



Universiteit
Leiden
The Netherlands

Neuroimaging biomarkers in genetic frontotemporal dementia : towards a timely diagnosis

Feis, R.A.

Citation

Feis, R. A. (2020, October 14). *Neuroimaging biomarkers in genetic frontotemporal dementia : towards a timely diagnosis*. Retrieved from <https://hdl.handle.net/1887/137726>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/137726>

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

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/137726> holds various files of this Leiden University dissertation.

Author: Feis, R.A.

Title: Neuroimaging biomarkers in genetic frontotemporal dementia : towards a timely diagnosis

Issue Date: 2020-10-14



Chapter 4

Frontotemporal dementia mutation carriers: symptom onset prediction using MRI-based and cognitive features

Published in *Brain Communications* 2020;2(2):fcaa079 as:

Classification using fractional anisotropy predicts conversion in genetic frontotemporal dementia, a proof of concept

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

Abstract

Frontotemporal dementia (FTD) is a highly heritable and devastating neurodegenerative disease. About 10–20% of all FTD is caused by known pathogenic mutations, but a reliable tool to predict clinical conversion in mutation carriers is lacking. In this retrospective proof-of-concept case-control study, we investigate whether magnetic resonance imaging (MRI)-based and cognition-based classifiers can predict which mutation carriers from genetic frontotemporal dementia families will develop symptoms ('convert') within 4 years.

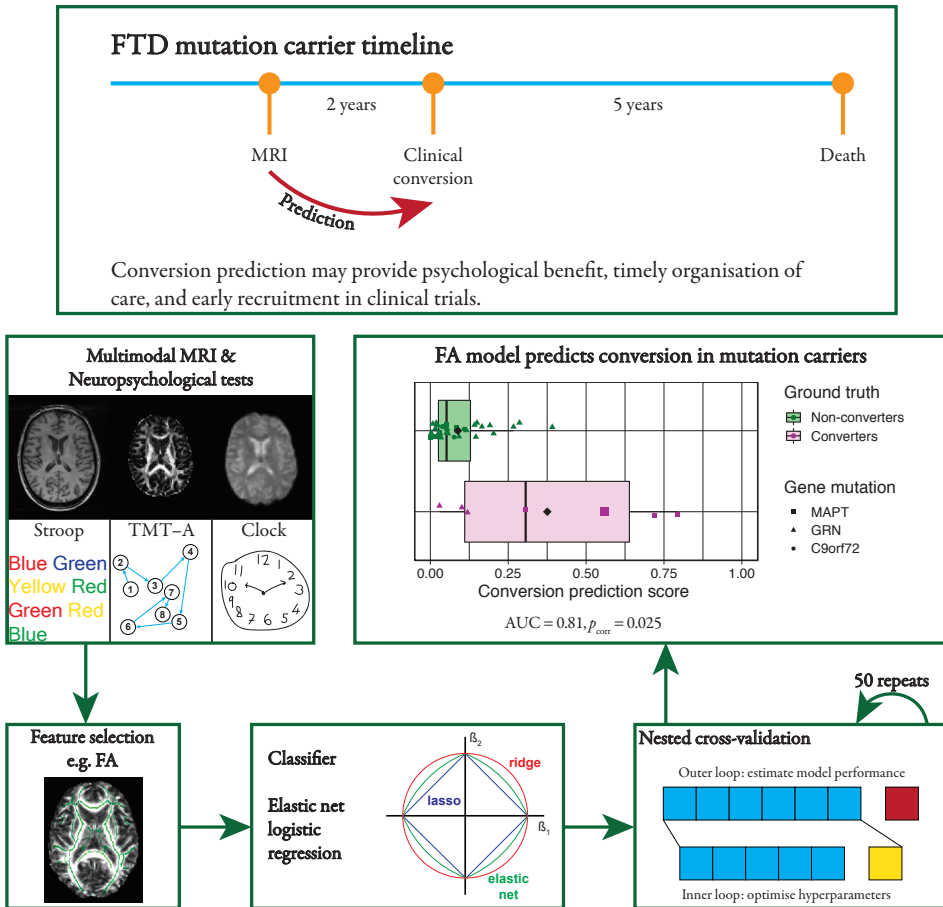
From genetic FTD families, we included 42 presymptomatic FTD mutation carriers. We acquired anatomical, diffusion-weighted imaging, and resting-state functional MRI, as well as neuropsychological data. After 4 years, seven mutation carriers had converted to FTD ('converters'), while 35 had not ('non-converters'). We trained regularised logistic regression models on baseline MRI and cognitive data to predict conversion to FTD within 4 years, and quantified prediction performance using area under the receiver operating characteristic curves (AUC).

The prediction model based on fractional anisotropy, with highest contribution of the forceps minor, predicted conversion to FTD beyond chance level (AUC = 0.81, family-wise error corrected $p = 0.025$ vs. chance level). Other MRI-based and cognitive features did not outperform chance level.

Even in a small sample, fractional anisotropy predicted conversion in presymptomatic FTD mutation carriers beyond chance level. After validation in larger data sets, conversion prediction in genetic FTD may facilitate early recruitment into clinical trials.

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

Graphical abstract



AUC, area under the receiver operating characteristic curve; C9orf72, chromosome 9 open reading frame 72; FA, fractional anisotropy; FTD, frontotemporal dementia; GRN, progranulin; MAPT, microtubule-associated protein tau; MRI, magnetic resonance imaging; p_{corr} , family-wise error rate corrected p -value; TMT-A, trail making test A.

