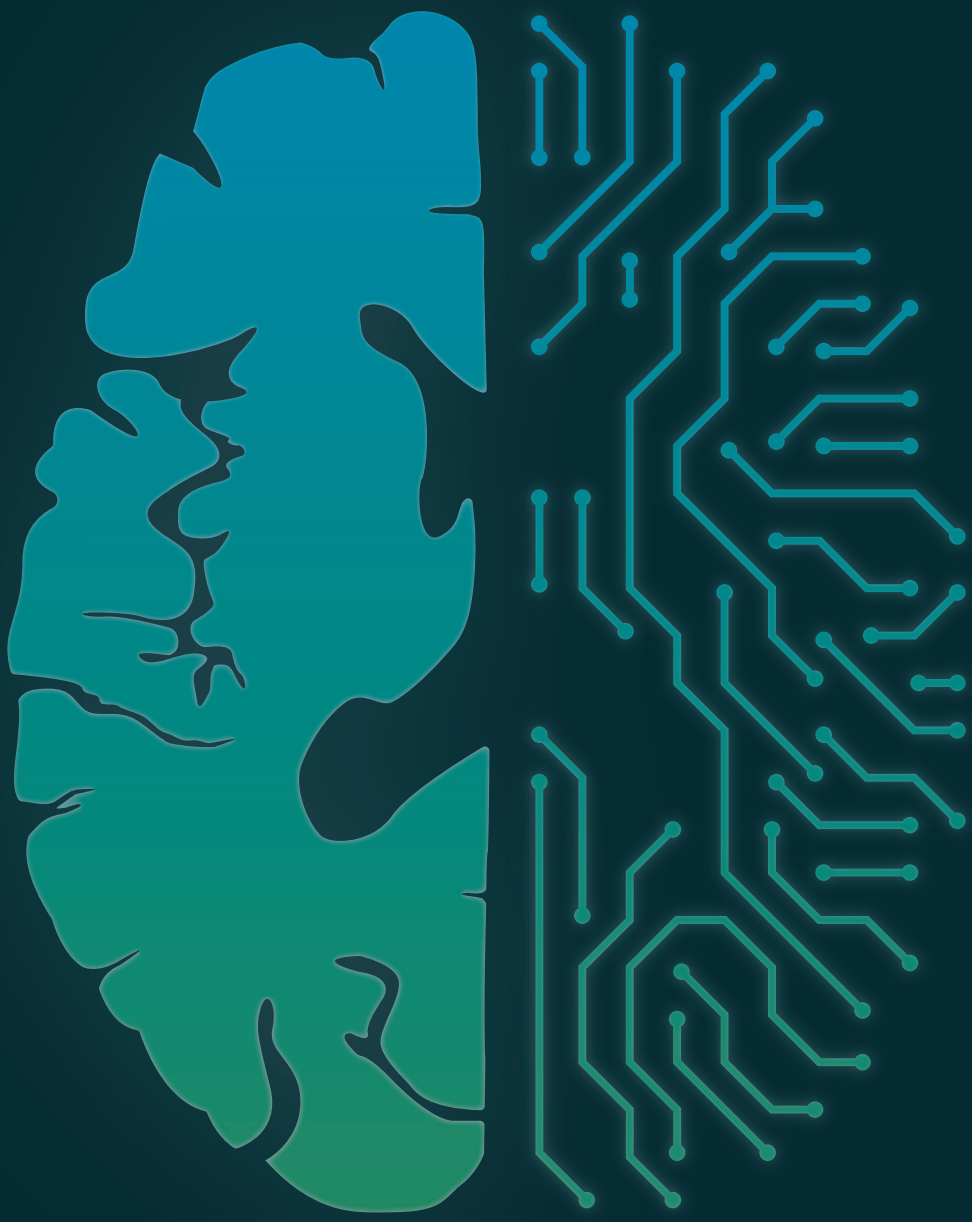
Neuroimaging biomarkers in genetic frontotemporal dementia : towards a timely diagnosis

Feis, R.A.



Chapter 2

Frontotemporal dementia mutation carriers: presymptomatic MRI-based classification

Published in *NeuroImage: Clinical 2019;22:101718* as:

Single-subject classification of presymptomatic frontotemporal dementia mutation carriers using multimodal MRI

Rogier A. Feis, Mark J.R.J. Bouts, Jessica L. Panman, Lize C. Jiskoot, Elise G.P. Dopper, Tijn M. Schouten, Frank de Vos, Jeroen van der Grond, John C. van Swieten, Serge A.R.B. Rombouts

Abstract

Classification models based on magnetic resonance imaging (MRI) may aid early diagnosis of frontotemporal dementia (FTD) but have only been applied in established FTD cases. Detection of FTD patients in earlier disease stages, such as presymptomatic mutation carriers, may further advance early diagnosis and treatment. In this study, we aim to distinguish presymptomatic FTD mutation carriers from controls on an individual level using multimodal MRI-based classification.

Anatomical MRI, diffusion-weighted imaging (DWI) and resting-state functional MRI data were collected in 55 presymptomatic FTD mutation carriers (8 microtubule-associated protein tau, 35 progranulin, and 12 chromosome 9 open reading frame 72 repeat expansion) and 48 familial controls. We calculated grey and white matter density features from anatomical MRI scans, diffusivity features from DWI, and functional connectivity features from resting-state functional MRI. These features were applied in a recently introduced multimodal behavioural variant FTD (bvFTD) classification model, and were subsequently used to train and test unimodal and multimodal carrier-control models. Classification performance was quantified using area under the receiver operating characteristic curves (AUC).

The bvFTD model was not able to separate presymptomatic mutation carriers from controls beyond chance level (AUC = 0.582, $p = 0.078$). In contrast, one unimodal and several multimodal carrier-control models performed significantly better than chance level. The unimodal model included the radial diffusivity feature and had an AUC of 0.642 ($p = 0.032$). The best multimodal model combined radial diffusivity and white matter density features (AUC = 0.684, $p = 0.004$).

