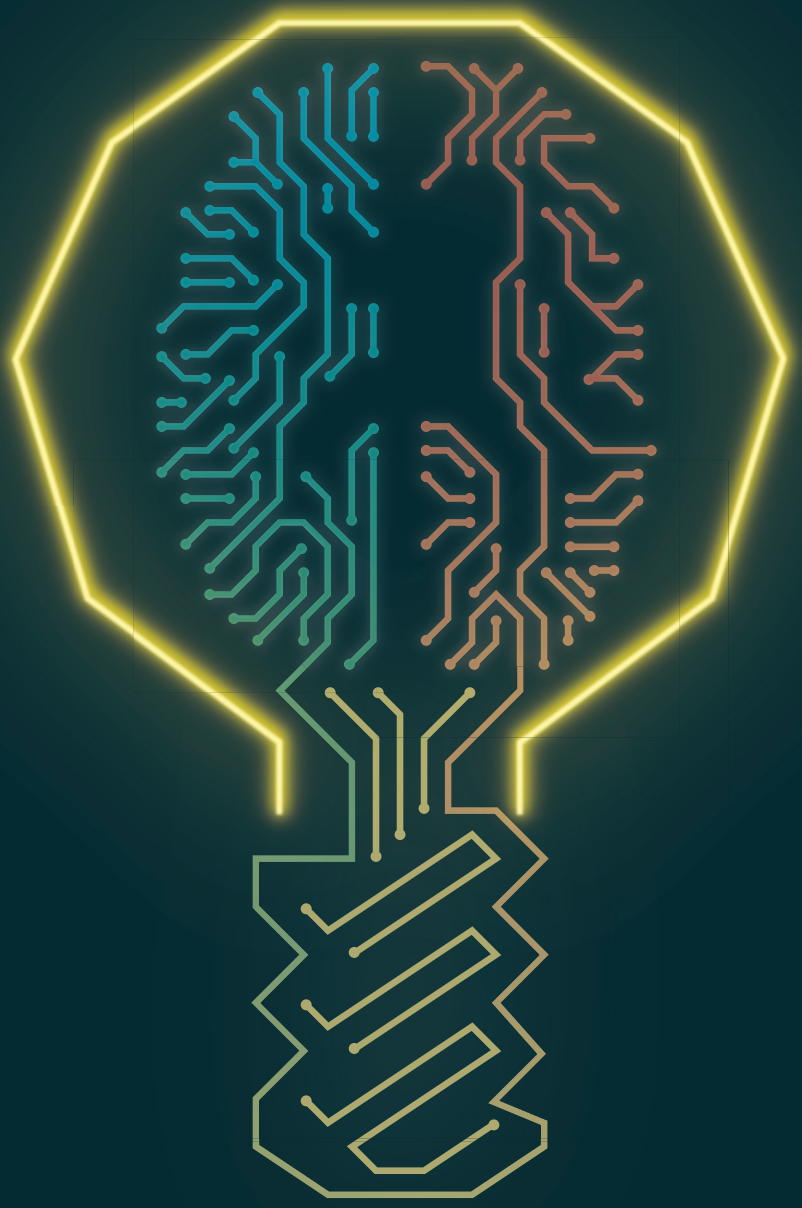
Neuroimaging biomarkers in genetic frontotemporal dementia : towards a timely diagnosis

Feis, R.A.



Chapter 7

General discussion

Main findings

In this dissertation, we aimed to advance the development of potential magnetic resonance imaging (MRI-)based biomarkers for the early detection of frontotemporal dementia (FTD). We focused on machine learning with multimodal MRI features to study potential early FTD biomarkers in the single-subject setting that corresponds with the clinical diagnostic process. Furthermore, we aimed to facilitate the combination of existing resting-state functional MRI (rs-fMRI) data from different centres through structured noise reduction. Lastly, we tested whether presymptomatic MRI changes were different between cognitively healthy subjects at genetic risk for respectively FTD and Alzheimer's disease (AD). In this chapter, the main findings of the studies in this dissertation are summarised and discussed. Critical considerations and future directions for neuroimaging research in FTD are also examined.