Introduction

Frontotemporal dementia (FTD) is a highly heritable and devastating neurodegenerative disease that often occurs at a presenile age (Ratnavalli et al., 2002; Harvey et al., 2003; Hogan et al., 2016). Patients typically present with behavioural symptoms (behavioural variant FTD; bvFTD; Rascovsky et al., 2011) or with language disorders (primary progressive aphasia; PPA; Gorno-Tempini et al., 2011), but may also develop amyotrophic lateral sclerosis (ALS; Lomen-Hoerth et al., 2002) or parkinsonian syndromes such as corticobasal syndrome (CBS) and progressive supranuclear palsy (PSP; Josephs et al., 2006). About 10–20% of all FTD is caused by three known pathogenic mutations: microtubule-associated protein tau (*MAPT*), progranulin (*GRN*), and chromosome 9 open reading frame 72 (*C9orf72*) repeat expansion (Seelaar et al., 2008; Rohrer et al., 2009; Benussi et al., 2015). These mutations have a high penetrance and an autosomal dominant inheritance pattern. As such, roughly 50% of family members from a genetic FTD pedigree will develop an FTD variant.

Uncertainty about if and when a person will develop FTD is a major burden (Riedijk et al., 2009; Cohn-Hokke et al., 2018), and some subjects undergo genetic testing to eliminate this uncertainty. However, when a mutation is found, this raises the question when symptoms will emerge. A reliable prognosis may provide psychological benefit, and contributes to the timely organisation of adequate care. Moreover, conversion prediction in genetic FTD may facilitate early recruitment into clinical trials for disease modifying treatments, which could lead to improved trial efficacy (Tsai & Boxer, 2016). However, accurate conversion prediction is currently not possible. Expected time to symptom onset, calculated using the ages of FTD onset of family members, has been used as proxy in research settings (Rohrer et al., 2015; Jiskoot et al., 2018a), but these estimates are inaccurate due to the large variation in age of onset between and within genetic FTD families (van Swieten & Heutink, 2008; Jiskoot et al., 2018a). A more reliable clinical tool is needed to predict conversion in genetic FTD.

Magnetic resonance imaging (MRI)-based classification is a promising diagnostic biomarker for FTD (Raamana et al., 2014; Klöppel et al., 2015; Koikkalainen et al., 2016; Bron et al., 2017; Meyer et al., 2017; Bouts et al., 2018), and distinguishes presymptomatic FTD mutation carriers from controls beyond chance (Feis et al., 2019a). Classification with cognitive measures has also been advocated as diagnostic FTD biomarker (Wang et al., 2016). Therefore, MRI- and cognition-based classification may be suitable candidates for conversion prediction in genetic FTD. In recent observational studies, FTD mutation carriers were followed up until conversion, and multimodal MRI data were analysed using mass univariate techniques such as voxel-based morphometry and diffusion tensor imaging (DTI; Jiskoot et al., 2019), or using classification scores calculated from multimodal MRI and a classification model based on FTD patients (Feis et al., 2019b). These studies suggest that brain changes as found on MRI appear relatively explosively in the final years before symptom onset. However, it is yet unknown whether conversion in such subjects could have been predicted based on pre-conversion MRI scans or neuropsychological data.

Here, we study whether MRI-based features and cognitive measures have predictive value for FTD conversion in genetic FTD. To this end, we train prediction models on MRI-based features and cognitive features to predict which presymptomatic FTD mutation carriers develop symptoms within 4 years ('converters'), and which mutation carriers do not ('non-converters').

Methods

Participants

This retrospective case-control study included FTD mutation carriers from the FTD-Risk Cohort (FTD-RisC; Dopper et al., 2014; Meeter et al., 2016a; Papma et al., 2017; Jiskoot et al., 2016, 2018b, 2019), a longitudinal study that follows healthy, 50% at-risk family members of genetic FTD patients on a 2-year basis. For the current study, we included the data of 43 FTD mutation carriers, who entered FTD-RisC between May 2010 and November 2014. Inclusion criteria were age at MRI scan between 40 and 70 years old, availability of a 3-dimensional T_1 -weighted MRI (3DT_{1w}) scan, a diffusion-weighted imaging (DWI) data set, a resting-state functional MRI T_2^* -weighted (rs-fMRI) scan, and availability of clinical follow-up after 2 and 4 years. Genotyping, as described in previous work (Dopper et al., 2014; Jiskoot et al., 2016), revealed nine *MAPT* mutation carriers, 28 *GRN* mutation carriers, and six *C9orf72* repeat expansion carriers. For seven FTD mutation carriers who had converted to FTD (four *MAPT* and three *GRN* mutation carriers; ‘converters’; see ‘Conversion’ section below), we included the MRI data of their last visit before conversion. For one converter (*MAPT* mutation carrier), these MRI data were excluded due to the presence of artefacts, and we included MRI data of the previous visit. The time between MRI and conversion ranged between 11 and 41 months (mean = median = 22 months, SD = 10.5 months). More details on this group’s demographics are provided in there results. The remaining mutation carriers were confirmed to be asymptomatic after 4 years follow-up in accordance with established diagnostic criteria for bvFTD (Rascovsky et al., 2011), PPA (Gorno-Tempini et al., 2011), and ALS (Ludolph et al., 2015), and were termed ‘non-converters’. Follow-up after 6 years was not complete for this group: 25 non-converters were followed up after 6 years and all remained asymptomatic at that point. For non-converters, we included the FTD-RisC baseline MRI data. One non-converter with a *GRN* mutation was excluded due to incomplete neuropsychological data. Other exclusion criteria were: current or past neurologic (other than dementia) or primary psychiatric disorders, history of drug abuse, large image artefacts, and gross brain pathology other than atrophy.