FTD mutation carriers can be separated from controls with a modest AUC even before symptom onset, using a newly created carrier-control classification model, while this was not possible using a recent bvFTD classification model. A multimodal MRI-based classification score may therefore be a useful biomarker to aid earlier FTD diagnosis. The exclusive selection of white matter features in the best performing model suggests that the earliest FTD-related pathological processes occur in white matter.

Keywords: frontotemporal dementia; *MAPT* protein, human; *GRN* protein, human; *C9orf72*, human; diffusion tensor imaging; resting-state functional MRI; multimodal MRI; classification; machine learning

Introduction

Frontotemporal lobar degeneration is a common cause of early-onset dementia with a similar prevalence to Alzheimer's disease in the presenile population (Ratnavalli et al., 2002; Harvey et al., 2003; Rabinovici & Miller, 2010; Rascovsky et al., 2011; Seelaar et al., 2011). Although there are clinical disease criteria for the different clinical variants of frontotemporal dementia (FTD; Gorno-Tempini et al., 2011; Rascovsky et al., 2011), diagnosis is often complicated and delayed by clinical heterogeneity. This hinders clinicians in providing accurate prognosis, effective disease management, and developing new treatments (Mohs et al., 2001; Mendez et al., 2007; Mendez, 2009; Pressman & Miller, 2014).

Multimodal magnetic resonance imaging (MRI) has been suggested as a promising biomarker to improve on diagnostic standards in FTD. In FTD patients, MRI revealed specific patterns of neurodegeneration, involving grey matter and white matter atrophy (Whitwell & Jack, 2005; Chao et al., 2007; Rabinovici et al., 2008; Seeley et al., 2009; Zhang et al., 2011; Pan et al., 2012; Whitwell et al., 2012, 2015; Risacher & Saykin, 2013; Frings et al., 2014; Möller et al., 2015a), differences in diffusion tensor imaging (DTI) measures (Zhang et al., 2009, 2011; Agosta et al., 2012; McMillan et al., 2012, 2014; Mahoney et al., 2014; Möller et al., 2015b; Daianu et al., 2016), and differences in functional connectivity (Zhou et al., 2010; Farb et al., 2013; Lee et al., 2014; Zhou & Seeley, 2014; Hafkemeijer et al., 2015, 2016).

These patterns have subsequently been utilised on an individual level to create MRI-based classification algorithms that can discriminate between FTD patients and control subjects (Davatzikos et al., 2008; McMillan et al., 2014; Raamana et al., 2014; Koikkalainen et al., 2016; Wang et al., 2016; Bron et al., 2017; Meyer et al., 2017; Bouts et al., 2018). Accurate classification of FTD patients using MRI measures is an important step towards a more substantiated diagnostic standard. However, most classification models are based on established FTD cases, limiting generalisability in patients who are at an earlier disease stage. Still, detection of these early-stage FTD cases is necessary to facilitate precise subject recruitment into clinical trials and potential early treatment with disease-modifying drugs (Huey et al., 2008).

In order to characterise FTD pathophysiology at an earlier stage, presymptomatic carriers of autosomal dominant FTD gene mutations were compared to controls in MRI group analyses (Borroni et al., 2008, 2012; Whitwell et al., 2011a; Rohrer et al., 2013, 2015; Dopper et al., 2014; Premi et al., 2014; Lee et al., 2017; Papma et al., 2017; Bertrand et al., 2018; Cash et al., 2018). Carriers of the three most common FTD gene mutations microtubule-associated protein tau (*MAPT*), progranulin (*GRN*), and chromosome 9 open reading frame 72 (*C9orf72*) repeat expansion show brain alterations on MRI, even well before symptom onset. In these subjects, white matter diffusivity changes (Borroni et al., 2008; Dopper et al., 2014; Papma et al., 2017) and functional connectivity changes (Whitwell et al., 2011a; Borroni et al., 2012; Dopper et al., 2014; Premi et al., 2014) are often, but not exclusively (Lee et al., 2017; Papma et al., 2017; Bertrand et al., 2018), found in the absence of grey matter atrophy, suggesting that changes in the functional architecture and white matter tracts may precede structural deterioration in the grey matter (Rohrer et al., 2013). Nonetheless, multicentre analyses of a large international cohort show grey matter loss in *MAPT* mutation carriers, *GRN* mutation carriers, and *C9orf72* repeat expansion carriers even before 'conversion' to symptomatic FTD (Rohrer et al., 2015; Cash et al., 2018). Although these presymptomatic group differences give insight into the pathophysiological mechanisms of FTD, individual heterogeneity complicates its utility in FTD diagnosis. Therefore, translation from group differences to single-subject classification models is imperative.

The present study brings two research areas together: we combine machine learning with presymptomatic FTD mutation carriers to study individual classification of FTD-pathology at an early stage. Our aim is to distinguish individual presymptomatic FTD mutation carriers from healthy controls using multimodal MRI.

Methods

Design

In order to distinguish presymptomatic FTD mutation carriers from controls, we applied two models. First, we applied a recent behavioural variant FTD (bvFTD-)control classification model (Bouts et al., 2018) to our MRI data to investigate whether the model separates presymptomatic mutation carriers from controls. We shall refer to this model as the 'bvFTD model'. In a second analysis, we trained new classification models on the presymptomatic mutation carriers and controls' data using cross-validation. We shall refer to these models as 'carrier-control models'. MRI preprocessing, feature selection and classification were performed identically to previous work (Bouts et al., 2018).

