



Universiteit
Leiden
The Netherlands

Multimodal MRI-based classification of Alzheimer's disease

Vos, F. de

Citation

Vos, F. de. (2021, December 9). *Multimodal MRI-based classification of Alzheimer's disease*. Retrieved from <https://hdl.handle.net/1887/3245855>

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/1887/3245855>

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

3

A comprehensive analysis of resting state fMRI measures to classify individual patients with Alzheimer's disease

Published in Neurolmage, 2018; 167, 62-72.

Frank de Vos, Marisa Koini, Tijn M. Schouten, Stephan Seiler,
Jeroen van der Grond, Anita Lechner, Reinhold Schmidt, Mark de Rooij,
Serge A. R. B. Rombouts

Abstract

Alzheimer's disease (AD) patients show altered patterns of FC (FC) on resting state functional magnetic resonance imaging (fMRI) scans. It is yet unclear which resting state fMRI measures are most informative for the individual classification of AD patients. We investigated this using resting state fMRI scans from 77 AD patients ($MMSE = 20.4 \pm 4.5$) and 173 controls ($MMSE = 27.5 \pm 1.8$). We calculated i) FC matrices between resting state components as obtained with independent component analysis (ICA), ii) the dynamics of these FC matrices using a sliding window approach, iii) the graph properties (e.g., connection degree, and clustering coefficient) of the FC matrices, and iv) we distinguished five FC states and administered how long each subject resided in each of these five states. Furthermore, for each voxel we calculated v) FC with 10 resting state networks using dual regression, vi) FC with the hippocampus, vii) eigenvector centrality, and viii) the amplitude of low frequency fluctuations (ALFF). These eight measures were used separately as predictors in an elastic net logistic regression, and combined in a group lasso logistic regression model. We calculated the area under the receiver operating characteristic curves (AUC) to determine classification performance. The AUC values ranged between 0.51 and 0.84 and the highest were found for the FC matrices (0.82), FC dynamics (0.84) and ALFF (0.82). The combination of all measures resulted in an AUC of 0.85. We show that it is possible to obtain moderate to good AD classification using resting state fMRI scans. FC matrices, FC dynamics and ALFF are most discriminative and the combination of all the resting state fMRI measures improves classification accuracy slightly.

Keywords: resting state fMRI, Alzheimer's disease, classification, independent component analysis, dual regression, dynamic functional connectivity

3.1 Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder characterised by widespread grey matter atrophy (Jack et al., 2004), specifically hippocampal atrophy is considered to be the hallmark of AD (Morra et al., 2009). In order to develop a cure, or to slow down the disease progression, it is essential to diagnose AD in an early stage (Prince et al., 2011).

AD patients differ in their pattern of functional connectivity (FC) as shown by resting state functional magnetic resonance imaging (fMRI) scans. They have decreased FC between the hippocampus and several regions throughout the neocortex (Allen et al., 2007; Wang et al., 2006), reduced FC within the default mode network (Binnewijzend et al., 2012; Greicus et al., 2004), and increased FC within the frontal networks (Agosta et al., 2012). AD patients also have different large-scale FC matrices (Brier et al., 2012) and graph properties derived from these matrices (Sanz-Arigita et al., 2010; Supekar et al., 2008). In addition, AD patients differ in the dynamics of their FC and their dwell time in specific FC states (Jones et al., 2012). Furthermore, AD patients have less signal in the low frequency domain (0 - 0.1 Hz) of their resting state signal (Han et al., 2011).

These FC differences might exist in an early stage of AD, even before the presence of brain atrophy and cognitive decline (Buckner et al., 2005; Sheline and Raichle, 2013). For instance, cognitively normal elderly with increased amyloid binding, an important AD indicator, have decreased FC between the precuneus and several regions within the default mode network, and these effects are similar to those observed in AD patients (Sheline et al., 2010a). Carriers of the APOE $\epsilon 4$ gene, who are at genetic risk for AD, have reduced FC between the precuneus and the hippocampus (Sheline, et al., 2010b), and increased FC within the default mode network (Filippini et al., 2009).

Resting state fMRI might be used for the diagnosis or even early detection of AD and it is important to investigate this potential (Buckner et al., 2005; Sperling, 2011). AD biomarkers can be evaluated using individual classification studies. Resting state fMRI-based AD classification studies have progressed through the use of machine learning techniques. Machine learning techniques enable the incorporation of many predictors into one predictive model and they automatically select the relevant ones. So far, AD has been classified moderately to good using FC matrices (Challis et al., 2015; Chen et al., 2011; Schouten et al., 2016) and their graph properties (Khazaee et al., 2015), FC dynamics (Wee et al., 2016), FC within the default mode network (Koch et al., 2012) and the amplitude of low frequency fluctuations (ALFF; Dai et al., 2012).

It is not known which of these resting state fMRI measures is best for AD classification. Moreover, the combination of different resting state fMRI measures might improve AD classification (Dai et al., 2012; de Vos et al., 2016; Mesrob et al., 2012; Schouten et al., 2016; Sui et al., 2013). In this study, we will use a wide range of resting state fMRI measures in combination with machine learning techniques to classify AD patients and controls. These measures include FC with several resting state networks (RSNs), FC with the hippocampus, FC matrices and their graph properties, FC dynamics, FC states, and the ALFF within the resting state signal. We will determine the most accurately predicting measures and combine them to investigate whether this increases the classification accuracy.

3.2 Materials and methods

Participants

Our dataset consisted of 77 clinically diagnosed probable AD patients and 173 cognitively normal elderly controls (see Table 1). The AD patients were scanned at the Medical University of Graz as a part of the prospective registry on dementia (PRODEM; see also Seiler et al., 2012). The inclusion criteria for PRODEM are: dementia diagnosis according to DSM-IV criteria (American Psychiatric Association, 2000), non-institutionalisation or need for 24-hour care, and the availability of a caregiver who agrees to provide information on the patients' and his or her own condition. Patients were excluded if they were unable to sign an informed consent or if co-morbidities were likely to preclude termination of the study. We used the baseline scans from the PRODEM study, and only included patients that were diagnosed with AD in line with the NINCDS-ADRDA Criteria (McKhann et al., 1984), and for which anatomical MRI and resting state fMRI scans were available. The controls were scanned at the same scanning site, over the same time period, with the same scanning protocol as a part of the Austrian stroke prevention study. The Austrian Stroke Prevention Study is a community-based cohort study on the effects of vascular risk factors on brain structure and function in elderly participants without a history or signs of stroke and dementia on the inhabitants of Graz, Austria (see also Schouten et al., 2016).