Standard protocol approvals, registrations, and patient consents

Participants and clinical investigators were blinded to the participants’ genetic status, except for those that underwent predictive testing at their own request. For converters, genetic counselling was offered to the patient and family members, and genetic status was unblinded to confirm the presence of the pathogenic mutation. 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 (Feis et al., 2019b).

Conversion

Conversion was determined in a multidisciplinary consensus meeting of the Erasmus Medical Centre FTD Expertise Centre, involving neurologists (JCvS), neuropsychologists (JLP, JMP, LCJ), neuroradiologists, a clinical geneticist, and a care consultant (Feis et al., 2019b). In the consensus meetings, information from the medical history, neuropsychological assessment, and MRI of the brain were reviewed. The timing of symptom onset was estimated from heteroanamnesic information provided by knowledgeable informants (e.g., siblings, spouses). After the clinical diagnosis was

made, and if the subject and family agreed, genetic status was unblinded for confirmation. Four *MAPT* mutation carriers and one *GRN* mutation carrier converted to bvFTD, while two *GRN* mutation carriers converted to non-fluent variant PPA (nfvPPA). BvFTD converters presented with progressive behavioural deterioration, functional decline, and frontal and/or temporal lobe atrophy on structural MRI, fulfilling the international diagnostic consensus criteria for bvFTD with definite frontotemporal lobar degeneration pathology (Rascovsky et al., 2011). The two nfvPPA converters presented with isolated language difficulties and no impairment in daily living activities. Both showed a non-fluent, halting speech with sound errors and agrammatism, fulfilling the diagnostic criteria for nfvPPA (Gorno-Tempini et al., 2011). For more detailed information on conversion criteria, and for a full description of the converters' clinical profile, see Jiskoot et al. (2018b, 2019).

MRI data acquisition

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 (Feis et al., 2019a). 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 diffusivity, and rs-fMRI sequence 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. Scan parameters are provided in **Table 4.1**.

Table 4.1 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	$0.88 \times 0.88 \times 1.20$	$2.00 \times 2.00 \times 2.00$	$2.75 \times 2.75 \times 2.99^b$
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 b_0 image.

^b Including 10% interslice gap.

Image preprocessing

Preprocessing of MRI scans was performed similarly to previous work (Bouts et al., 2018; Feis et al., 2019a, b). All registration and segmentation steps were critically reviewed and errors were corrected accordingly.

For $3DT_1w$ images, we performed bias field correction (N4ITK; Tustison et al., 2010), brain extraction (FSL BET; Smith, 2002), nonlinear registration to the MNI152 $2 \times 2 \times 2$ mm T_1 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). Our choice to use SPM segmentation was based on the segmentation tool comparison by Kazemi & Noorizadeh, 2014.

Preprocessing of DWI data sets included correction of motion and eddy-current induced distortions (eddy correct; Leemans & Jones, 2009), and voxel-wise calculation of the measures fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AxD; largest eigenvalue), and radial diffusivity (RD; average of the two remaining eigenvalues) using 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 that extracted values originate from white matter.

For rs-fMRI data, preprocessing consisted of 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, followed by automatic identification and removal of motion artefacts using ICA-AROMA (version 0.3 beta; Pruim et al., 2015). 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 3DT₁w image using boundary-based registration (Greve & Fischl, 2009). Next, resulting registration parameters were concatenated to the 3DT₁w-to-MNI152 template registration parameters to obtain the final registration parameters.

MRI feature selection

Selection of MRI features was performed as described previously (Bouts et al., 2018; Feis et al., 2019a).

Cortical grey matter density (GMD) and white matter density (WMD) were calculated as weighted means of their respective regional grey matter or white matter probability (SPM segmentation) weighted by the probability of a voxel belonging to a specific cortical region. Cortical probabilities were derived from the 48 Harvard-Oxford probabilistic anatomical brain atlas (split into left and right), and white matter probabilities were derived from the Johns-Hopkins University white matter tractography atlas for 20 white matter tracts. Voxels with atlas probability values under 25% were excluded. We estimated these features in native space by transforming the atlas masks to the images' native space using the inverted non-linear registration parameters. For deep grey matter regions, GMD values were calculated as the regions' volume (FIRST segmentations) divided by the total intracranial volume. This resulted in a GMD feature vector of 110 weighted mean GMD values (48 left cortical, 48 right cortical, and 14 deep grey matter regions) and in a feature vector of 20 weighted mean WMD values per subject. While grey matter volume, as estimated by voxel-based morphometry, would also be a useful feature for prediction, we chose the simpler GMD for generalisation purposes, as it is not reliant on a study-specific template.

DTI features were calculated as weighted mean FA, MD, AxD, or RD values per tract per subject. First, we projected each subject's FA, MD, AxD, and RD values onto the TBSS group skeleton on a voxel-wise basis. Next, we weighted the mean values per tract by the probability of a voxel belonging to that specific tract, derived from 20 tracts of the Johns-Hopkins University white matter tractography atlas. Voxels with atlas probability values under 25% were excluded. This resulted in

four feature vectors of each 20 weighted mean values per subject.

To calculate functional connectivity features, all processed rs-fMRI images were combined in a temporally concatenated group-level ICA (MELODIC; Beckmann & Smith, 2004), with dimensionality fixed at 20 components and an ICA threshold of 0.99 (Smith et al., 2013). This means 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 regions. Correlations between the mean time courses of all pairs of components were subsequently calculated. Functional connectivity was calculated as full correlations (FCor), and as sparse, L1-regularised, partial correlations (PCor) between the mean time courses of all pairs of components. Partial correlations were calculated using the graphical lasso algorithm (J. Friedman et al., 2008). This procedure resulted in two feature vectors of each $(20 \times 19) \div 2 = 190$ (partial) correlations per subject.