Clinical translation of MRI-based biomarkers for early FTD diagnosis

The translation of promising biomarkers to clinical practice requires validation in a single-subject setting. Accordingly, classification studies with FTD patients and controls are becoming increasingly important in the FTD neuroimaging field. Classifiers based on grey matter volume (Raamana et al., 2014; Klöppel et al., 2015; Koikkalainen et al., 2016; Bron et al., 2017; Canu et al., 2017; Meyer et al., 2017; Bouts et al., 2018), white matter diffusion (Bron et al., 2017; Canu et al., 2017; Bouts et al., 2018), and functional connectivity (Canu et al., 2017; Bouts et al., 2018) distinguish FTD patients from controls with high accuracy. However, previous classification studies relied on the current diagnostic criteria for patient inclusion, and may therefore not generalise to early FTD disease stages. On the other hand, multimodal MRI differences between groups of presymptomatic FTD mutation carriers and controls have been investigated (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; Jiskoot et al., 2018a; Panman et al., 2019), though single-subject classification and prediction studies were lacking in the preclinical stage.

In **chapter 2**, we studied classification in presymptomatic FTD mutation carriers (microtubule-associated protein tau [*MAPT*] and progranulin [*GRN*] mutation carriers, and chromosome 9 open reading frame 72 [*C9orf72*] repeat expansion carriers) and familial non-carriers using MRI-based classification models. We found that several carrier-control models, which were based on white matter features, separated mutation carriers from controls beyond chance level. Importantly, this shows that single-subject classification is possible in the presymptomatic stage of genetic FTD. However, a classification model trained on symptomatic bvFTD patients and controls (Bouts et al., 2018) did not outperform chance level when applied to the presymptomatic sample. This model included grey matter density (GMD), white matter diffusion, and functional connectivity features, highlighting that features that are important for the classification of symptomatic FTD patients are not necessarily useful in earlier stages.

In **chapter 3**, we investigated how MRI-based classification scores develop over time as FTD mutation carriers approach symptom onset. We calculated MRI-based classification scores at each time point, and used linear mixed effects models to compare presymptomatic mutation carriers with non-carriers. We found that the progression of classification scores over time was similar for presymptomatic FTD mutation carriers and non-carriers. Within the mutation carrier group, we compared mutation carriers that developed symptoms during the study's follow up (termed 'converters') with mutation carriers that remained presymptomatic ('non-converters'). We found

that classification scores in the converter group increased significantly more than in the non-converter group. The results in **chapter 3** show that MRI-based classification is sensitive to FTD-related changes over time. Also, our results suggest that presymptomatic FTD mutation carriers remain relatively stable over time, until they are within a few years of symptom onset, when brain changes begin to rapidly develop.

Since the converters had a large increase in classification scores over time compared to non-converters, we hypothesised that it might be possible to use MRI features to predict future symptom onset in FTD mutation carriers. In **chapter 4**, we included MRI data of 42 presymptomatic FTD mutation carriers, and trained classifiers to separate those that developed symptoms within four years post-MRI from those that remained presymptomatic after four years. Although only seven subjects converted within the four years follow-up, it was still possible to separate these converters from non-converters beyond chance level using the fractional anisotropy (FA) feature. This proof of concept implies that accurate symptom onset prediction may be possible in genetic FTD once larger cohorts become available.