MR acquisition

All participants were scanned on a Siemens Magnetom TrioTim 3T MRI scanner. The anatomical T1-weighted images were acquired with the following parameters: TR = 1900 ms, TE = 2.19 ms, flip angle = 9°, and an isotropic voxel size of 1 mm. The

resting state fMRI session was conducted, acquiring 150 volumes with TR = 3000 ms, TE = 30 ms, flip angle = 90°, 40 axial slices, with an isotropic voxel size of 3 mm. The participants were instructed to lie still with their eyes closed, and to stay awake.

MRI preprocessing

The MRI data were preprocessed using the FMRIB Software Library (FSL, version 5.0) (Jenkinson et al., 2012; Smith et al., 2004). For the anatomical MRI this included brain extraction, bias field correction, and non-linear registration to standard MNI152 template (Grabner et al., 2006). For the resting state fMRI data this included brain extraction, motion correction, a temporal high pass filter with a cut-off point of 100 seconds, and spatial smoothing. The mean framewise displacement as calculated by MCFLIRT from FSL (Jenkinson et al., 2002) ranges from 0.02 to 0.42 mm (mean = 0.10, SD = 0.06) for the control subjects, and from 0.03 to 0.55 mm (mean = 0.13, SD = 0.11) for the AD patients ($p < 0.05$). To control for head motion, we applied motion correction using MCFLIRT (Jenkinson & Smith, 2002), and regressed the motion parameters out of the fMRI data. Additionally, we used the FMRIB's ICA-based Xnoisefilter (FIX, version 1.06) to automatically identify and remove noise components from the fMRI data (Salimi-Khorshidi et al., 2014), thereby increasing the signal to noise ratio (Griffanti et al., 2016). For the spatial smoothing, we used a smoothing kernel with a full width half maximum of 3 mm. We performed minimal smoothing, because this is recommended prior to running an independent component analysis (ICA) in order to reduce the probability of finding spurious components (Jenkinson, 2015).

Table 1. Sample demographics

	Controls		AD ¹ patients		X ²
Gender (♂/♀)	74/99 (57% ♀)		31/46 (60% ♀)		n.s. ²
	min - max	mean ± SD	min - max	mean ± SD	t-test
Age	47 - 83	66.1 ± 8.7	47 - 83	68.6 ± 8.6	p<0.05
Education (years)	9 - 18	11.5 ± 2.8	4 - 20	10.8 ± 3.2	n.s.
Disease duration (months)	-	-	2 - 156	26.7 ± 24.5	-
MMSE ³	22 - 30	27.5 ± 1.8	10 - 28	20.4 ± 4.5	p<0.001
CDR ⁴	-	-	0.5 - 2	0.8 ± 0.3	-
GDS ⁵	0 - 11	2.1 ± 2.1	0 - 10	2.6 ± 2.6	n.s.

¹AD = Alzheimer's disease, ²MMSE = mini-mental state examination, ³CDR = clinical dementia rating,

⁴GDS = geriatric depression scale.

Resting state fMRI measures

We calculated eight types of measures from the resting state fMRI data. For most of those eight types of measures we calculated more than one variety, resulting in a total of 31 measures. These resting state fMRI measures are listed in Table 2, along with the number of values they comprise. These values are used as predictors in the classification analyses. Figure 1 summarises the procedures used to calculate the resting state fMRI measures. A more elaborated description is written below.

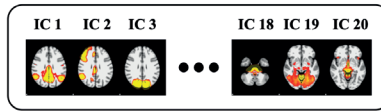
Table 2. List of resting state fMRI measures used for Alzheimer's disease classification

Resting state measure	# of predictors
1: FC¹ matrices	
1a. 20 X 20 full correlation	190
1b. 70 X 70 full correlation	2415
1c. 20 X 20 sparse partial correlation	190
1d. 70 X 70 sparse partial correlation	2415
FC dynamics	
2a. SD ² of 20 X 20 full correlation FC matrix	190
2b. SD of 70 X 70 full correlation FC matrix	2415
2c. SD of 20 X 20 sparse partial correlation FC matrix	190
2d. SD of 70 X 70 sparse partial correlation FC matrix	2415
3: FC states	
3a. FC states of 20 X 20 full correlation FC matrix	5
3b. FC states of 70 X 70 full correlation FC matrix	5
3c. FC states of 20 X 20 partial correlation FC matrix	5
3d. FC states of 70 X 70 partial correlation FC matrix	5
4: Graph metrics	
4a. Graph metrics of 20 X 20 full correlation FC matrix	124
4b. Graph metrics of 70 X 70 full correlation FC matrix	424
4c. Graph metrics of 20 X 20 partial correlation FC matrix	124
4d. Graph metrics of 70 X 70 partial correlation FC matrix	424
5: FC with resting state networks	
5a. FC with visual network 1	190981
5b. FC with visual network 2	190981
5c. FC with visual network 3	190981
5d. FC with default mode network	190981
5e. FC with the cerebellum	190981
5f. FC with sensorimotor network	190981
5g. FC with auditory network	190981
5h. FC with executive control network	190981
5i. FC with frontoparietal network 1	190981
5j. FC with frontoparietal network 2	190981
6: FC with Hippocampus	
6a. FC with left hippocampus	190981
6b. FC with right hippocampus	190981
7: Eigenvector centrality	
Fast eigenvector centrality mapping	190981
8: ALFF³	
8a. ALFF	190981
8b. fALFF ⁴	190981
All resting state fMRI measures combined	2,876,251

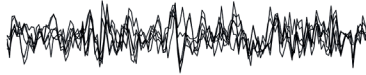
¹FC = functional connectivity, ²SD = standard deviation, ³ALFF = amplitude of low frequency fluctuations, ⁴fALFF = fractional amplitude of low frequency fluctuations.