Participants

This retrospective study partially included previously published (Dopper et al., 2014; Jiskoot et al., 2016; Papma et al., 2017) and newly acquired data from the Erasmus Medical Centre and Leiden University Medical Centre. Participants and clinical investigators were blinded to the participants' DNA status. The study was conducted in accordance with regional regulations and the Declaration of Helsinki. The Erasmus Medical Centre and Leiden University Medical Centre local medical ethics committees approved the study, and every participant provided written informed consent.

For the current study, we included 55 presymptomatic FTD mutation carriers (8 *MAPT*, 35 *GRN*, and 12 *C9orf72* repeat expansion) and 48 healthy familial controls (6 *MAPT* family, 31 *GRN* family, and 11 *C9orf72* family) between May 2010 and March 2016. These subjects were recruited from a cohort of healthy first-degree relatives of FTD patients with either an *MAPT* mutation, *GRN* mutation, or *C9orf72* repeat expansion (FTD-Risk Cohort; FTD-RisC), and visited the Erasmus Medical Centre for a one-day assessment in order to ascertain asymptomatic status, collect clinical data, and determine DNA status as described before (Dopper et al., 2014; Jiskoot et al., 2016; Papma et al., 2017). Participants were considered asymptomatic in the absence of (1) behavioural, cognitive, or neuropsychiatric change reported by the participant or knowledgeable informant, (2) cognitive disorders on neuropsychiatric tests, (3) motor neuron disease signs on neurologic examination, and (4) other FTD (Gorno-Tempini et al., 2011; Rascovsky et al., 2011) or amyotrophic lateral sclerosis (Ludolph et al., 2015) criteria. Healthy controls were assumed to have equal FTD risk as the general population. For a more detailed description of the recruitment protocol, see earlier work (Dopper et al., 2014; Jiskoot et al., 2016; Papma et al., 2017). Inclusion criteria for the current study were: age between 40 and 70 years, and availability of a 3-dimensional T_1-weighted MRI ($3DT_1w$) scan, a diffusion-weighted imaging (DWI) data set, and a resting-state functional T_2*-weighted MRI (rs-fMRI) scan. Exclusion criteria were: current or past neurologic or psychiatric disorders, history of drug abuse, large image artefacts, and gross brain pathology other than atrophy.

For details on the sample on which the bvFTD model was trained, please refer to Bouts et al. (2018). In short, 23 bvFTD patients and 35 controls between 40 and 80 years old were included to undergo a clinical assessment and MRI between November 2009 and November 2012. The MRI acquisition protocol was similar to the protocol applied in the current sample of mutation carriers and controls. Image processing steps were identical to processing steps in the current sample.

MRI data acquisition

All subjects were scanned at the Leiden University Medical Centre using a 3 T MRI scanner (Achieva, Philips Medical Systems, Best, The Netherlands) with an 8-channel SENSE head coil. The imaging protocol included a whole-brain near-isotropic $3DT_1w$ sequence for cortical and subcortical tissue type segmentation, a DWI sequence for assessments of white matter diffusion metrics, and an rs-fMRI 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. For scan parameters, see **Table 2.1**.

Table 2.1 MRI sequence parameter settings

	$3DT_1w$	DWI[a]	rs-fMRI
Slices, *n*	140	70	38
TR, ms	9.8	8,250	2,200
TE, ms	4.6	80	30
Flip angle, °	8	90	80
Matrix, mm	256 × 256	128 × 128	80 × 80
Voxel size, mm	0.88 × 0.88 × 1.20	2.00 × 2.00 × 2.00	2.75 × 2.75 × 2.99[b]
Duration, min	4.57	8.48	7.28

Scan protocol of whole-brain near-isotropic 3-dimensional T_1-weighted ($3DT_1w$), diffusion-weighted imaging (DWI), and resting-state functional T_2*-weighted MRI (rs-fMRI) on a 3 T scanner at the Leiden University Medical Centre.
TR, repetition time; TE, echo time.
[a] 60 directions, b = 1,000, one b0 image.
[b] Including 10% interslice gap.

Image preprocessing

For $3DT_1w$ images, the following preprocessing steps were performed: bias field correction (N4ITK; Tustison et al., 2010), brain extraction (FSL BET; Smith, 2002), nonlinear registration to the MNI152 2 × 2 × 2 mm T1 template (FNIRT; Anderson et al., 2007), tissue type segmentation (SPM12; Friston et al., 2007) and segmentation of deep grey matter structures, including the bilateral thalamus, caudate nucleus, putamen, globus pallidum, nucleus accumbens, amygdala, and hippocampus (FIRST; Patenaude et al., 2011).

Preprocessing for DWI data sets included correction of motion and eddy-current induced distortion (eddy correct; Leemans & Jones, 2009), and calculation of voxel-wise measures of fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AxD, largest eigenvalue), and radial diffusivity (RD, average of the two remaining eigenvalues, DTIFIT; Smith et al., 2004). A global mean FA image was created by nonlinearly registering FA maps to the FMRIB58_FA template, and tract-based spatial statistics (FSL TBSS; Smith et al., 2006) was used to extract FA, MD, AxD, and RD values using the standard FSL TBSS skeleton. The skeleton was thresholded at 0.2 to ensure skeleton extracted values originate from white matter.

For rs-fMRI data, preprocessing included motion correction (Jenkinson et al., 2002), brain extraction, spatial smoothing using a Gaussian kernel with a full width at half maximum of 3 mm, grand mean intensity normalisation, motion artefact removal, and high-pass temporal filtering (cut-off frequency = 0.01 Hz). Motion artefacts were removed using a single-session independent

component analysis (ICA) to decompose the rs-fMRI data into distinct statistically independent components. Subsequently, motion-related components were automatically identified and removed using the ICA-based automatic removal of motion artefacts (ICA-AROMA, version 0.3-beta; Pruim et al., 2015) procedure. Registration to standard space was performed in two steps. First, a temporal mean image calculated from the 4D rs-fMRI volume was registered to the $3DT_1$w image using boundary-based registration (Greve & Fischl, 2009). Next, resulting registration parameters were concatenated to the $3DT_1$w-to-MNI152 template registration parameters to obtain the final registration parameters.