Finally, we created one multimodal feature vector by concatenating all feature vectors together. As such, we had nine feature vectors in total: two 3DT_{1w} feature vectors, four DTI feature vectors, two rs-fMRI feature vectors, and one multimodal imaging feature vector.

Neuropsychological assessment and cognitive feature selection

We screened global cognitive functioning by means of the Mini-Mental State Examination (MMSE; Folstein et al., 1975). Experienced neuropsychologists (JLP, JMP, LCJ) administered neuropsychological tests within six cognitive domains: language, attention and mental processing speed, executive functioning, social cognition, memory, and visuoconstruction. The language domain was assessed using the 60-item Boston naming test (BNT; Kaplan et al., 1978), verbal semantic association test (SAT; Visch-Brink et al., 2005), ScreeLing phonology (Doesborgh et al., 2003), and categorical fluency (Thurstone & Thurstone, 1962). The domain of attention and mental processing speed was evaluated using the trail making test (TMT-)A (Army Individual Test Battery, 1944), the mean of Stroop colour-word tests I and II (Stroop, 1935), Wechsler adult intelligence scale III (WAIS-III) digit span forwards (Wechsler, 2005), and letter digit substitution test (LDST; Jolles et al., 1995). We measured the executive functioning domain using TMT-B (Army Individual Test Battery, 1944), Stroop colour-word test III (Stroop, 1935), WAIS-III digit span backwards (Wechsler, 2005), modified Wisconsin card sorting test (WCST) concepts (Nelson, 1976), letter fluency (Thurstone & Thurstone, 1962), and WAIS-III similarities (Wechsler, 2005). The social cognition domain was assessed by means of Happé theory of mind cartoons, Happé non-theory of mind cartoons (Happé et al., 1999), and Ekman faces (Ekman & Friesen, 1976). We evaluated the memory domain using the immediate response and recall of the Dutch Rey auditory verbal learning test (RAVLT; Rey, 1958), and the visual association test (VAT; Lindeboom et al., 2002). Lastly, we tested the visuoconstruction domain by means of clock drawing (Royall et al., 1998) and WAIS-III block design (Wechsler, 2005).

Raw test scores were added into one cognitive feature vector per domain. This resulted in a language feature vector with four features, an attention feature vector with four features, an executive feature vector with six features, a social feature vector with three features, a memory feature vector with three features, and a visuoconstruction feature vector with two features. Finally, we added these feature vectors together to create a multidomain feature vector, bringing the total number of feature vectors on seven: language, attention, execution, social, memory, visuoconstruction, and multidomain.

Classifier

In order to identify future converters ($n = 7$) from non-converters ($n = 35$) at baseline, we trained prediction models using MRI features ($n = 9$) and, separately, cognitive features ($n = 7$). 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; Feis et al., 2019a). 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 (i.e., L1 penalty) and for the sum of the squared value of the coefficients (i.e., L2 penalty). 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 in the model without penalty to ensure that estimated feature regression coefficients were conditional on subject age and sex. A prediction score of 0 represented a non-converter and 1 represented a converter.

Cross-validation

We trained our conversion prediction models in a stratified nested seven-fold cross-validation scheme to make sure that the proportion of converters and non-converters was the same in each fold (i.e., one converter and five non-converters per fold). In the outer loop, one part of the data (i.e., one of the seven folds) was set apart as a test set and served to test the generalised prediction performance of the elastic net regression model. The remaining parts (six of the seven folds) were used to train the model. Within the training set of the outer loop, we performed a nested cross-validation to optimise the model's hyperparameters without overestimating prediction performance (Varma & Simon, 2006; Kriegeskorte et al., 2009). The resulting optimal hyperparameters were used in the training set of the outer loop to train the model, and the prediction performance was then tested in the test set of the outer loop. This process was repeated seven times to make sure that each subject was part of the test set exactly once. Since the test set of the outer loop was neither used for model training, nor for parameter optimisation, we reduced the risk of overestimating the generalisation performance as much as possible (Kriegeskorte et al., 2009; Schouten et al., 2016; Bouts et al., 2018; Feis et al., 2019a). The entire prediction procedure was repeated 50 times to average prediction outcome variability resulting from random partitioning in training and test folds. All prediction analyses and evaluations were implemented in R version 3.3.2 (R core 2016, GLMnet package; Friedman et al., 2010).

Prediction performance

For both analyses, we quantified prediction performances using receiver operating characteristic curves. Receiver operating characteristic curves were calculated by shifting the threshold for predicting an individual as converter from 0 to 1, and plotting the true positive rate (sensitivity) vs. the false positive rate ($1 - \text{specificity}$) for each intermediate point. The area under this receiver operating characteristic curve (AUC) is a measure of prediction 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 prediction accuracy, given equal class distribution and equal penalty for false positive and false negative predictions. We averaged AUC, accuracy, sensitivity, and specificity values from the 50 times repeated nested cross-validations (Schouten et al., 2016; Bouts et al., 2018; Feis et al., 2019a).

Statistical analysis

Statistical group analyses of demographic data were performed using R (R Core 2016, Vienna, Austria). We tested for differences between converters and non-converters using unpaired t -tests (age and education), the Mann-Whitney U test (MMSE scores [0–30]) and the chi-square test (sex distribution).