Our results in **chapter 2–4** are consistent with the hypothesis that white matter changes predate and exceed grey matter atrophy in FTD. In presymptomatic FTD mutation carriers, differences in white matter diffusion tensor imaging (DTI) metrics are found in absence of grey matter atrophy (Borroni et al., 2008; Dopper et al., 2014), or exceed grey matter atrophy in magnitude (Pievani et al., 2014; Cash et al., 2018; Jiskoot et al., 2018a; Panman et al., 2019). In the years around symptom onset, grey matter atrophy starts increasing (Jiskoot et al., 2019), though white matter DTI changes remain more extensive than grey matter atrophy in clinical FTD (Agosta et al., 2012; Zhang et al., 2013; Mahoney et al., 2014). Although we did not statistically compare the different classification models to each other, it is striking that white matter features were the only ones to outperform chance level in our presymptomatic studies (**chapter 2, 4**). The best unimodal classification model to separate FTD mutation carriers from controls was based on the radial diffusivity (RD) feature (**chapter 2**), while multimodal models that outperformed chance included RD, white matter density (WMD), mean diffusivity (MD), and/or axial diffusivity (AxD). FA was the only feature that could predict which FTD mutation carriers would develop symptoms in four years' time (**chapter 4**). A model trained on bvFTD patients and controls that also included the GMD feature could not discriminate between presymptomatic mutation carriers and controls, neither cross-sectionally (**chapter 2**), nor longitudinally (**chapter 3**). However, it did show a different classification score progression between mutation carriers that developed symptoms during follow-up and mutation carriers that remained presymptomatic (**chapter 3**). This result might be explained by the notion that grey matter atrophy only becomes detectable around the time of symptom onset. Together, our results and previous studies indicate that white matter DTI changes are likely the most promising MRI biomarker for early diagnosis in genetic FTD.

Resting-state functional MRI data harmonisation across centres

In **chapter 5**, we applied 'FIX' (FMRIB's ICA-based X-noiseifier), a novel clean-up tool for structured noise, on two MRI data sets of healthy controls from different centres to investigate whether it can improve rs-fMRI data harmonisation between centres. We showed that FIX removes structured noise, while retaining physiologically driven functional connectivity differences. FIX reduced nearly all functional connectivity differences to non-significant levels, with the exception of functional connectivity differences in the visual network. These visual network differences were likely physiological differences secondary to the experimental design, as one sample was scanned with eyes open, while the other was scanned with eyes closed.

Recently, several other data harmonisation tools have been proposed, such as ICA-AROMA (Pruim et al., 2015) and ComBat (Johnson et al., 2007; Yu et al., 2018). FIX and ICA-AROMA are both specifically designed for the denoising of rs-fMRI data using independent component analysis (ICA), while ComBat has been validated for DTI (Fortin et al., 2017), cortical thickness (Fortin et al., 2018), and functional connectivity (Yu et al., 2018) analyses. Further research should establish whether the combination of FIX and ComBat further enhances the quality and reproducibility of rs-fMRI, or whether combining these steps would prove redundant.

In new multicentre studies, the standardisation of protocols across centres is important to reduce noise as much as possible (Wegner et al., 2008; Zivadinov & Cox, 2008; Glover et al., 2012). However, our results in **chapter 5** and other data harmonisation studies (Pruim et al., 2015; Fortin et al., 2017, 2018; Yu et al., 2018) suggest that reanalysis of existing non-standardised rs-fMRI data may also be possible across sites. This possibility will especially be important to facilitate the valid comparisons of rare diseases and at-risk populations.

Neuroimaging biomarkers for differential diagnosis of FTD and AD

On MRI, FTD and AD are characterised by divergent patterns of structural and functional neurodegeneration in terms of grey matter atrophy (Zhang et al., 2011; Möller et al., 2015b), white matter degeneration (Zhang et al., 2011; Möller et al., 2015b; Daianu et al., 2016), and functional connectivity (Zhou et al., 2010). Grey matter atrophy is greater in FTD patients in the orbitofrontal, inferior and medial frontal gyrus, the anterior cingulate gyrus, caudate nucleus, and the nucleus accumbens (Zhang et al., 2011; Möller et al., 2015b), while AD patients have more grey matter atrophy in the bilateral occipital gyri and the left precuneus (Zhang et al., 2011). White matter changes are more abundant in FTD patients than AD patients. FTD patients have stronger diffusion abnormalities than AD patients in the uncinate fasciculi, forceps minor, and anterior thalamic radiation, while there are no areas that clearly show the opposite pattern (Zhang et al., 2011; Möller et al., 2015b; Daianu et al., 2016). Furthermore, functional connectivity in the default mode network and salience network are inversely affected in AD and FTD. In AD, functional connectivity in the default mode network is disrupted, while functional connectivity in the salience network is increased. Conversely, functional connectivity in the salience network is disrupted in FTD, while functional connectivity in the default mode network is increased (Zhou et al., 2010).