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

Feature selection

Cortical grey matter density (GMD) and white matter density (WMD) were calculated as a weighted average of their respective regional white matter or grey matter probability (SPM segmentation) weighted by the probability of a voxel being part of that specific tract or region. The latter probabilities were derived from the 48 Harvard-Oxford probabilistic anatomical brain atlas cortical regions (split into left and right) and from the Johns Hopkins University white matter tractography atlas for 20 white matter tract regions. Voxels with region probability values less than 25% were excluded. This provided a measure of brain atrophy of a specific grey matter region or white matter tract. For deep grey matter regions, GMD values were calculated as the regions' volume (FIRST segmentations) divided by total intracranial volume. This resulted in a feature vector of 110 average GMD values (48 left cortical, 48 right cortical, and 14 deep grey matter regions), and a feature vector of 20 average WMD values per subject.

DTI-based features were calculated by projecting each subject's FA, MD, AxD, and RD values onto the TBSS group skeleton on a voxel-wise basis. Like the WMD features, the 20 white matter tracts of the probabilistic Johns Hopkins University white matter tractography atlas were then used to calculate a weighted mean value per tract per subject. This resulted in 4×20 feature vectors of mean FA, MD, AxD, and RD values per subject.

In order to calculate the functional connectivity features, all processed rs-fMRI images were combined in a temporally concatenated ICA (Beckmann & Smith, 2004), with dimensionality fixed at 70 components and an ICA threshold of 0.99 (Smith et al., 2013). 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 grey matter 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 (FCor) or as sparse, L1-regularised, partial correlations (PCor) between the components' time courses. Partial correlations were calculated using the graphical lasso algorithm (J. Friedman et al., 2008). The functional connectivity measures resulted in two feature vectors of each $(70 \times 69) \div 2 = 2{,}415$ (partial) correlations per subject. Finally, we concatenated all feature vectors into one vector per subject.

BvFTD model

For our first analysis, a bvFTD patient-control classification model (Bouts et al., 2018) was applied to each subject's extracted feature vector. We applied the best performing, multimodal model that

discriminated bvFTD patients from controls, which included the features GMD, FCor, FA, and MD (Bouts et al., 2018), as well as age and sex. Each subject's feature vector was fed into the model, resulting in a classification score from 0 to 1, where 0 represents a control subject and 1 represents a bvFTD patient. Extrapolated to our subjects, these scores showed how alike our presymptomatic FTD mutation carriers and healthy controls are to bvFTD patients.

Carrier-control models

For the second analysis, feature vectors were used to train a logistic elastic net regression algorithm (Zou & Hastie, 2005; Friedman et al., 2010; Schouten et al., 2016; Bouts et al., 2018). The elastic net regression procedure estimates a sparse regression model that includes only a subset of the provided features by imposing a penalty for including features and for the weight of each feature. This way, elastic net provides a solution for the imbalance between the large number of features and the small number of subjects. Age and sex were included into the model without penalty to ensure that estimated feature regression coefficients were conditional on subject age and sex. Here, a classification score of 0 represented a control subject and 1 represented a presymptomatic FTD mutation carrier.

Cross-validation

Similarly to previous work (Schouten et al., 2016; Bouts et al., 2018), we trained our carrier-control model in a nested 10-fold cross-validation scheme to reduce classification bias. One part of the data (e.g., 10%) was set apart as a test set and served to test the generalised classification performance of the elastic net regression model. The remaining parts (90%) were used to train the model. However, in addition to the classification performance, we also wanted to determine the optimal penalty size without overestimating classification performance (Varma & Simon, 2006; Kriegeskorte et al., 2009). To this end, we used a second, nested, 10-fold cross-validation loop on the training set over a grid of hyperparameters to determine the optimal penalty. In the nested loop, we estimated the model's hyperparameters that corresponded with the lowest binomial deviance, a goodness-of-fit measure that evaluates the difference between the predicted and actual observations. Next, these hyperparameters and corresponding penalties were used to train a model using the training set of the outer loop. Finally, the classification performance was tested on the test set of the outer loop. This process was repeated ten times to make sure that each subject was part of the test set at least once. Since the test set of the outer loop was neither used for model training, nor for parameter optimisation, potential prediction bias was reduced as much as possible (Kriegeskorte et al., 2009). The entire classification procedure was repeated 50 times to average classification outcome variability resulting from random partitioning in training and test folds. All classification analyses and evaluations were implemented in R version 3.3.2 (R core 2016, GLMnet package; Friedman et al., 2010).

Classification performance

For both analyses, we quantified classification performances using receiver operating characteristic (ROC) curves. ROC curves were calculated by shifting the threshold for classifying an individual as patient (bvFTD model analysis) or mutation carrier (carrier-control model analysis) from 0 to 1, and plotting the true positive rate (sensitivity) vs. the false positive rate (1 – specificity) for each intermediate point. The area under this ROC curve (AUC) is a measure of classification performance insensitive to the distribution between the groups (Fawcett, 2006). Additionally, we

calculated the optimal operating point on the curve to calculate the model's sensitivity, specificity, and classification accuracy, given equal class distribution, and equal penalty for false positive and false negative predictions. For the carrier-control model analysis, we averaged AUC, accuracy, sensitivity, and specificity values from the 50 times repeated nested cross-validations.

Multimodal classification

To obtain the best multimodal carrier-control model using several feature vectors, we performed step-wise feature concatenation as previously described (Schouten et al., 2016; Bouts et al., 2018). First, we assessed classification performance for each feature separately. Subsequently, we added a new feature to the best performing feature combination (i.e., highest AUC) of the previous step until all features were included in the model. The best performing feature combination will be referred to as the multimodal carrier-control model.