Functional connectivity matrices

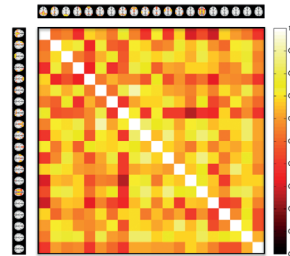
1A
group ICA
components



1B
subject time
courses per
component

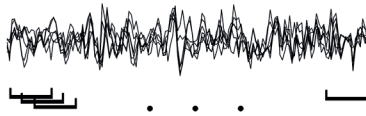


1C
connectivity matrix

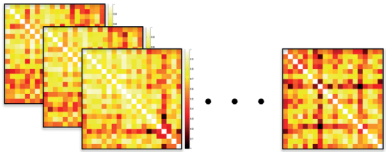


Functional connectivity dynamics

2A
subject time
courses per
component

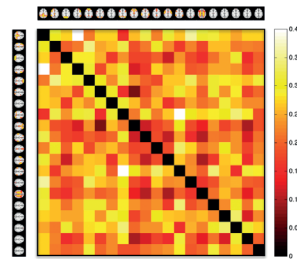


2B
sliding window
connectivity
matrices



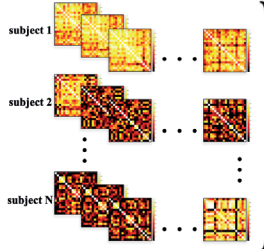
2C

SD of sliding window connectivity matrices

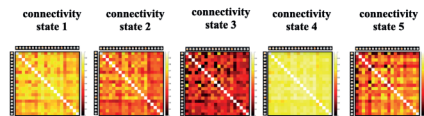


Functional connectivity states

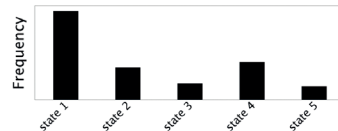
3A
sliding
window
connectivity
matrices for
all subjects



3B
K-means
clustering
into K= 5
connectivity
states

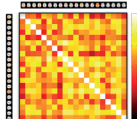


3C
count
subjects'
number of
windows
within each
state

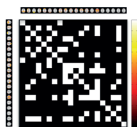


Graph metrics

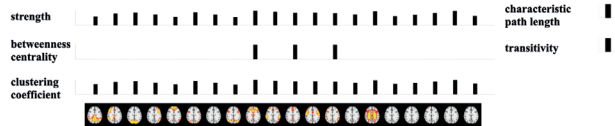
4A
connectivity
matrix



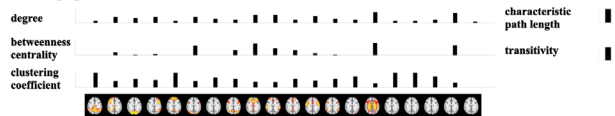
4B
binarised
connectivity
matrix



4C graph metrics from connectivity matrix

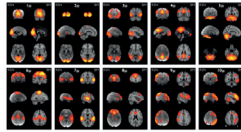


4D graph metrics from binarised matrix

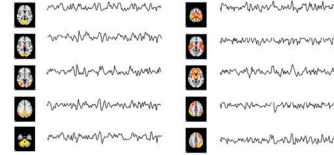


Functional connectivity with resting state networks

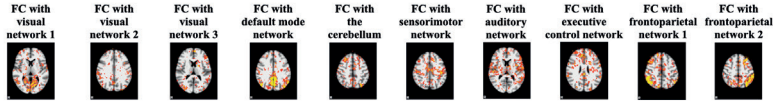
5A
resting state
network
templates from
Smith et al.
(2009)



5B
Dual regression
step 1 yields
subject-specific
time courses for
each template

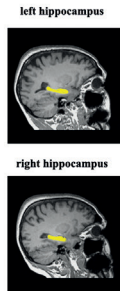


5C
Dual regression
step 2 yields
subject-specific
functional
connectivity
(FC) maps

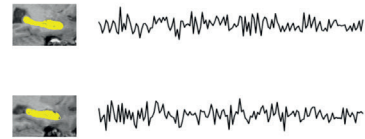


Functional connectivity with the hippocampus

6A
For each subject
use the structural
scan to segment
the left and right
hippocampus



6B
Register the hippocampus
to functional space and
calculate the mean time
course of the resting state
scan within the
hippocampus mask

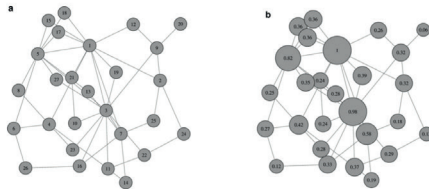


6C
Correlate the hippocampus
time course with each
voxel's time course to
obtain a whole brain
hippocampus connectivity
map



Eigenvector centrality

7
Eigen vector
centrality for
each voxel



Eigen vector centrality map

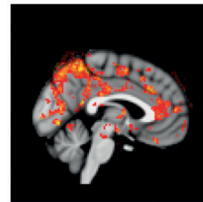
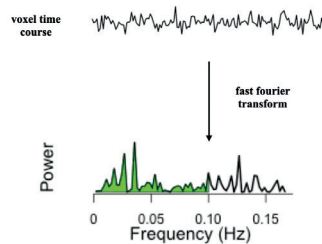


Figure from Wink et al. (2012). Left: a simulated network consisting of 27 nodes. Right: the Eigen vector centrality values of the nodes. Eigenvector centrality attributes large values to nodes that are central within the network.

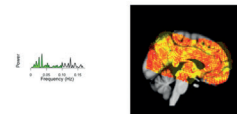
Amplitude of low frequency fluctuations

8A
Apply fast fourier
transform to the
time course of
each voxel to
obtain the power
spectral density



8B
Calculate the
magnitude within
the low frequency
band (0 - 0.1 Hz)

8C
Amplitude of
low frequency
fluctuations
(ALFF)



8D
Fractional ALFF.
Power within the
low frequency
band divided by
the total power

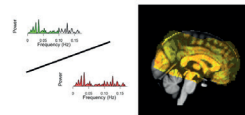


Figure 1. The procedures for calculating the eight resting state fMRI modalities.

Functional connectivity matrices

For each participant, we calculated FC between RSNs. We used temporal concatenation ICA in FSL MELODIC (Beckmann & Smith, 2004) to obtain RSNs. First, we registered the functional data of all participants to standard space and concatenated them along the time dimension. We then performed a low and a high dimensional ICA on the concatenated data set, forcing a solution with 20 and 70 components respectively. The components of these two ICA solutions are shown in Figure S2 in the supplementary materials. We registered the resulting ICA component weight maps back to subject space, weighted them by the subject specific grey matter density maps, and multiplied them with the functional data. We then calculated the mean time courses for the components and used these for the FC analysis. We calculated both full and partial correlation matrices. For the partial correlation matrices, we used the graphical lasso algorithm (Friedman et al., 2008) implemented in MATLAB (MATLAB 2013a, The MathWorks Inc., Natick, MA, 2000). We set the λ parameter at 100, because this setting works best in most cases for fMRI functional connectivity (Smith et al., 2011). For each participant, we thus calculated four FC matrices. The two 20 by 20 matrices each contain $(20 * 19)/2 = 190$ unique elements, and the 70 by 70 matrices each contain $(70 * 69)/2 = 2415$ unique elements. We used these elements as predictors for classification.