To compare MRI and cognitive prediction models' AUC values vs. chance level, we used permutation tests ($N = 5,000$; Noirhomme et al., 2014). We used the maximum t -statistic method to correct for multiple comparisons within, respectively, the nine MRI-based and the seven cognitive prediction analyses. For each permutation, we calculated the maximum absolute t -statistic within the family of tests, which resulted in a maximum t -distribution of 5,000 maximum absolute t -statistics. The observed t -statistic of each analysis was then compared to this maximum t -distribution in order to obtain a family-wise error rate corrected p -value. The alpha level required for statistical significance was set at 0.05.

Data availability

Raw data were generated at the Leiden University Medical Centre. The derived data, as well as scripts, that support the findings of this study are available from the corresponding author upon request.

Results

Demographics

Seven converters and 35 non-converters met the inclusion criteria (**Table 4.2**). At the time of MRI scan, converters did not differ from non-converters in terms of age ($p = 0.85$), sex distribution ($p = 0.40$), and MMSE scores ($p = 0.50$). However, converters had higher levels of education than non-converters ($p = 0.015$). More information on the converter group is shown in **Table 4.3**.

Table 4.2 Participant demographics

	Converters ($n = 7$) ^a	Non-converters ($n = 35$)	p -value
Age, mean (SD) years	51.7 (8.8)	51.0 (8.4)	0.85
Sex, n (%) ♀	4 (57%)	26 (74%)	0.40
Education, mean (SD) years ^b	15.2 (0.8)	13.7 (2.9)	0.015
MMSE, median (range) points	29 (27–30)	30 (24–30)	0.50

MMSE, Mini-Mental State Examination.

^a 4 microtubule-associated protein tau mutation carriers, 3 progranulin mutation carriers.

^b Education values were missing for one converter.

Table 4.3 Converter demographics

	Gene	Mutation	Age at MRI	Sex	Months to conversion	FTD variant	Prediction score ^a
Converter 1	<i>GRN</i>	S82VfsX174	65	Female	23	bvFTD	0.12
Converter 2	<i>MAPT</i>	P301L	54	Female	28	bvFTD	0.31
Converter 3	<i>MAPT</i>	P301L	57	Male	18	bvFTD	0.79
Converter 4	<i>GRN</i>	S82VfsX174	56	Female	11	nfvPPA	0.10
Converter 5	<i>GRN</i>	S82VfsX174	49	Female	22	nfvPPA	0.03
Converter 6	<i>MAPT</i>	G272V	41	Male	41 ^b	bvFTD	0.56
Converter 7	<i>MAPT</i>	G272V	42	Male	11	bvFTD	0.72

bvFTD, behavioural variant FTD; FTD, frontotemporal dementia; *GRN*, progranulin; *MAPT*, microtubule-associated protein tau; MRI, magnetic resonance imaging; nfvPPA, non-fluent variant primary progressive aphasia.

^a Prediction scores based on the FA prediction model range from 0 to 1, with 0 representing non-converters and 1 representing converters.

^b MRI data nearer to symptom onset was excluded due to artefacts for this subject.

Prediction performance

Prediction with MRI features yielded mixed results (**Table 4.4**). The model based solely on FA features (FA model) significantly outperformed chance level with an AUC of 0.81 ($p = 0.025$ vs. chance level). For the FA model, subjects' prediction scores are shown in **Figure 4.1**. Converters with a *MAPT* mutation seemed to have higher conversion prediction scores than those with a *GRN* mutation. Similarly, male mutation carriers seemed to have higher prediction scores than female

mutation carriers. There seemed to be no clear correlation (Pearson's $r = 0.02$) between conversion prediction scores and time from MRI to conversion. However, the above three observations should be noted with care, since the small sample size did not permit for meaningful statistical testing.

The FA model's beta weights for the 50 cross-validation repeats are shown in **Figure 4.2**. The highest beta weights were found in the forceps minor. Other white matter tracts with above average beta weights included the forceps major, right corticospinal tract, right inferior longitudinal fasciculus, right anterior thalamic radiation, and right uncinate fasciculus. For the two WM tracts with the largest beta weights (i.e., forceps minor and major), we plotted subjects' mean FA value across the tract in **Figure 4.3**. Corresponding with the beta weights, the differences between converters and non-converters were larger for the forceps minor than for the forceps major.

The other DTI features (i.e., MD, AxD, RD) did not outperform chance level, nor did grey and white matter density features, and functional connectivity features (i.e., FCor, PCor). Concatenating all features together in a multimodal MRI model did not improve performance compared to the FA model (AUC = 0.63, $p = 0.55$ vs. chance level).

Prediction with cognitive features resulted in poor performances (**Table 4.5**). The best performing cognitive model was based on the executive domain (AUC = 0.35). Adding multiple domains together in a multidomain cognitive model did not improve performance (AUC = 0.34).

Table 4.4 MRI features' performance

MRI modality	AUC	Min – max	Sensitivity	Specificity	Accuracy	FWERC p -value (AUC > chance)
GMD	0.668	0.563 – 0.731	0.746	0.599	0.623	0.375
WMD	0.438	0.306 – 0.567	0.571	0.517	0.526	0.976
<i>EA</i>	<i>0.812</i>	<i>0.608 – 0.906</i>	<i>0.769</i>	<i>0.761</i>	<i>0.762</i>	<i>0.025</i>
MD	0.677	0.616 – 0.743	0.711	0.693	0.696	0.322
AxD	0.435	0.327 – 0.563	0.626	0.491	0.514	0.977
RD	0.625	0.465 – 0.722	0.689	0.630	0.640	0.590
FCor	0.336	0.204 – 0.449	0.446	0.513	0.501	1.000
PCor	0.403	0.233 – 0.502	0.514	0.527	0.525	0.996
Multimodal	0.626	0.547 – 0.710	0.557	0.781	0.744	0.552

Converters (seven presymptomatic FTD-RisC mutation carriers that developed symptoms within 4 years after assessment) vs. non-converters (35 presymptomatic FTD-RisC mutation carriers that remained cognitively healthy after 4 years). Multimodal represents a combination of all MRI features. Min – max AUC values represent the variance across the 50 repeats. *Italic*: mean AUC significantly higher than chance level after family-wise error rate correction.