Statistical analysis

Statistical analyses of non-imaging data were performed using R (R Core 2016, Vienna, Austria). We tested for carrier-control differences using independent *t*-tests (age and education), the Mann-Whitney U test (mini-mental state examination [MMSE] scores [0–30]) and the chi-square test (sex distribution). Classification scores were compared using Mann-Whitney U tests for overall carrier-control contrasts, and Kruskal Wallis H tests and Dunn post-hoc tests for comparisons between all four groups (*MAPT*, *GRN*, *C9orf72*, and controls). To compare models' AUC values against chance level, we used permutation tests (N = 5,000; Noirhomme et al., 2014). In order to correct for multiple comparisons, we took the maximum AUC difference of the family of tests for each permutation. Then we compared the observed AUC difference to the new distribution of maximum AUC differences to get a family-wise error rate corrected p-value. The alpha level required for statistical significance was set at 0.05.

Results

Demographics

In total, 103 subjects met the inclusion criteria (**Table 2.2**). Mean age was similar for mutation carriers (52.0 ± 8.6 years) and healthy controls (54.2 ± 7.5 years). The proportion of female participants between mutation carriers (67%) and healthy controls (58%) was not different (p = 0.35). Education level was similar between groups (mutation carriers, 13.6 ± 2.9 years; healthy controls, 13.2 ± 2.4 years). MMSE was similarly distributed between groups (median [min–max], mutation carriers: 30 [24–30], healthy controls: 29 [24–30]).

Table 2.2 Participant demographics

	Mutation carriers (n = 55)	**Controls (n = 48)**	***p*-value**
Gene, *n MAPT* / *GRN* / *C9orf72*	8 / 35 / 12	–	–
Age, mean (SD) years	52.0 (8.6)	54.2 (7.5)	0.18
Sex, *n* (%) ♀	37 (67%)	28 (58%)	0.35
Education, mean (SD) years[a]	13.6 (2.9)	13.2 (2.4)	0.42
MMSE, median (range) points	30 (24–30)	29 (24–30)	0.53

C9orf72, chromosome 9 open reading frame 72; *GRN*, progranulin; *MAPT*, microtubule-associated protein tau; MMSE, mini-mental state examination.

[a] Education values were missing for one mutation carrier and two controls.

BvFTD model

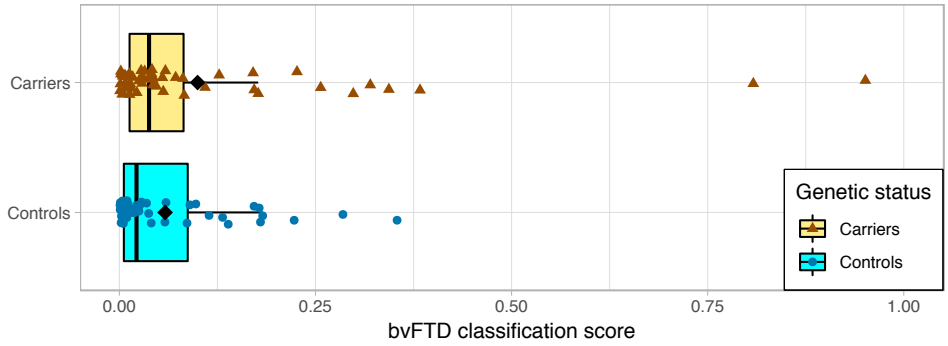
Application of the bvFTD model resulted in low bvFTD classification scores for most subjects (**Figure 2.1A**), and the bvFTD classification scores were not significantly different between presymptomatic mutation carriers (median = 0.038) and controls (median = 0.022, p = 0.15). ROC analysis of the bvFTD classification scores resulted in an AUC of 0.582, which was not significantly better than chance level (p = 0.078). Separated by gene (**Figure 2.1B**), there were no differences between the four groups' bvFTD classification scores (p = 0.37). BvFTD classification scores of the original patients and controls used for cross-validation of the bvFTD model were added as reference (**Figure 2.1C**, data courtesy of Bouts et al., 2018).

Figure 2.1 Classification results bvFTD model (opposite page)

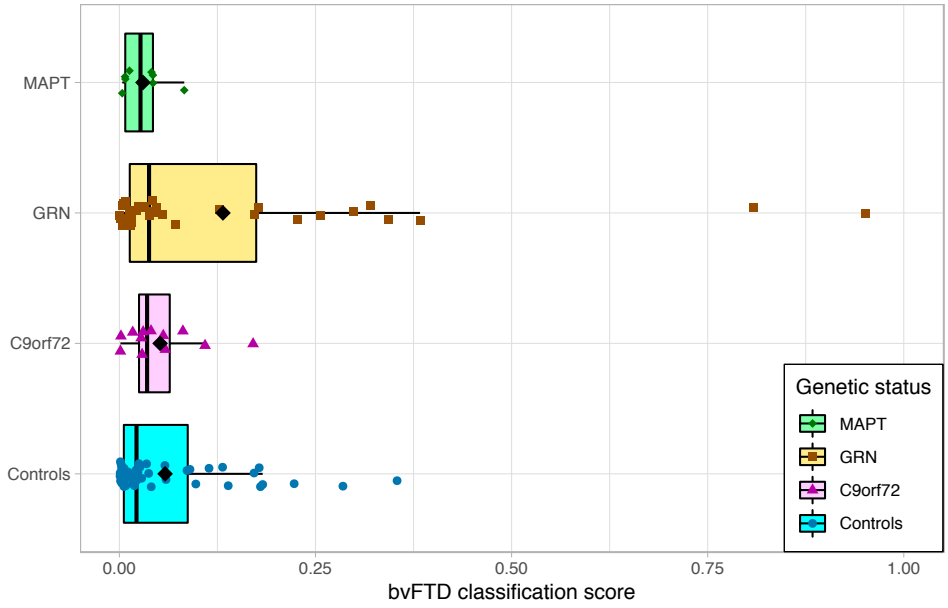
Box and scatter plot of each subject's bvFTD classification score on a scale from 0 (representing control) to 1 (representing bvFTD patient) after application of the bvFTD model. Groups are defined by mutation carrier status (**A**) and genetic status (**B**). Classification scores were not significantly different for mutation carriers and controls (p = 0.15), and did not differ between the four genetic groups (p = 0.37). Classification score results of the bvFTD patients and controls on which the bvFTD model was trained were added for reference (**C**, data courtesy of Bouts et al., 2018).