If present in early FTD and AD stages, these diverging patterns might serve as biomarkers to differentiate between FTD and AD. In **chapter 6**, we investigated grey matter atrophy, white matter diffusion, and functional connectivity in cognitively healthy subjects at risk for FTD (i.e., *MAPT* and *GRN* mutation carriers) and AD (i.e., apolipoprotein E ε4 [*APOE4*] carriers). However, we found no evidence for diverging patterns between the at-risk groups for FTD and AD in our cross-sectional sample. To gain more insight into the early differences between FTD and AD neuropathological processes, longitudinal studies in genetic at-risk groups may be required. Alternatively, memory clinic cohorts, especially patients without dementia diagnosis at first presentation, may provide the means to longitudinally study early stages of sporadic FTD and AD (Handels et al., 2012).

Explosive onset of neuropathology in genetic FTD

In genetic FTD, widespread presymptomatic changes were found in large samples of mutation carriers (Rohrer et al., 2015; Cash et al., 2018; Jiskoot et al., 2018a). However, in smaller samples, similar presymptomatic changes were generally not found (Borroni et al., 2008; Whitwell et al.,

2011a; Dopper et al., 2014; Panman et al., 2019), unless subjects were near symptom onset (Jiskoot et al., 2019). These results have led to the hypothesis that detectable neurodegenerative processes may start relatively shortly before symptom onset in genetic FTD. Our results further corroborate this hypothesis. We did not find convincing presymptomatic differences on multimodal MRI between presymptomatic FTD mutation carriers and non-carriers (**chapter 6**), and classification performance was modest though above chance level (**chapter 2**). Moreover, classification scores progressed similarly for presymptomatic mutation carriers and controls, unless the mutation carriers were near symptom onset, in which case the scores rose steeply (**chapter 3**). Finally, it was possible to predict symptom onset in presymptomatic mutation carriers within a timespan of four years beyond chance level, even in a small sample (**chapter 4**). These results might seem slightly at odds with large group studies in the Genetic Frontotemporal dementia Initiative (GENFI; Rohrer et al., 2013), which reported grey matter volume loss from up to 10 years before estimated symptom onset (Cash et al., 2018), and white matter DTI changes from up to 30 years before estimated symptom onset (Jiskoot et al., 2018a). Indeed, the acquisition of large data sets enables the detection of very small differences in group studies. However, these early and small differences are not necessarily useful for classification in the single-subject setting due to small effect sizes and large individual variance.

So far, few FTD mutation carriers have developed symptomatic FTD in longitudinal studies (**chapter 3, 4**). Given the explosive rate in which neurodegeneration develops in the years prior to symptom onset, continuation of these studies is crucial to develop more accurate and robust classification and prediction models.

Critical considerations and future directions

Sample size

For machine learning purposes (**chapters 2, 3,** and particularly **chapter 4**), we had relatively few data. Cross-validation crucially minimises overfitting in classification analyses, but leads to model uncertainty, especially in small sample sizes (Varoquaux, 2018). This resulted in low power to find classification models that statistically outperformed chance level, which may have resulted in false negative findings in **chapters 2, 4**. On the other hand, *MAPT* and *GRN* mutations, and *C9orf72* repeat expansion are each associated with its distinct neurodegenerative pattern (Seelaar et al., 2011; Whitwell et al., 2012; Jiskoot et al., 2018a), and one might argue that pooling them may have biased our outcomes towards the largest group, in our case the *GRN* mutation carriers, and introduced additional heterogeneity. As such, our small sample size led to methodological imperfections from both clinical and statistical perspectives. We balanced these two interests to present a unique proof of concept that presymptomatic changes in FTD mutation carriers can be detected on the single-subject level.

