**, 433–447.
- [65] Bron EE, Smits M, Papma JM, Steketee RME, Meijboom R, de Groot M, van Swieten JC, Niessen WJ, Klein S (2016) Multiparametric computer-aided differential diagnosis of Alzheimer's disease and frontotemporal dementia using structural and advanced MRI. *Eur. Radiol.* 1–11.
- [66] Brier MR, Thomas JB, Snyder AZ, Benzinger TL, Zhang D, Raichle ME, Holtzman DM, Morris JC, Ances BM (2012) Loss of intranetwork and internetwork resting state functional connections with Alzheimer's disease progression. *J Neurosci* **32**, 8890–8899.

Tables

Table 1. Demographics.

	Control	AD	bvFTD
N	35	30	23
Age	60.8±6.1	66.9±7.8*	63.5±7.6
Sex (female, n, %)	16 (45.7%)	12 (40.0%)	6 (26.1%)
MMSE	28.8±1.5	23.4±3.1*	24.5±3.2*
Disease duration (months)	0.0±0.0	35.2±24.9*	52.9±52.1*
Center (VUMC, n, %)	21 (60.0%)	20 (66.7%)	16 (69.6%)

*versus control subjects, $p < 0.01$

Mean ± standard deviation or n (%). AD: Alzheimer's disease, bvFTD: behavioral variant frontotemporal dementia, MMSE: mini-mental state examination.

Table 2. Mean AUC values for AD versus bvFTD classifications using single or combination of multiple MRI-derived measures

Combined with:	MRI Measure							
	MD	corr	FA	GMD	AxD	pcor	RD	WMD
-	<u>0.708</u>	0.674	0.623	0.614	0.636	0.561	0.652	0.614
MD		0.803	0.707	0.637	0.656	0.631	0.665	0.712
MD + corr			0.811	0.806	0.801	0.800	0.805	0.809
MD + corr + FA				0.806	0.797	0.805	0.803	0.806
MD + corr + FA + GMD					0.807	0.802	0.804	0.800
MD + corr + FA + GMD + AxD						0.810	0.805	0.808
MD + corr + FA + GMD + AxD + pcor							0.803	0.801
MD + corr + FA + GMD + AxD + pcor + RD								0.801

Multiparametric models result from stepwise adding measures to the best performing classification model of the previous step, starting with the best performing single MRI measure (i.e. MD, underlined). Bold: best performing model. Italic: mean AUC significantly above chance level

after family-wise error rate correction. GMD: gray matter density, WMD: white matter density, FA: fractional anisotropy, MD: mean diffusivity, AxD: axial diffusivity, RD: radial diffusivity, corr: full correlations between ICA components, pcor: L1-regularized partial correlations between ICA components, AUC: area under the ROC curve.

Table 3. Classification performance values of AD versus bvFTD classifications using single or multiple MRI-derived measures.

Measure	AUC	Min - Max	Sensitivity	Specificity	Accuracy	Cutoff
GMD	0.614*	0.535 - 0.722	0.642	0.651	0.647	0.448
WMD	0.614*	0.533 - 0.712	0.657	0.635	0.645	0.451
FA	0.623*	0.530 - 0.725	0.670	0.635	0.650	0.449
<u>MD</u>	<u>0.708[#]</u>	<u>0.625 - 0.775</u>	<u>0.740</u>	<u>0.710</u>	<u>0.723</u>	<u>0.439</u>
AxD	0.636*	0.554 - 0.713	0.672	0.649	0.659	0.435
RD	0.652*	0.557 - 0.730	0.710	0.652	0.677	0.434
corr	0.674 ^{#*}	0.599 - 0.767	0.656	0.657	0.657	0.455
pcor	0.561*	0.455 - 0.639	0.597	0.652	0.628	0.448
MD + corr + FA	0.811^{##}	0.755 - 0.862	0.746	0.801	0.777	0.425

Mean, minimum, and maximum area under the ROC curve (AUC) after 50 times 10-fold cross-validation. Bold: best performing model, Underlined: best performing single measure-based model. Mean sensitivity, specificity, and classification accuracy were derived from the

optimal operating point on the ROC (cutoff). GMD: gray matter density, WMD: white matter density, FA: fractional anisotropy, MD: mean diffusivity, AxD: axial diffusivity, RD: radial diffusivity, corr: full correlations between ICA components, pcor: L1-regularized partial correlations between ICA components, AUC: area under the ROC curve. *P<0.05 versus multiparametric model, #P<0.05 versus chance, ##P<0.001 versus chance.