BvFTD, behavioural variant frontotemporal dementia; *C9orf72*, chromosome 9 open reading frame 72; *GRN*, progranulin; *MAPT*, microtubule-associated protein tau.

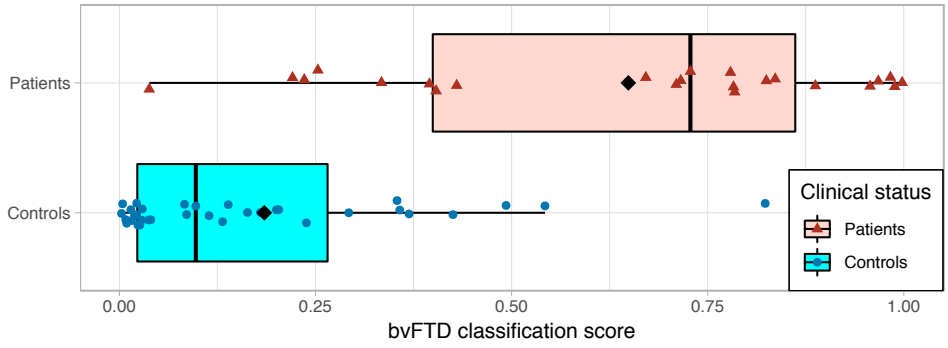
A. Classification by carrier status



B. Classification by genetic status



C. Classification by disease status



Carrier-control models

The best performing unimodal carrier-control models included RD, WMD, and MD, with AUCs of 0.642, 0.592, and 0.587, respectively. Of these models, only the RD model outperformed chance after family-wise error rate correction ($p = 0.032$, **Table 2.3**). Step-wise concatenation resulted in the best performing multimodal model, which included the features RD and WMD, and outperformed chance with an AUC of 0.684 ($p = 0.004$). Classification performance did not improve when additional features were added to this model (**Table 2.4**). Interestingly, all multimodal models that outperformed chance level used exclusively white matter features from diffusion-weighted and/or structural scans: RD, WMD, MD, and AxD).

Application of the best performing multimodal carrier-control model resulted in mutation carrier classification scores (**Figure 2.2**), which were different between mutation carriers (median = 0.589) and controls (median = 0.435, $p < 0.001$, **Figure 2.2A**). Furthermore, there was a difference between the four groups' mutation carrier classification scores ($p = 0.008$) when separated by gene (**Figure 2.2B**). Post-hoc tests revealed that *GRN* mutation carriers had higher mutation carrier classification scores than controls (Bonferroni family-wise error rate corrected $p = 0.009$). The other groups did not differ from each other.

Table 2.3 ROC characteristics

Modality	AUC	Min – max	Sensitivity	Specificity	Accuracy	FWERC *p*-value (AUC > chance)
GMD	0.502	0.427 – 0.554	0.538	0.514	0.554	0.895
WMD	0.592	0.547 – 0.656	0.595	0.585	0.663	0.190
FA	0.494	0.408 – 0.554	0.528	0.504	0.535	0.897
MD	0.587	0.523 – 0.629	0.591	0.579	0.642	0.240
AxD	0.554	0.500 – 0.598	0.568	0.559	0.578	0.448
RD	*0.642*	*0.583 – 0.673*	*0.637*	*0.629*	*0.669*	*0.032*
FCor	0.509	0.454 – 0.554	0.537	0.527	0.590	0.818
PCor	0.505	0.457 – 0.551	0.542	0.492	0.540	0.836
Multimodal	*0.684*	*0.629 – 0.722*	*0.664*	*0.640*	*0.735*	*0.004*

Presymptomatic mutation carriers vs. controls classification. Multimodal represents the best combination from our step-wise multimodal procedure (i.e., RD and WMD). Italic: mean AUC significantly higher than chance level after family-wise error rate correction.

AUC, area under the receiver operating characteristic curve; AxD, axial diffusivity; FA, fractional anisotropy; FCor, full correlations between ICA components; FWERC, family-wise error rate corrected; GMD, grey matter density; ICA, independent component analysis; MD, mean diffusivity; PCor, L1-regularised partial correlations between ICA components; RD, radial diffusivity; WMD, white matter density.