Longitudinal MRI data acquisition is expensive and presents a psychological burden for these rare mutation carriers. Therefore, our results are important as a proof of concept that longitudinal MRI acquisition may aid the development of early diagnostic FTD biomarkers. When more data become available, stratification of classification analyses across mutations, underlying frontotemporal lobar degeneration (FTLD) pathologies, and clinical FTD syndromes may lead to more accurate and robust models through increased homogeneity. Moreover, the combination of overlapping (Elahi et al., 2017) and diverging (Whitwell et al., 2010; Galantucci et al., 2011; Zhang et al., 2011; Mahoney et al., 2014; Tu et al., 2015; Daianu et al., 2016; Omer et al., 2017) neuroimaging changes associated with the different FTD syndromes may facilitate hierarchical classification in presymptomatic FTD mutation carriers, as has already been done in symptomatic AD and FTD variants (Kim et al., 2019). The low prevalence of FTD mutation carriers and heterogeneity within and between mutations are considerable challenges for the validation of potential biomarkers. However, advances in data harmonisation techniques will make it easier to combine neuroimaging data (**chapter 5**; Fortin et al., 2017, 2018; Yu et al., 2018) within and perhaps between large consortia, such as GENFI (Rohrer et al., 2013; Europe and Canada), and the Advancing Research and Treatment for Frontotemporal Lobar Degeneration / Longitudinal Evaluation of Familial Frontotemporal Dementia Subjects (ARTFL/LEFFTDS; Staffaroni et al., 2019; USA and Canada) cohorts.

Diffusion imaging

In this dissertation, especially **chapters 2, 4**, we speculate that white matter diffusion metrics are the most promising biomarkers for early diagnosis in genetic FTD. We investigated white matter diffusion characteristics using the DTI model (Basser et al., 1994) in conjunction with tract-based spatial statistics (TBSS; Smith et al., 2006). However, both methods have their limitations. For example, when using a study-specific TBSS skeleton (e.g., in **chapter 6**), white matter pathology in the tract centres of patients or mutation carriers may influence the skeletonisation step, thus biasing results (Bach et al., 2014). In **chapters 2–4**, we used a predefined white matter skeleton to circumvent this problem. More general problems with DTI include the inability of DTI metrics such as FA to differentiate between intrinsic white matter properties and fibre coherence (Dell'Acqua & Tournier, 2019). Recent advances in diffusion MR acquisition, such as multiband excitation (Feinberg et al., 2010; Sotiropoulos et al., 2013), and modelling, such as spherical deconvolution (Tournier et al.,

2007; Dell'Acqua et al., 2010; Zhang et al., 2012; Raffelt et al., 2015; Dell'Acqua & Tournier, 2019), provide solutions to these problems, but have not been extensively used yet in clinical populations (Mito et al., 2018). We believe it is crucial to adopt these new diffusion MRI techniques in future studies involving FTD patients and presymptomatic mutation carriers to validate our results using novel white matter metrics that better capture microstructural white matter changes.

Classification analyses

For our classification analyses in **chapters 2–4**, we used elastic net (Zou & Hastie, 2005; Friedman et al., 2010), a type of regularised logistic regression, on multimodal MRI features in a repeated nested cross-validation scheme. Here, we review some strengths and limitations to this approach, and discuss the features we used.

Elastic net combines L1 (lasso; Tibshirani, 1996) and L2 (ridge; Hoerl & Kennard, 1970) penalties to regularise feature inclusion and the weight of each feature. The resulting models only include a subset of all features, which is essential when the number of features exceeds the number of subjects. However, though regularisation aids model performance and generalisation, it hinders the interpretability of the classification model (Shmueli, 2010). When redundant information is encoded in multiple correlated features, regularised regression models will minimise the sum of prediction error and penalty by including only one of these features in the classification model. Therefore, the features' beta weights in the classification model do not reliably indicate how meaningful the included features are. In neuroimaging classification studies, high beta weights in certain features, such as hippocampal atrophy, are sometimes claimed to verify the model's biological validity. In our work, we have tried to refrain from bold claims based on the classification models' beta weights.