Dynamics of functional connectivity matrices

We also calculated the dynamics of the above-described FC matrices using a sliding window approach (Chang and Glover 2011; Hutchinson et al., 2013; Jones et al., 2012). We used a window size of 33 seconds, similar to Jones et al. (2012) and Rashid et al. (2014), because it was shown that time windows as short as 30 seconds can provide reasonably good connectivity estimates (Shirer et al. 2012). We shifted the windows one volume at a time, resulting in 140 windows (Jones et al., 2012; Rashid et al., 2014). Within each window we calculated the four FC matrices as described in the previous paragraph. Then we calculated the standard deviation of the FC matrices over all the windows. This resulted in four matrices of standard deviations for each subject, with equal size as the FC matrices. We used the elements of these matrices as predictors for classification.

Functional connectivity states

For each of the four types of FC matrices we distinguished five 'FC states' and administered how long each subject resided in each of these five states. Functional connectivity states are patterns of FC that reoccur in time across participants (Allen et al., 2012; Jones et al., 2012; Rashid et al., 2014). In order to determine the FC states,

we clustered the sliding window FC matrices using k-means clustering. So, for each of the four types of FC matrices we clustered the 250 (number of subjects) * 140 (number of windows) = 35000 sliding window matrices. We created $k = 5$ clusters like Jones et al. (2012) and Rashid et al. (2015) and we used the Manhattan distance criterion like Allen et al. (2012). Then, for each participant we counted the number of sliding window matrices that were assigned to each of the five FC states. The five frequency values for each of the four types of FC matrices were used as predictors for classification.

Graph metrics

For each of the four types of FC matrices we calculated commonly used graph metrics. We used both the original and the binarised version of the FC matrices. Binary links denote the presence or absence of connections, while the original values contain information about the connection strengths (Rubinov & Sporns, 2010). Current network methods cannot quantify the role of negative connections in network organisation (Rubinov & Sporns, 2010) and therefore we absolutised the negative links. We binarised the full correlation matrices by maintaining the 20% largest absolute correlations within each matrix (Khazaee et al., 2015). Since the sparse partial correlation matrices are sparse from itself, we did not apply a binarisation threshold, but binarised the matrices by transforming all values greater than zero to 1. We used the Brain Connectivity Toolbox (Rubinov & Sporns, 2010) available for MATLAB (MATLAB 2013a, The MathWorks Inc., Natick, MA, 2000) to calculate the graph metrics. For the original connectivity matrix, we calculated the connection strength, weighted betweenness centrality, and weighted clustering coefficient for every node in the network and the weighted characteristic path length and weighted transitivity for the entire network (Rubinov & Sporns, 2010). For the binarised connectivity matrix we calculated the connection degree, betweenness centrality, and clustering coefficient for every node in the network and the characteristic path length and transitivity for the entire network (Rubinov & Sporns, 2010). So, in total we calculated 10 different graph measures, six measures for every node and four measures for the entire network. This resulted in $6*20 + 4*1 = 124$ predictors for the 20*20 FC matrices and $6*70 + 4*1 = 424$ predictors for the 70*70 FC matrices.

Whole brain functional connectivity with resting state networks

We calculated whole brain FC with 10 RSNs using the dual regression approach in FSL (Filippini et al., 2009). We used templates that were obtained using an independent data set to increase the reproducibility of our findings (Griffanti, 2016). We used the RSN templates that were obtained using an ICA by Smith et al. (2012). These RSNs

are freely available online (http://www.fmrib.ox.ac.uk/analysis/brainmap+rsns/PNAS_Smith09_rsn10.nii.gz) as spatial maps in standard space. These 10 RSNs include three visual networks, the default mode network, the cerebellum, a sensorimotor network, an auditory network, an executive control network and two frontoparietal networks. Additionally, we included the white matter (WM) and cerebrospinal fluid (CSF) maps provided by FSL (Jenkinson et al., 2012; Smith et al., 2004) as confound maps. Those 12 spatial maps (10 RSNs plus two confound maps) were then used in a dual regression analysis. First, for each subject, the 12 spatial maps were regressed (as spatial regressors in a multiple regression) into the subjects' 4D space-time dataset. This results in a set of subject-specific time series, one for each spatial map. Next, those time series were regressed (as temporal regressors, again in a multiple regression) into the same 4D dataset, resulting in 12 subject-specific spatial maps, one for each RSN and one for each of the two confound maps. These subject-specific spatial maps represent whole brain FC with the RSNs. We used the voxel-wise whole brain FC results for the ten RSNs as predictors for classification.

Whole brain functional connectivity with hippocampus

For each participant, we calculated whole brain FC with the left and with the right hippocampus (Allen et al., 2007; Wang et al., 2006). We first calculated the time course of the hippocampus for each participant. To this end we segmented the hippocampus in the anatomical scan using FSL First. We eroded the segmented hippocampus with three voxel layers to ascertain that only hippocampus voxels were included. The eroded hippocampus was then affine registered to the functional data and we calculated the mean time course of the functional data within the hippocampus mask. Then, for each participant we regressed the time course of the hippocampus, along with the mean WM and CSF time courses as confound regressors, into the functional data using multiple regression. This resulted in a whole brain FC map with the hippocampus. We performed this analysis for both the left and the right hippocampus and used the two resulting whole brain FC maps as predictors for classification.

Eigenvector centrality

For each participant, we calculated an eigenvector centrality map. Eigenvector centrality attributes a value to each voxel in the brain such that a voxel receives a large value if it is strongly correlated with many other voxels that are themselves central within the network (Lohmann et al., 2010). We used the fastECM algorithm (Wink et al., 2012; Binnewijzend et al., 2014) to calculate a whole brain eigenvector centrality map in standard space for each participant.

Amplitude of low frequency fluctuations

We calculated ALFF (Biswal et al., 2010; Zang et al., 2007) and fractional ALFF (fALFF) (Zou et al., 2008) for each participant. We used the REST software package (Song et al., 2011) to calculate whole brain ALFF and fALFF maps. ALFF was defined as the power within the 0 - 0.1 Hz frequency band and fALFF was defined as the power within the 0 - 0.1 Hz frequency band divided by the power of the whole frequency spectrum. For standardisation purposes, we divided the voxels' ALFF/fALFF values by the mean ALFF/fALFF within a subjects' whole brain (Zang et al., 2007).