Table 4. Classification performance values of control versus bvFTD classifications using single or multiple MRI-derived measures.

Measure	AUC	Min - Max	Sensitivity	Specificity	Accuracy	Cutoff
GMD	0.858 ^{##}	0.827 - 0.896	0.806	0.865	0.841	0.400
WMD	0.584*	0.504 - 0.655	0.601	0.633	0.620	0.399
<u>FA</u>	<u>0.862^{##}</u>	<u>0.810 - 0.903</u>	<u>0.842</u>	<u>0.845</u>	<u>0.844</u>	<u>0.427</u>
MD	0.614* [#]	0.508 - 0.708	0.628	0.621	0.624	0.399
AxD	0.723* [#]	0.599 - 0.779	0.676	0.727	0.707	0.396
RD	0.809* ^{##}	0.763 - 0.858	0.793	0.765	0.776	0.383
corr	0.648*	0.534 - 0.716	0.649	0.658	0.654	0.387
pcor	0.613*	0.504 - 0.729	0.563	0.708	0.651	0.414
FA + GMD + corr	0.922	0.877 - 0.954	0.867	0.896	0.884	0.437

Mean, minimum, and maximum area under the ROC curve (AUC) after 50 times 10-fold cross validation. Bold: best performing model. Underlined: best performing single measure-based model. Mean sensitivity, specificity, and classification accuracy were derived from the optimal operating point on the ROC (cutoff). GMD: gray matter density, WMD: white matter density, FA: fractional anisotropy, MD: mean diffusivity, AxD: axial diffusivity, RD: radial diffusivity, corr: full correlations between ICA components, pcor: L1-regularized partial correlations between ICA components, AUC: area under the ROC curve. Bold: best performing model, *P<0.05 versus multiparametric model, #P<0.05 versus chance, ##P<0.001 versus chance.

Table 5. Classification performance values of control versus AD classifications using single or multiple MRI-derived measures.

Measure	AUC	Min - Max	Sensitivity	Specificit		
				y	Accuracy	Cutoff
<u>GMD</u>	<u>0.940^{##}</u>	<u>0.913 - 0.962</u>	<u>0.901</u>	<u>0.899</u>	<u>0.900</u>	<u>0.384</u>
WMD	0.722 ^{##\$*}	0.644 - 0.770	0.691	0.671	0.681	0.406
FA	0.777 ^{\$*}	0.734 - 0.811	0.745	0.729	0.736	0.435
MD	0.670 ^{##\$*}	0.601 - 0.734	0.623	0.713	0.671	0.512
AxD	0.790 ^{\$*}	0.731 - 0.821	0.704	0.787	0.749	0.459
RD	0.832 [*]	0.784 - 0.865	0.827	0.749	0.785	0.369
corr	0.708 ^{##\$*}	0.641 - 0.738	0.631	0.792	0.718	0.527
pcor	0.657 ^{\$*}	0.621 - 0.704	0.630	0.677	0.655	0.481
GMD + AxD + FA + RD + MD	0.941	0.910 - 0.966	0.898	0.904	0.901	0.402

Mean, minimum, and maximum area under the ROC curve (AUC) after 50 repetitions. Bold: best performing model. Underlined: best performing single measure-based model. Mean sensitivity, specificity, and classification accuracy were derived from the optimal operating point on the ROC (cutoff). GMD: gray matter density, WMD: white matter density, FA: fractional anisotropy, MD: mean diffusivity, AxD: axial diffusivity, RD: radial diffusivity, corr: full correlations between ICA components, pcor: L1-regularized partial correlations between ICA components, AUC: area under the ROC curve. Bold: best performing model, *P<0.05 versus multiparametric model, §P<0.05 versus GMD, #P<0.05 versus chance, ##P<0.001 versus chance.