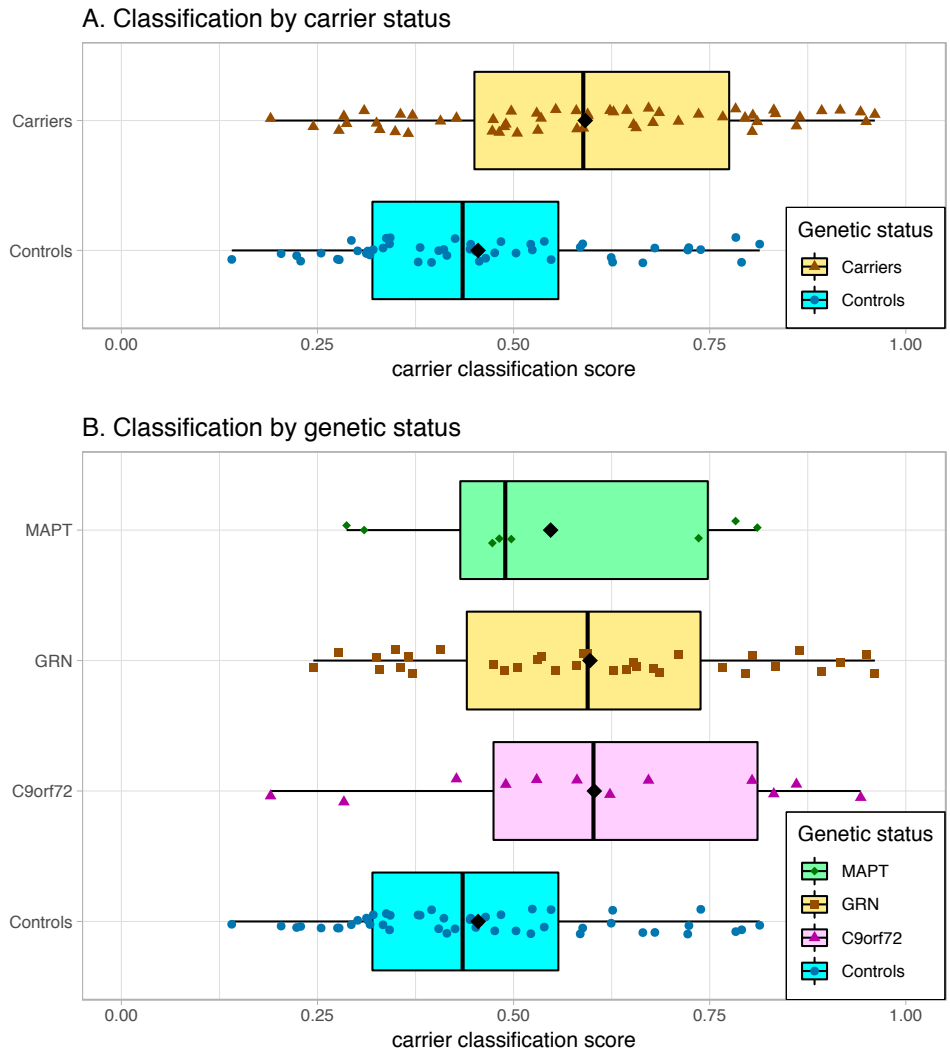


Figure 2.2 Classification results carrier-control model

Box and scatter plot of each subject's mutation carrier classification score on a scale from 0 (representing control) to 1 (representing presymptomatic mutation carrier) after application of the best performing carrier-control model including the features RD and WMD. Mutation carriers had significantly higher scores than controls (**A**, $p < 0.001$). Furthermore, there was an omnibus difference between the four genetic groups (**B**, $p = 0.008$), and post-hoc tests revealed higher scores for *GRN* mutation carriers than for controls ($p = 0.009$).

C9orf72, chromosome 9 open reading frame 72; *GRN*, progranulin; *MAPT*, microtubule-associated protein tau; RD, radial diffusivity; WMD, white matter density.

Table 2.4 Multimodal classification performance

Step\combined with:	**RD**	**WMD**	**MD**	**AxD**	**GMD**	**FA**	**FCor**	**PCor**
1: –	*0.642*	0.592	0.587	0.554	0.502	0.494	0.509	0.505
2: RD	–	***0.684***	0.621	0.600	0.565	0.564	0.516	0.501
3: RD + WMD	–	–	*0.658*	*0.640*	0.596	0.607	0.520	0.501
4: RD + WMD + MD	–	–	–	0.627	0.612	0.614	0.524	0.502
5: RD + WMD + MD + AxD	–	–	–	–	0.612	0.609	0.531	0.507
6: RD + WMD + MD + AxD + GMD	–	–	–	–	–	0.585	0.534	0.509
7: RD + WMD + MD + AxD + GMD + FA	–	–	–	–	–	–	0.530	0.500
8: RD + WMD + MD + AxD + GMD + FA + FCor	–	–	–	–	–	–	–	0.510

Mean AUC values from 50 repetitions. Multimodal models result from step-wise addition of measures to the best performing classification model of the previous step, starting with the best performing single measure (i.e., RD). Bold: best performing model. Italic: mean AUC significantly higher than chance level after family-wise error rate correction.

AUC, area under the receiver operating characteristic curve; AxD, axial diffusivity; FA, fractional anisotropy; FCor, full correlations between ICA components; GMD, grey matter density; ICA, independent component analysis; MD, mean diffusivity; PCor, L1-regularised partial correlations between ICA components; RD, radial diffusivity; WMD, white matter density.

Discussion

This study investigated whether presymptomatic FTD mutation carriers with mutations in *MAPT*, *GRN*, or with a *C9orf72* repeat expansion can be individually distinguished from healthy controls using MRI. Using a recently introduced MRI-based classification model trained on established bvFTD patients and controls, nearly all FTD mutation carriers and controls had low classification scores. The bvFTD model was therefore not able to separate mutation carriers from controls beyond chance level. However, MRI-based classification models that were trained on our own sample were able to separate mutation carriers from controls better than chance level. In our carrier-control model, the RD feature proved sufficient to separate mutation carriers from controls better than chance, but the best performing model used RD in combination with the WMD feature. All models that outperformed chance used exclusively white matter features, such as DTI features and WMD, supporting the hypothesis that white matter alterations are the first to appear in preclinical FTD pathology.