AUC, area under the receiver operating characteristic curve; FA, fractional anisotropy; FCor, full correlations between 20 ICA components; FTD-RisC, frontotemporal dementia Risk Cohort; FWERC, family-wise error rate corrected; GMD, grey matter density; ICA, independent component analysis; MD, mean diffusivity; PCor, L1-regularised partial correlations between 20 ICA components; RD, radial diffusivity; WMD, white matter density.

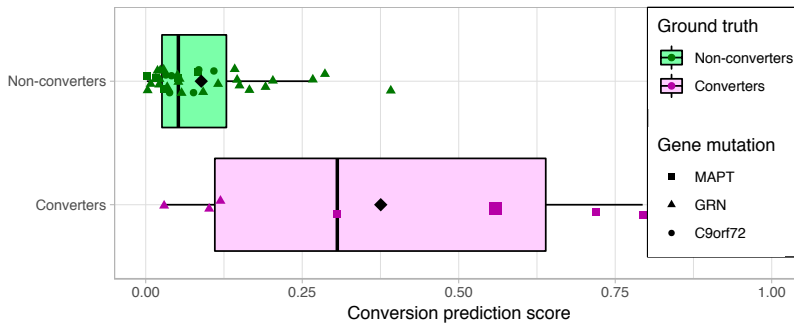


Figure 4.1 Converters' and non-converters' conversion prediction scores

Box and scatter plot of each subject's conversion prediction score on a scale from 0 (representing non-converter) to 1 (representing converter) after applying the FA model. Different gene mutations were represented with different shapes. The converter with the longest time between MRI and conversion (i.e., 41 months) was annotated with increased size. These conversion prediction scores result in a performance of 0.81 AUC for the FA model ($p = 0.025$ vs. chance level). *C9orf72*, chromosome 9 open reading frame 72; FA, fractional anisotropy; *GRN*, progranulin; *MAPT*, microtubule-associated protein tau.

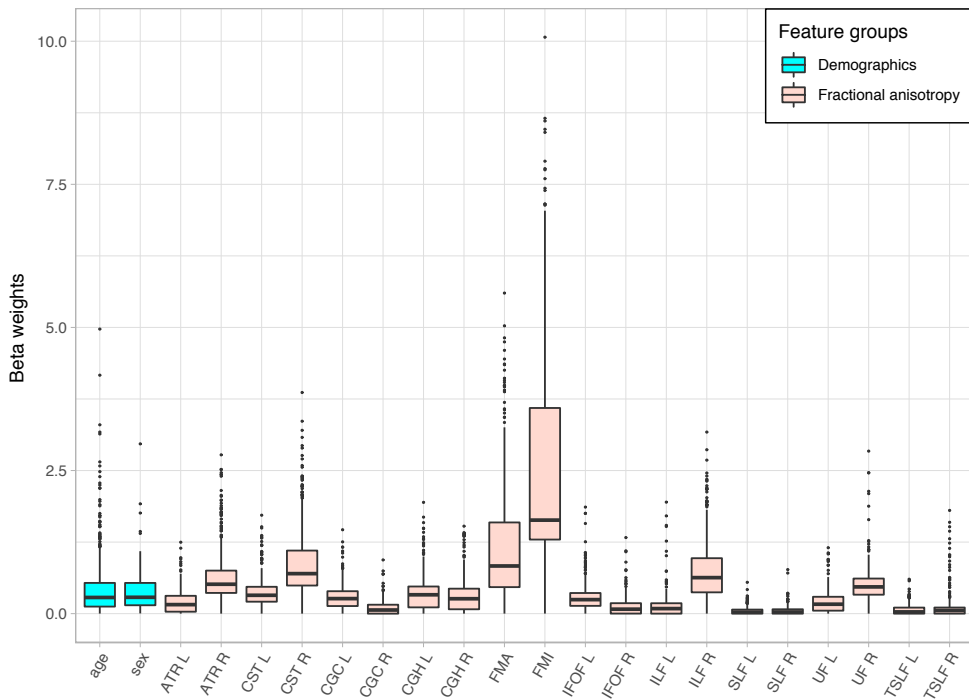


Figure 4.2 FA model beta weights

Box plots show the FA model's standardised beta weights for the 50 cross-validation repeats. Demographics (age and sex, in blue) were included in the model without penalty, while the FA features (red) were regularised.

ATR, anterior thalamic radiation; CST, corticospinal tract; CGC, cingulum in the cingulate gyrus area; CGH, cingulum in the hippocampal area; FA, fractional anisotropy; FMA, forceps major; FMI, forceps minor; IFOF, inferior fronto-occipital fasciculus; ILF, inferior longitudinal fasciculus; L, left; R, right; SLF, superior longitudinal fasciculus; TSLF, temporal projection of the SLF; UF, uncinate fasciculus.