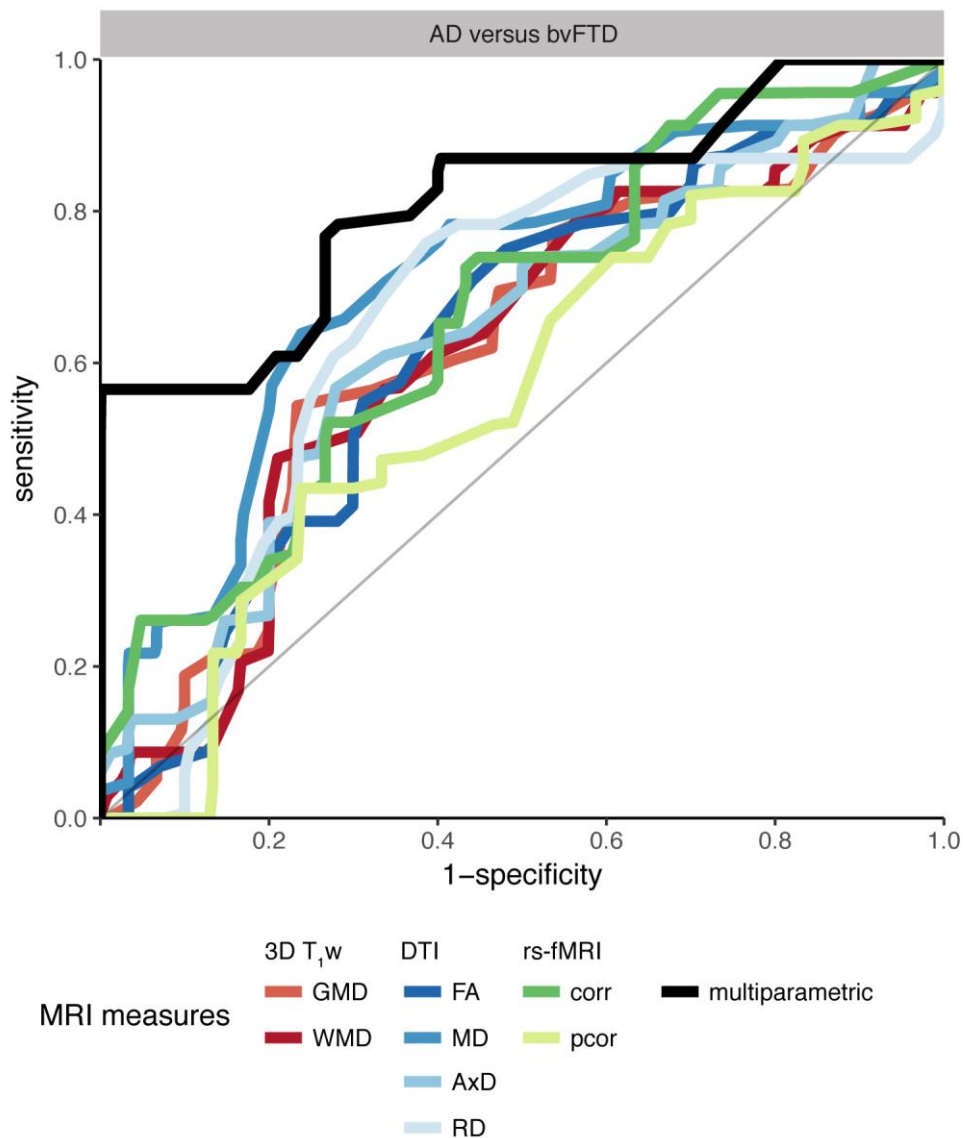


Figure 1. Receiver-operating characteristic curves of classifications between AD and bvFTD patients. For each differential diagnosis, measurements of gray matter density (GMD), white matter density (WMD), fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AxD), radial diffusivity (RD), full correlation between rs-fMRI derived independent components (cor), and L1-regularized partial correlation between rs-fMRI-derived independent components (pcor) were separately assessed. The multiparametric curve resulted from step-wise combining MRI measures, each time adding a measure to the combination with the highest AUC values of the previous step. Highest AUC was obtained by combining MD, corr, and FA measurements (mean AUC [min-max]) 0.811 [0.755 – 0.862]). The diagonal line represents random classification performance.