Statistical analyses

For each of the 31 groups of predictors of the eight resting state fMRI modalities we used an elastic net logistic regression model to classify the subjects as either AD or control. Elastic net regression is commonly used for neuroimaging classification studies (Teipel et al., 2017a; Nir et al., 2016; Trzepacz et al., 2016). We used the glmnet package (Friedman et al., 2010; Zou & Hastie, 2005) available for R (R version 3.1.2, R Core Team, 2014). Elastic net regression uses penalties to hinder the predictors from entering the regression model (Friedman et al., 2010; Zou & Hastie, 2005). Thus, only the most relevant predictors will enter the regression model, which is helpful if the number of predictors outnumbers the number of subjects. Elastic net regression uses a combination of an L1 (LASSO) (Tibshirani, 1996) and L2 (Ridge) (Hoerl and Kennard, 1970) penalty. Therefore, two hyper parameters should be set: the α parameter determines the relative weight of the two different penalties and λ determines the size of those penalties. Elastic net logistic regression has already been used for AD classification (de Vos et al., 2016; Schouten et al., 2016; Schouten et al., 2017; Teipel et al., 2015; Trzepacz et al., 2014). For the combined classification model, we concatenated the 31 groups of predictors, resulting in a combined set containing 2,876,251 predictors. These predictors were jointly included in the group lasso model (Simon et al., 2013) and we informed the group lasso with an index vector that indicates the group membership of the predictors. The group lasso is similar to the elastic net, but sparse with respect to groups of predictors. This improves interpretation of the combined model, because a modality is either entirely included or excluded from the prediction model. We used the SGL package (Simon et al., 2013) available for R (R version 3.1.2, R Core Team, 2014).

We used cross-validation to ensure that we are not overfitting the prediction models. In our case there are two potential sources of overfitting. We could either include too many predictors in our logistic regression model or we could overestimate the

classification accuracy by looping over all the values of the hyper parameters and only pick the best result. To ascertain that we are not subject to any of these two sources of overfitting we used a nested cross-validation approach (Krstajic et al., 2014). We used the inner loop of the nested cross-validation to tune the hyper parameters and the outer loop to fit and test the logistic regression model. For both the inner and outer loop we used 10-fold cross-validation, thus using 90 percent of the subjects in the training set and 10 percent in the test set, and repeating this 10 times such that all subjects were part of the test set once.

We made receiver operating characteristic (ROC) curves and calculated the area under the curve (AUC) as a measure of classification performance. The AUC is invariant to the class distribution (Bradley, 1997; Fawcett, 2004), which is an advantage since the number of control subjects is larger than the number AD patients. We also calculated sensitivity, specificity and balanced accuracy values for those classification cut-offs that resulted in the highest balanced accuracy. We repeated the cross-validation procedure 10 times to get a more reliable cross-validation error (Krstajic et al., 2014) and extracted the mean AUC value.

In order to statistically compare the AUC values, we used bootstrap tests for paired AUCs (Hanley and McNeil, 1983) implemented in the pROC package (Robin et al., 2011) available for R (R version 3.1.2, R Core Team, 2014). For the comparison of the different resting state fMRI measures we used two-sided tests, because we have not formulated any directed hypotheses for these comparisons. To compare the combined model with the single measures we applied one-sided hypothesis tests, because we hypothesised that the combined model would outperform the single measures. We present uncorrected p-values and Bonferroni corrected p-values. The Bonferroni correction was applied separately to the inter measure comparisons and the comparisons of the single measures with the combined model.

3.3 Results

Classification results

Figure 2 shows the AUC values for the 31 different types of resting state fMRI measures and the combined model. Table 3 also presents values for sensitivity, specificity and balanced accuracy. The AUC values range between 0.51 and 0.84. The functional connectivity matrices (AUC values between 0.72 and 0.82) and the FC dynamics

(AUC values between 0.72 and 0.84) distinguish AD patients and controls quite well. Particularly the sparse partial correlations between the 70 ICA components (AUC = 0.82) and the standard deviations of these sparse partial correlations over time (AUC = 0.84) have high AUC values. Also, the ALFF measures are discriminative for AD. ALFF has an AUC value of 0.82, and fALFF has an AUC value of 0.69. The FC states (AUC values between 0.55 and 0.74) and the graph metrics (AUC values between 0.70 and 0.79) have reasonable classification accuracies. Functional connectivity with the 10 RSNs (AUC values between 0.52 and 0.71) mostly performs poorly, except for FC with the default mode network (AUC = 0.70) and the executive control network (AUC = 0.71). FC with the left (AUC = 0.59) and right (AUC = 0.51) hippocampus result in poor classification performances and Eigenvector centrality mapping results in moderate classification performance (AUC = 0.69). As shown in Figure 2 on the right, the combination of all the resting state fMRI measures using the group lasso model results in an AUC value of 0.85, which is higher than any of the measures used alone. Combining resting state fMRI measures thus seems beneficial, although the effect is small.