In an effort to improve on the FTD diagnostic criteria, single-subject classification using MRI measures has recently received significant attention (Davatzikos et al., 2008; McMillan et al., 2014; Raamana et al., 2014; Möller et al., 2015b; Koikkalainen et al., 2016; Wang et al., 2016; Bron et al., 2017; Meyer et al., 2017; Bouts et al., 2018). A recent multimodal classification study incorporated structural, DTI, and arterial spin labelling data to classify FTD (behavioural and language variants) from cognitively normal controls, and achieved an AUC of 0.96 (Bron et al., 2017). Another classification study included tissue density, DTI, and rs-fMRI measures, and achieved an AUC of 0.92 for bvFTD vs. cognitively normal controls (Bouts et al., 2018). These high classification performances are promising, but they are based on established FTD cases. It is unclear how FTD patient models generalise to earlier FTD stages, where brain alterations are less distinct. To test this, we applied a bvFTD model (Bouts et al., 2018) on FTD mutation carriers in a presymptomatic stage. We hypothesised that if the bvFTD model would be able to recognise early-stage FTD pathology, our presymptomatic FTD mutation carriers would have higher classification scores than controls. We found that it was not possible to separate mutation carriers from controls significantly better than chance using this model, as most mutation carriers and controls had very low bvFTD classification scores. This could indicate that presymptomatic differences present in FTD mutation carriers are too subtle to be picked up by a classification model that was trained established bvFTD patients. However, it could also mean that most of our mutation carriers were still too far from conversion to have significant FTD-related changes. Since the bvFTD model was trained on patients, it stands to reason that classification of mutation carriers and controls becomes more accurate as mutation carriers approach conversion. Vice versa, one might expect the mutation carriers with high classification scores to be closer to symptom onset than mutation carriers with lower classification scores. Although it was not statistically significant, there was a trend towards older age in mutation carriers with a bvFTD classification score higher than 0.25 than in the rest of the mutation carrier group (data not shown). It can therefore not be entirely ruled out that age is partly associated with a higher bvFTD score. Longitudinal research is warranted to formally test whether this model captures presymptomatic FTD-related changes as mutation carriers approach conversion.

By training classifiers on presymptomatic FTD mutation carriers and controls, we obtained a unimodal carrier-control model based on the RD feature and several multimodal carrier-control models that significantly outperformed chance level. This suggests that classification models should be trained using early-stage FTD patients or presymptomatic FTD mutation carriers instead of

advanced FTD cases, in order to be sensitive to early-stage FTD pathology. Furthermore, our carrier-control models demonstrate that MRI-based machine learning is powerful enough to detect subtle pathological changes associated with FTD even before symptom onset and on a single-subject level. Although classification performance beyond chance level is an important finding, it must be noted that AUCs of 0.642 and 0.684 are modest and far from sufficient for diagnostic use in the clinic. This is at least partly explained by our heterogeneous sample, as we included mutation carriers of several genes in order to obtain sufficient sample size for robust cross-validation. Heterogeneity further arose from the uncertain time to onset in our sample. Investigating a uniform population a few years before symptom onset might lead to higher classification performance, but these data were not available to us.

On a pathological level, it has been argued that neurodegeneration in FTD starts in the white matter (Rohrer et al., 2013; Suri et al., 2014; Möller et al., 2016; Canu et al., 2017). Our results support this hypothesis, as the only unimodal model that outperformed chance was based on the RD feature, which was furthermore included in all multimodal models that significantly outperformed chance. Additionally, all models that outperformed chance level were based on DTI features, while three models also included WMD. This means that our carrier-control model was able to combine subtle white matter differences from the diffusion-weighted scans and the structural $3DT_1w$ scan to classify a subject as mutation carrier or healthy control. In our sample, inclusion of GMD and functional connectivity features did not aid classification, but in fact drove down classification performance. For instance, adding FCor or PCor to the model resulted in classification scores near 0.5. This reinforces the notion that feature selection is important for MRI-based machine learning.

In addition to the uncertain time to onset, there were several other limitations. Firstly, the bvFTD model was trained on a relatively small sample of 23 bvFTD patients and 35 controls. A model based on a larger sample might capture the heterogeneity of bvFTD pathology more completely, which could benefit generalisation to our presymptomatic sample. Furthermore, the model was trained on sporadic bvFTD patients, while it was applied to *MAPT* mutation carriers, *GRN* mutation carriers, and *C9orf72* repeat expansion carriers. Since correlations between genetics, pathology and phenotype are not fully elucidated (Mann & Snowden, 2017), care must be taken not to overinterpret our results. Specifically, pathological changes associated with non-behavioural variants (Seelaar et al., 2011) may be insufficiently recognised by the bvFTD model. Lastly, we used nested cross-validation to estimate out-of-sample performance for the carrier-control model, which minimises prediction bias (Kriegeskorte et al., 2009). Still, measuring performance on a separate validation cohort would further increase the validity of this study.

Conclusion

Our data show that presymptomatic FTD mutation carriers can be distinguished from healthy controls on an individual level using a new multimodal MRI-based carrier-control classification model, while this was not possible using a recent bvFTD classification model. A multimodal MRI-based classification score may therefore be a useful biomarker to aid earlier FTD diagnosis. Successful single-subject recognition of early-stage or presymptomatic FTD may facilitate more precise subject recruitment into clinical trials. Furthermore, our multimodal MRI-based carrier-control classification model supports the hypothesis that FTD-related neurodegeneration starts in the white matter.

Acknowledgements

Our appreciation goes out to all FTD-RisC participants and families. The authors of this work were supported by the Leiden University Medical Centre MD/PhD Scholarship (to RAF), ZonMw programme Memorabel project 733050103, JPND PreFrontAls consortium project 733051042 (to JCvS), and a VICI grant 016–130–667 from The Netherlands Organisation for Scientific Research (NWO; to SARBR). The views expressed are those of the authors and not necessarily those of the funding sources. The funding sources were not involved in the design of the study; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication. The authors report no conflict of interest.