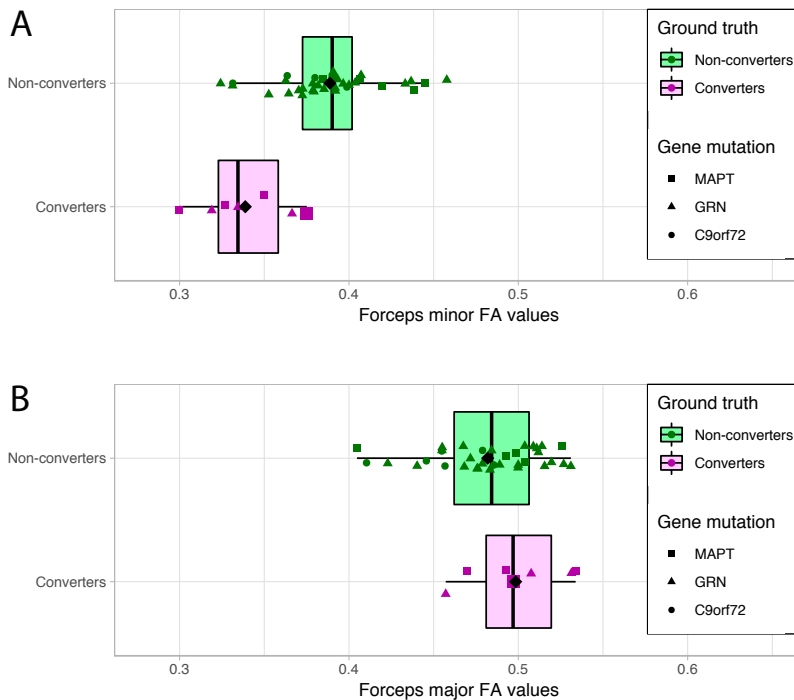


Figure 4.3 Mean FA values in two white matter tracts

Box and scatter plot of each subject's mean FA value in the forceps minor (A), and forceps major (B). Different FTD gene mutations were represented with different shapes. The converter with the longest time between MRI and conversion (i.e., 41 months) was annotated with increased size.

C9orf72, chromosome 9 open reading frame 72; FA, fractional anisotropy; *GRN*, progranulin; *MAPT*, microtubule-associated protein tau.

Table 4.5 Cognitive features' performance

Cognitive domain	AUC	Min – max	Sensitivity	Specificity	Accuracy	FWERC <i>p</i> -value (AUC > chance)
Language	0.282	0.180 – 0.371	0.394	0.515	0.495	0.999
Attention	0.343	0.208 – 0.441	0.517	0.451	0.462	0.992
Executive	0.349	0.184 – 0.482	0.534	0.439	0.455	0.987
Social	0.344	0.208 – 0.449	0.474	0.488	0.486	0.992
Memory	0.272	0.143 – 0.359	0.491	0.414	0.427	0.999
Visuoconstruction	0.288	0.186 – 0.424	0.429	0.477	0.469	0.999
Multidomain	0.340	0.208 – 0.449	0.477	0.481	0.480	0.994

Converters (seven presymptomatic FTD-RisC mutation carriers that developed symptoms within 4 years after assessment) vs. non-converters (35 presymptomatic FTD-RisC mutation carriers that remained cognitively healthy after 4 years). Multidomain represents a combination of all cognitive features. Min – max AUC values represent the variance across the 50 repeats.

AUC, area under the receiver operating characteristic curve; FTD-RisC, frontotemporal dementia Risk Cohort; FWERC, family-wise error rate corrected.

Discussion

This study describes conversion prediction within 4 years in FTD mutation carriers using MRI-based and cognitive measures. We found that it was possible to distinguish converters from non-converters beyond chance level with an AUC of 0.81 using FA. Other MRI-based and cognitive measures did not outperform chance level.

The FA model's performance reaffirms the potential of diffusion scans for early FTD diagnosis. Group differences in FA and other DTI metrics exist in presymptomatic FTD mutation carriers compared to controls (Borroni et al., 2008; Dopper et al., 2014; Pievani et al., 2014; Lee et al., 2017; Papma et al., 2017; Bertrand et al., 2018; Jiskoot et al., 2018a, 2019). More specifically with regard to conversion, two recent FTD-RisC studies observed how mass univariate MRI measures and MRI-based classification scores develop as subjects approach conversion, as compared to subjects who did not convert during follow-up. Jiskoot et al. found that grey matter atrophy and white matter DTI changes become noticeable around two years before symptom onset, and longitudinal FA decline was primarily located in the genu of the corpus callosum (Jiskoot et al., 2019), the area with the highest beta values in our current study. Feis et al. combined multimodal MRI to calculate a single classification score at each time point using a classification model that was trained on bvFTD patients and controls. They found that the classification scores of presymptomatic mutation carriers and non-carriers did not differ over time, but that the classification scores of converters rose faster than those of non-converting mutation carriers (Feis et al., 2019b). Together, these studies suggest that FTD-related deterioration on MRI accelerates when subjects are near conversion. However, they fail to address the most important question: is it possible, using presymptomatic cross-sectional data, to predict future conversion in individual FTD mutation carriers? We expand on this by showing that individual differences in FA have predictive value for FTD onset within 4 years in FTD mutation carriers.