Supplementary Material

Supplementary Table 1. MRI sequence parameter settings

	Slices	TR (ms)	TE (ms)	Flip Angle (°)	Matrix (mm)	Voxel size (mm)	Duration (min)
VUmc (Signa HDxt, GE Healthcare)							
3DT1w	180	7.8	3	12	256 x 256	0.98 x 0.98 x 1.00	4.57
DTI*	45	13000	87.8	90	128 x 128	2.00 x 2.00 x 2.40	7.48
rs-fMRI	34	1800	35	80	64 x 64	3.30 x 3.30 x 3.30 [^]	6.04
LUMC (Achiva, Philips Medical Systems)							
3DT1w	140	9.8	4.6	8	256 x 256	0.88 x 0.88 x 1.20	4.57
DTI [#]	70	8250	80	90	128 x 128	2.00 x 2.00 x 2.00	8.48
rs-fMRI	38	2200	30	80	80 x 80	2.75 x 2.75 x 2.99 [^]	7.28

Scan protocol of whole-brain near-isotropic 3DT₁-weighted (3DT_{1w}), diffusion tensor imaging (DTI), and resting state functional MRI T₂*-weighted MRI (rs-fMRI) on 3T scanners at the VU University Medical center (VUmc) and the Leiden University Medical Center (LUMC).

* 30 directions, b=1000, one b₀ image, [#] 60 directions, b=1000, five b₀ images, [^]10% interslice gap

Supplementary Table 2. Mean AUC values for control versus bvFTD classifications using single or combination of multiple MRI-derived measures.

Combined with:	MRI Measure							
	FA	GMD	corr	WMD	AxD	pcor	MD	RD
-	<u>0.862</u>	0.858	0.648	0.584	0.723	0.613	0.614	0.809
FA		0.918	0.855	0.825	0.865	0.873	0.861	0.861
FA + GMD			0.922	0.919	0.920	0.914	0.920	0.918
FA + GMD + corr				0.920	0.917	0.917	0.918	0.915
FA + GMD + corr + WMD					0.921	0.917	0.918	0.918
FA + GMD + corr + WMD + AxD						0.920	0.913	0.920
FA + GMD + corr + WMD + AxD + pcor							0.916	0.915
FA + GMD + corr + WMD + AxD + pcor + MD								0.918

Mean AUC values from 50 times 10-fold cross-validations. Multiparametric models result from stepwise adding measures to the best performing classification model of the previous step, starting with the best performing single MRI measure (i.e. FA, underlined). **Bold**: best performing model. *Italic*: mean AUC significantly above chance level after family-wise error rate correction. GMD: gray matter density, WMD: white matter density, FA: fractional anisotropy, MD: mean diffusivity, AxD: axial diffusivity, RD: radial diffusivity, corr: full correlations between ICA components, pcor: L1-regularized partial correlations between ICA components, AUC: area under the ROC curve.

Supplementary Table 3. Mean AUC values for control versus AD classifications using single or combination of multiple MRI-derived measures.

Combined with:	MRI measure							
	GMD	AxD	FA	RD	MD	pcor	corr	WMD
-	<u>0.940</u>	0.790	0.777	0.832	0.670	0.657	0.708	0.722
GMD		0.940	0.936	0.935	0.935	0.940	0.938	0.937
GMD + AxD			0.938	0.938	0.938	0.938	0.936	0.935
GMD + AxD + FA				0.938	0.935	0.934	0.937	0.936
GMD + AxD + FA + RD					0.941	0.940	0.936	0.939
GMD + AxD + FA + RD + MD						0.939	0.936	0.935
GMD + AxD + FA + RD + MD + pcor							0.937	0.935
GMD + AxD + FA + RD + MD + pcor + corr								0.936

Mean AUC values from 50 times 10-fold cross-validations. Multiparametric models result from stepwise adding measures to the best performing classification model of the previous step, starting with the best performing single MRI measure (i.e. GMD, underlined). Bold: best performing model. Italic: mean AUC significantly above chance level after family-wise error rate correction. GMD: gray matter density, WMD: white matter density, FA: fractional anisotropy, MD: mean diffusivity, AxD: axial diffusivity, RD: radial diffusivity, corr: full correlations between ICA components, pcor: L1-regularized partial correlations between ICA components, AUC: area under the ROC curve.