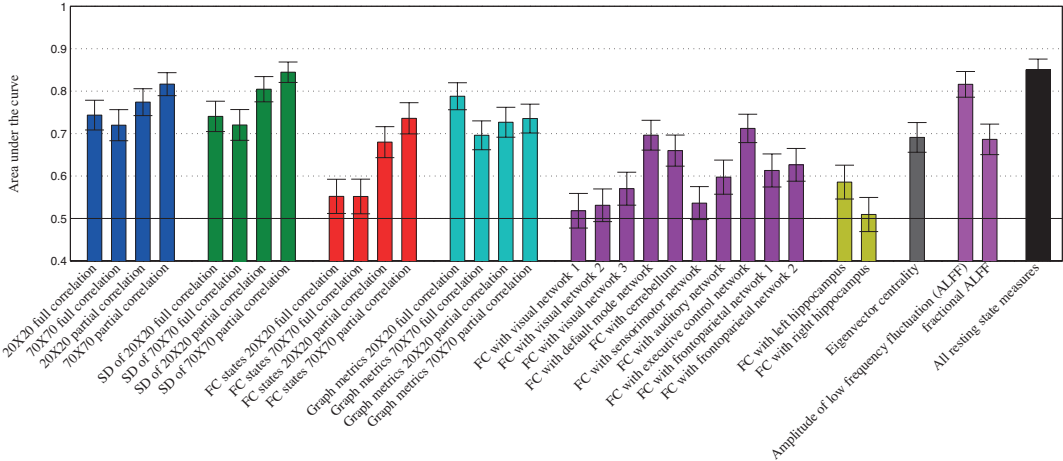


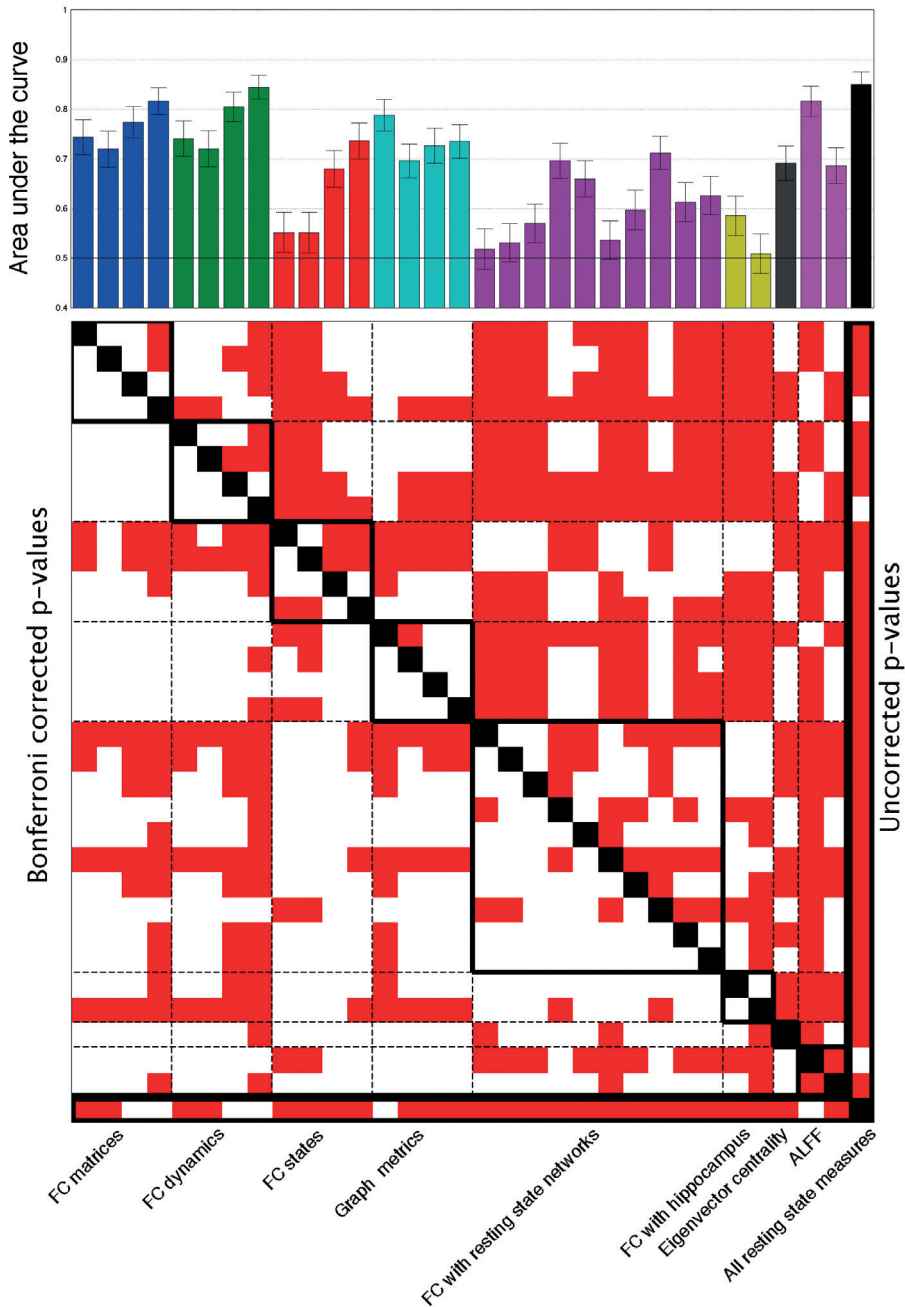
Figure 2. Area under the receiver operating characteristic curve (AUC) values for all the resting state fMRI measures. The wide bar on the right is the AUC value for the combination of all resting state fMRI measures. The error bars represent one standard error above and below the AUC values.

Table 3. Alzheimer's disease classification performance for the resting state fMRI Figure S1. Correlations between the 31 resting state fMRI measures. For each measure, we ran a principal component analysis (PCA) and we cross correlated the component scores of all the 31 first components..

Resting state measure	AUC ¹	Sensitivity	Specificity	Balanced accuracy
1: FC² matrices				
1a. 20 X 20 full correlation	0.74	0.73	0.68	0.71
1b. 70 X 70 full correlation	0.72	0.62	0.77	0.69
1c. 20 X 20 sparse partial correlation	0.77	0.68	0.76	0.72
1d. 70 X 70 sparse partial correlation	0.82	0.79	0.71	0.75
FC dynamics				
2a. SD ³ of 20 X 20 full correlation FC matrix	0.74	0.67	0.74	0.7
2b. SD of 70 X 70 full correlation FC matrix	0.72	0.70	0.69	0.69
2c. SD of 20 X 20 sparse partial correlation FC matrix	0.80	0.76	0.76	0.76
2d. SD of 70 X 70 sparse partial correlation FC matrix	0.84	0.83	0.73	0.78
3: FC states				
3a. FC states of 20 X 20 full correlation FC matrix	0.55	0.39	0.75	0.57
3b. FC states of 70 X 70 full correlation FC matrix	0.55	0.54	0.60	0.57
3c. FC states of 20 X 20 partial correlation FC matrix	0.68	0.60	0.71	0.66
3d. FC states of 70 X 70 partial correlation FC matrix	0.74	0.72	0.69	0.70
4: Graph metrics				
4a. Graph metrics of 20 X 20 full correlation FC matrix	0.79	0.79	0.68	0.74
4b. Graph metrics of 70 X 70 full correlation FC matrix	0.70	0.74	0.61	0.68
4c. Graph metrics of 20 X 20 partial correlation FC matrix	0.73	0.75	0.65	0.70
4d. Graph metrics of 70 X 70 partial correlation FC matrix	0.74	0.72	0.69	0.71
5: FC with resting state networks				
5a. FC with visual network 1	0.52	0.46	0.64	0.55
5b. FC with visual network 2	0.53	0.35	0.77	0.56
5c. FC with visual network 3	0.57	0.48	0.68	0.58
5d. FC with default mode network	0.70	0.67	0.66	0.67
5e. FC with the cerebellum	0.66	0.60	0.68	0.64
5f. FC with sensorimotor network	0.54	0.45	0.67	0.56
5g. FC with auditory network	0.60	0.68	0.52	0.60
5h. FC with executive control network	0.71	0.76	0.62	0.69
5i. FC with frontoparietal network 1	0.61	0.50	0.74	0.62
5j. FC with frontoparietal network 2	0.63	0.60	0.65	0.62
6: FC with Hippocampus				
6a. FC with left hippocampus	0.59	0.51	0.66	0.59
6b. FC with right hippocampus	0.51	0.35	0.74	0.55
7: Eigenvector centrality				
Fast eigenvector centrality mapping	0.69	0.66	0.66	0.66
8: ALFF⁴				
8a. ALFF	0.82	0.71	0.82	0.76
8b. fALFF ⁵	0.69	0.71	0.61	0.66
All resting state fMRI measures combined	0.85	0.86	0.71	0.79