Within the FA model, the forceps minor feature had the highest beta weights across 50 repeats, suggesting that this area was important for prediction. Accordingly, the difference in mean FA between converters and non-converters was larger in the forceps minor than in other tracts, e.g., the forceps major. The genu of the corpus callosum (i.e., part of the forceps minor) was recently found to be the most consistent white matter region in terms of DTI changes across FTD subtypes (Elahi et al., 2017). As our converter group included both bvFTD and nfvPPA converters, this might explain why the forceps minor had high beta weights, and could provide biological support for our model. Forceps minor changes may therefore be a promising marker for early FTD detection, although it might not discriminate between the different FTD subtypes. Other white matter tracts are affected more specifically in the different clinical syndromes. BvFTD is generally characterised by DTI changes in the anterior thalamic radiation, cingulum, and uncinate fasciculus (Zhang et al., 2011; Mahoney et al., 2014; Möller et al., 2015; Daianu et al., 2016), while the superior longitudinal fasciculus and corticospinal tract are typically affected in nfvPPA (Whitwell et al., 2010; Galantucci et al., 2011; Omer et al., 2017), and the inferior longitudinal and uncinate fasciculi show DTI changes in svPPA (Whitwell et al., 2010; Galantucci et al., 2011; Tu et al., 2015). Future studies should investigate whether this combination of shared and different white matter DTI changes between the FTD subtypes facilitate hierarchical classification. Although it is interesting to examine which features were important for prediction, care must be taken not to overinterpret the beta weights. Contrary to explanatory regression models, the beta weights of our regularised prediction model do not designate direct relationships between the features and prediction score, nor do they reflect mean differences between the groups (Shmueli, 2010). Each feature's effect is

conditional on the effects of all other features in the model, and multicollinearity between features may result in suppression of the effect of some features. For example, we cannot be sure whether the higher beta weights in the FA model for right hemispheric white matter tracts point towards asymmetric effects—such as are reported in *GRN* mutation carriers (Jiskoot et al., 2018a)—or whether the corresponding regions in the left hemisphere added redundant information, and were therefore suppressed in the model.

Apart from the FA model, no other MRI- or cognition-based prediction model significantly outperformed chance. However, it would be inappropriate to draw conclusions based on the non-significant results. Due to our small sample size, it is unclear whether these features were uninformative, or whether the study was underpowered to find significant results. Grey matter atrophy occurs at a later stage than white matter changes in *MAPT* and *GRN* mutation carriers (Borroni et al., 2008; Agosta et al., 2012; Rohrer & Rosen, 2013; Feis et al., 2019a; Jiskoot et al., 2019), but the question whether or not that precludes the use of grey matter features for early diagnosis requires further investigation. Similarly, we cannot comment on why the remaining diffusion features MD, AxD, and RD, which often yield comparable results to FA, did not outperform chance level. None of the cognitive features predicted conversion within 4 years. However, it should be noted that our selection of neuropsychological tests was arbitrary, and higher performance might be possible with different tests. Also, longitudinal cognitive assessments may be more specific and predictive of conversion than cross-sectional differences (Jiskoot et al., 2018b).

Important strengths of this study include the unique population and follow-up, which enabled the prediction of conversion in genetic FTD. Our clinical follow-up demonstrated that non-converters remained asymptomatic even after 4 (and in 25 cases after 6) years, ensuring that the labels ‘converter’ and ‘non-converter’ were truly separate. Furthermore, our methods have been validated in previous studies concerning classification in AD and FTD (Schouten et al., 2016; Bouts et al., 2018). The most important limitation to this study was sample size. Our index group consisted of seven converters, which is a small number in neuroimaging studies, and even more so in machine learning. This resulted in performance estimates with a large degree of uncertainty, which means that relatively high performances were necessary to significantly outperform chance level. For example, the AUC needed to significantly outperform chance level after family-wise error rate correction was 0.78 for the MRI analyses, and 0.74 for cognitive analyses. Another limitation to this study was the sample’s heterogeneity. We included subjects from *MAPT*, *GRN*, and *C9orf72* families in order to boost our sample size, even though there is evidence that each mutation has its own pattern of neurodegeneration over time due to different underlying pathology (Seelaar et al., 2011; Whitwell et al., 2012; Mann & Snowden, 2017; Jiskoot et al., 2018a). Indeed, converters with a *MAPT* mutation seemed to have higher conversion prediction scores than did converters with a *GRN* mutation, though this may also be due to sex differences. Stratification was not feasible due to our small sample size, but is necessary in future studies to show whether conversion prediction is possible in all FTD gene mutations or solely in *MAPT*. Just as different FTD mutations have different neurodegenerative profiles, so do the different clinical syndromes that constitute FTD (Seelaar et al., 2011; Whitwell et al., 2012; Rohrer & Rosen, 2013). Five subjects developed bvFTD, while two developed nfvPPA, increasing heterogeneity in our analyses. Despite these sources of heterogeneity, the FA-based model predicted conversion in 4 years beyond chance level, which is an important proof of concept that conversion prediction is possible in genetic FTD. Larger sample sizes might in the future facilitate hierarchical (Kim et al., 2019) or multilabel (Raamana et al., 2014; Klöppel et al., 2015; Koikkalainen et al., 2016; Bron et al., 2017; Canu et al., 2017) prediction to deal with these sources of heterogeneity. While the acquisition of a larger, similar cohort is costly and time-

consuming, our current results indicate that the acquisition of such data is meaningful and may lead to more accurate conversion prediction for genetic FTD. Lastly, the timing of conversion, and, therefore, the time between MRI and conversion, was based on heteroanamnesic information from knowledgeable informants, and as such may not be fully accurate. However, it should be noted that an exact time of conversion is near impossible to estimate in neurodegenerative diseases due to the gradual increase in symptomatology, and this inaccuracy is unlikely to have influenced our conclusions.

Conclusion

To conclude, we showed that FA predicted conversion within 4 years in presymptomatic FTD mutation carriers. This proof-of-concept study underlines the potential of MRI-based prediction in genetic FTD to contribute to a reliable early-stage FTD diagnosis, and should be replicated when larger sample sizes become available to corroborate our results.

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.