We performed all classification analyses in a nested cross-validation scheme to ensure we did not overfit our classification models (Kriegeskorte et al., 2009). We used the inner loop of the cross-validation to tune the hyperparameters, and the outer loop to fit and test the classification model. As such, the outer loop test set is used for neither hyperparameter tuning, nor training. The drawback of cross-validation is that data partitioning increases variance, resulting in increased model uncertainty, especially in smaller samples (Varoquaux, 2018). We repeated the entire cross-validation procedure 50 times to reduce the variability arising from the random partitioning in training and test folds.

In addition to applying a nested cross-validation approach, we took further care to avoid overfitting by using feature selection procedures that reduced the data's dimensionality without relying on class differences in our sample. To define grey and white matter features, we used anatomical atlases, while we used an unsupervised data-driven ICA (Beckmann & Smith, 2004) approach to define functional connectivity features. Our grey matter classification features consisted of GMD in 96 regions defined by the Harvard-Oxford cortical atlas, and grey matter volume in 14 subcortical volumes using FSL FIRST segmentations. White matter features included WMD and diffusion metrics (i.e., FA, MD, AxD, and RD) in 20 Johns-Hopkins university white matter atlas tracts. Functional connectivity features comprised full and partial (i.e., L1-regularised; J. Friedman et al., 2008) correlations between 70 (**chapter 2, 3**) or 20 (**chapter 4**) ICA components.

We chose these specific features based on earlier work, in which AD patients were classified from controls (Schouten et al., 2016; Bouts et al., 2018), and FTD patients were classified from controls and AD patients (Bouts et al., 2018). In the FTD field, no comparisons have been made between different features of grey matter structure, white matter diffusion, or functional connectivity. Therefore, higher performances than reported in this dissertation might be possible with a different

feature selection. However, in comparative AD classification studies for grey matter structure (de Vos et al., 2016), white matter diffusion (Schouten et al., 2017), and functional connectivity (de Vos et al., 2018) features, the features used here performed generally well.

Diagnostic uncertainty

In **chapters 3, 4**, we included the data of *MAPT* and *GRN* mutation carriers that had converted to FTD. Conversion was carefully determined by a multidisciplinary team according to the newest criteria for bvFTD (Rascovsky et al., 2011), primary progressive aphasia variants (Gorno-Tempini et al., 2011), and amyotrophic lateral sclerosis (Ludolph et al., 2015), which are also used in the clinic. The use of these diagnostic criteria may have introduced some circularity to our results, as atrophy on neuroimaging is one of the diagnostic criteria on which conversion was defined. However, a strength of this approach is that it facilitates the comparison to other studies and simplifies replication.

Conclusions

Neuroimaging group analyses in presymptomatic FTD mutation carriers have taught us much about the early pathways of FTLD neuropathology. This dissertation aimed to advance the development of presymptomatic MRI-based biomarkers for FTD towards clinical application through the use of machine learning. Importantly, eventual clinical implementation of machine learning seems straightforward, as machine learning can combine multiple complementary inputs (e.g., different MRI-based features) to calculate a single diagnostic score per subject. We showed that multimodal MRI-based biomarkers can identify presymptomatic FTD mutation carriers on the individual level, and are sensitive to increasing FTLD pathology over time. Moreover, we found that diffusion MRI can predict symptom onset within 4 years in genetic FTD. Finally, we showed that the reduction of structured noise in rs-fMRI aids the combination of such data in multicentre studies. Though our findings require replication in larger cohorts, this dissertation provides the optimistic prospect that neuroimaging FTD biomarkers may aid timely FTD diagnosis in individual patients.