¹AUC = area under the receiver operating characteristic curve, ²FC = functional connectivity,

³SD = standard deviation, ⁴ALFF = amplitude of low frequency fluctuations, ⁵fALFF = fractional amplitude of low frequency fluctuations.



Combined classification model

Figure 4 shows the contribution to the combined model for each of the 31 resting state fMRI measures. The y-axis represents the sum of the absolute standardised beta values for all the predictors within a resting state measure. A high value represents an important role for that group of predictors within the combined model. In order to quantify the spread of the contributions we fitted the group lasso model repeatedly on 100 bootstrap samples. The 100 results are represented by the boxplots. In line with the results of the single modalities, the FC matrices and the FC dynamics largely contribute the combined prediction model. There is also some contribution of the FC states and the graph metrics. Remarkably, ALFF hardly contributes, despite its discriminative power when used alone. There is considerable spread in the contribution of the resting state fMRI measures as shown by the 100 bootstrap results. None of the resting state fMRI measures contributes to the group lasso model in each bootstrap sample. However, it remains clear that the FC matrices and FC dynamics are important for the combined prediction model, whereas the other resting state fMRI measures contribute minimally or not at all.

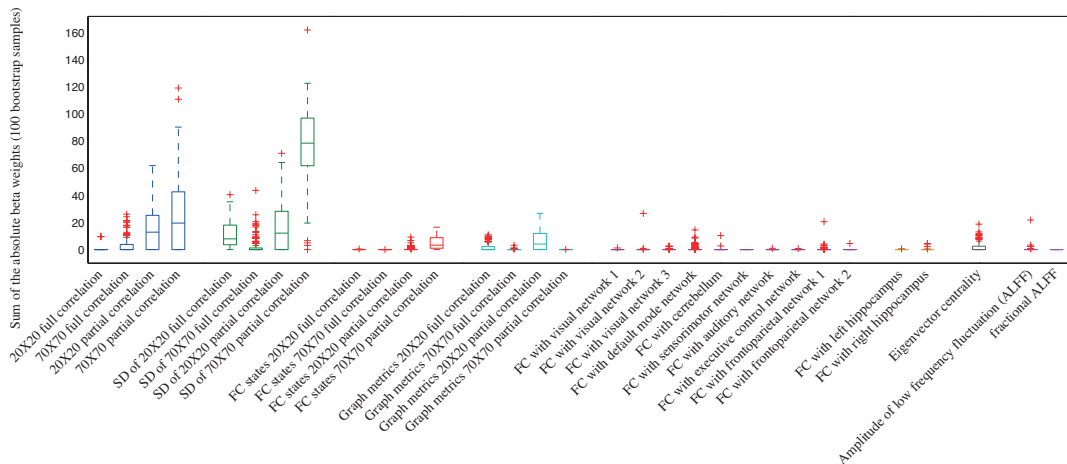


Figure 4. Importance of each resting state measure for the combined model. The combined model is fitted on 100 bootstrap samples to display the spread of the importance's. The importance is quantified by the sum of the absolute beta weights of all the predictors within a resting state measure category.

Supplementary analyses

Relation between different resting state fMRI measures

To explore relations between the resting state fMRI modalities, we calculated correlations between the 31 different resting state fMRI measures. These are presented in supplementary Figure S1. Straightforward calculation of correlation coefficients between the 31 resting state fMRI measures was not possible, because each resting state fMRI measure contains multiple predictors and the number of predictors is different for every measure. To overcome this problem, we ran a principal component analysis (PCA) for each of the 31 measures and cross correlated the component scores of all the 31 first components. Not surprisingly, resting state fMRI measures within the same modality are generally highly related. In addition, FC matrices, FC dynamics, FC states and ALFF appear to be related to each other.

Functional connections important for classification

To explore which of the ICA components were most important for AD classification, we plotted the mean beta values over all cross-validation folds and cross-validation repetitions for the FC matrices in supplementary Figure S3. Functional connectivity values between higher components have larger beta weights than FC values between lower components. Figure S2 shows that higher components are in fact real functional networks, whereas some lower components are noise components. This suggests that information on real functional networks was contributing to the classifier.

Percentage of non-zero parameters

For each resting state measure, we looked at the percentage of predictors that contributed to the classification model. Figure S4 shows the mean percentage of non-zero parameters over all cross-validation folds and cross-validation repetitions for each resting state measure. The percentages are mostly over 20%, indicating that for most measures many predictors are included in the classification model.

Voxel-wise vs. averaging over regions

The AUC values for resting state fMRI modalities one to four are mostly higher than the AUC values for resting state fMRI modalities five to eight. One notable difference between these two groups is the number of predictors. The number of predictors within resting state fMRI modalities one to four ranges from five to 2415 per category, whereas resting state fMRI modalities five to eight are voxel-wise maps and they contain 190981 predictors per category. To explore the possibility that the number

of predictors influences the classification performance, we averaged the voxel-wise maps over the 70 components as obtained by the high dimensional ICA and reran the classification analyses with the reduced number of predictors. Figure S5 shows both the original AUC values and the AUC values after averaging over the 70 components. The differences are small and the ranges of the different cross-validation repetitions are most of the time overlapping. The low classification performance for some of these categories seems not to be caused by the large number of predictors.

Optimal value analysis for the number of ICA components

We investigated the optimal number of ICA components for our connectivity analyses. We ran ICA analyses for 5 to 100 components with steps of five. For each number of components, we calculated connectivity matrices with both full or sparse partial correlations, and the dynamics of these connectivity matrices. The results are plotted in Figure S6. Calculating only five components seems to be too few, but upwards of 10 components the results are too diverse to draw conclusions on the optimal number of components.

3.4 Discussion

In this study, we determined the accuracy of different resting state fMRI measures for the individual classification of AD patients. We used machine learning techniques for efficient use of resting state fMRI measures in prediction models. FC matrices, FC dynamics, and ALFF show best discrimination between AD patients and control subjects. The combination of all the resting state fMRI measures improved the classification accuracy slightly, but not significantly. The FC matrices and the FC dynamics largely contribute to this combined model, whereas the other resting state fMRI measures are mostly redundant. This suggests that only FC matrices and FC dynamics need to be calculated to achieve optimal individual AD classification through a resting state fMRI scan.

FC matrices have been used successfully for AD classification before (Challis et al., 2015; Chen et al., 2011). Our results add to this conclusion and furthermore show that FC as calculated with sparse partial correlation results in higher classification accuracy than FC as calculated with full correlation. Likely, this is due to the fact that sparse partial correlations provide better FC estimates than full correlations (Smith et al., 2011). In addition, FC between 70 components resulted in somewhat higher classification

accuracy than FC between 20 components. This is in line with the observation that high dimensional ICA solutions provide a more specific representation of functional regions, and consequently FC between these regions results in better AD classification (Dipasquale et al., 2015). The dynamics of the FC matrices resulted in higher classification accuracy than the FC matrices itself. FC dynamics as opposed to static FC is a relatively unexplored domain in AD, but it has been shown that AD patients differ in their FC dynamics compared to controls (Chen et al., 2016; Jones et al., 2012; Wee et al., 2016). ALFF resulted in good classification accuracy, similar to Dai et al (2012). However, it did not provide additive value over the FC matrices and the FC dynamics for the combined model. For this reason, ALFF does not seem to be necessary for an AD classification model. FC states and graph metrics had reasonable classification accuracies, but they did not provide additive value for the combined model either. Furthermore, these two measures are derived from the FC matrices and FC dynamics, which themselves have higher classification accuracies. Functional connectivity states and graph metrics thus require more work to calculate and they do not seem to be beneficial over the simpler measures.

FC with the ten RSNs resulted mostly in poor classification accuracies. Exceptions are FC with the default mode network and FC with the executive control network. This corresponds to studies reporting abnormal FC in these networks in AD patients (Agosta et al., 2012; Binnewijzend et al., 2012; Greicus et al., 2004). FC with the hippocampus also resulted in poor classification accuracy, despite abnormal hippocampal connectivity patterns observed in AD patients (Allen et al., 2007; Supekar et al., 2008; Wang et al., 2006). These effects are probably not sufficiently consistent for AD classification. Possibly this is due to the fact that the hippocampus is not persistently connected with the cortex, but follows a context dependent connectivity pattern (Huijbers et al., 2012).

Some settings we have not explored. We used an ICA to determine regions as input to the FC analysis (Allen et al., 2012; Hutchison et al., 2013; Jones et al., 2012; Rashid et al, 2014), where others have used the automated anatomical labeling (AAL) atlas (Chen et al., 2011; Wee et al., 2016). We chose an ICA, because it is a data-driven approach that results in spatially independent components well suited for FC analyses. We used a group ICA and imposed the group components onto each subject (Dørum et al, 2017; Miller et al, 2016), because it is important to strive for the same parcellation in each subject in order to compare connectomes across subjects (Smith et al., 2013). We used the ICA components directly as nodes for the FC analysis. Others have used a follow up procedure to split ICA components into multiple nodes and use these nodes as input

to the FC analysis (Shirer et al., 2012; Jones et al., 2012; Shaw et al., 2015). We have not explored this option, but we obtained a similar result using the high dimensional ICA solution. When extracting a higher number of components, large networks split into multiple smaller networks. This can be observed in Figure S2. For example, component 1 of the 20 components solution splits into components 1, 6 and 13 of the 70 components solution. For the calculation of FC dynamics, we used blocked sliding windows covering 11 volumes (33 seconds) and we shifted the windows one volume at a time (Jones et al., 2012). Other methods have been reported, using tapered windows (Allen et al., 2012; Wee et al., 2016), different window sizes (Chen et al., 2016; Wee et al., 2016) and larger window shifts (Wee et al., 2016). Further, we quantified FC dynamics by calculating the standard deviations of the sliding window FC estimates. Alternatives are the dwell time in default mode network sub-configurations (Jones et al., 2012), graph measures obtained from the sliding window matrices (Wee et al., 2016), or higher order FC statistics that capture the covariance of the sliding window FC time series (Chen et al., 2016). We have not explored these methodological settings to study their effect on classification accuracy. For the analyses, we stuck to the default settings for that specific analysis as much as possible, and if there was no clear default setting, we based our choices on previous literature. We chose not to optimise the parameter settings within our study, because this would be computationally infeasible. Proper parameter optimisation must be performed using cross-validation, and in our case this would expand the cross-validation analyses considerably, because of the large number of predictors (~2 million) and the high number of parameters that can be optimised. In addition, the resting state fMRI scans used in this study covered 7.5 minutes, which is short for estimation of FC dynamics (Hindriks et al., 2016). It is not known whether classification accuracy improves with longer scan times.

For this study, we used only one sample to both train and test our prediction models. We have carefully used cross-validation techniques to prevent overfitting and obtain realistic accuracy estimates. Nevertheless, when applying these prediction models to other samples scanned at different scanner sites, we might find reduced classification accuracies, because these models are fine-tuned on the current sample. To evaluate the robustness of our classification models they have to be applied to a different sample. In addition, the current sample was not the result of random sampling from a prespecified population. The conclusions from the statistical tests that we have performed therefore only apply to the current sample.

3.5 Conclusion

In conclusion, we demonstrated the use of resting state fMRI scans for individual AD classification. The optimal combination of resting state fMRI measures comprises FC matrices and FC dynamics. These results may direct future studies that use resting state fMRI scans for the classification of patients with preclinical AD or mild cognitive impairment.

3.6 Acknowledgments

This study is supported by VICI grant no. 016.130.677 of the Netherlands Organisation for Scientific Research (NWO).

PRODEM is supported by funds of the Austrian Alzheimer society.

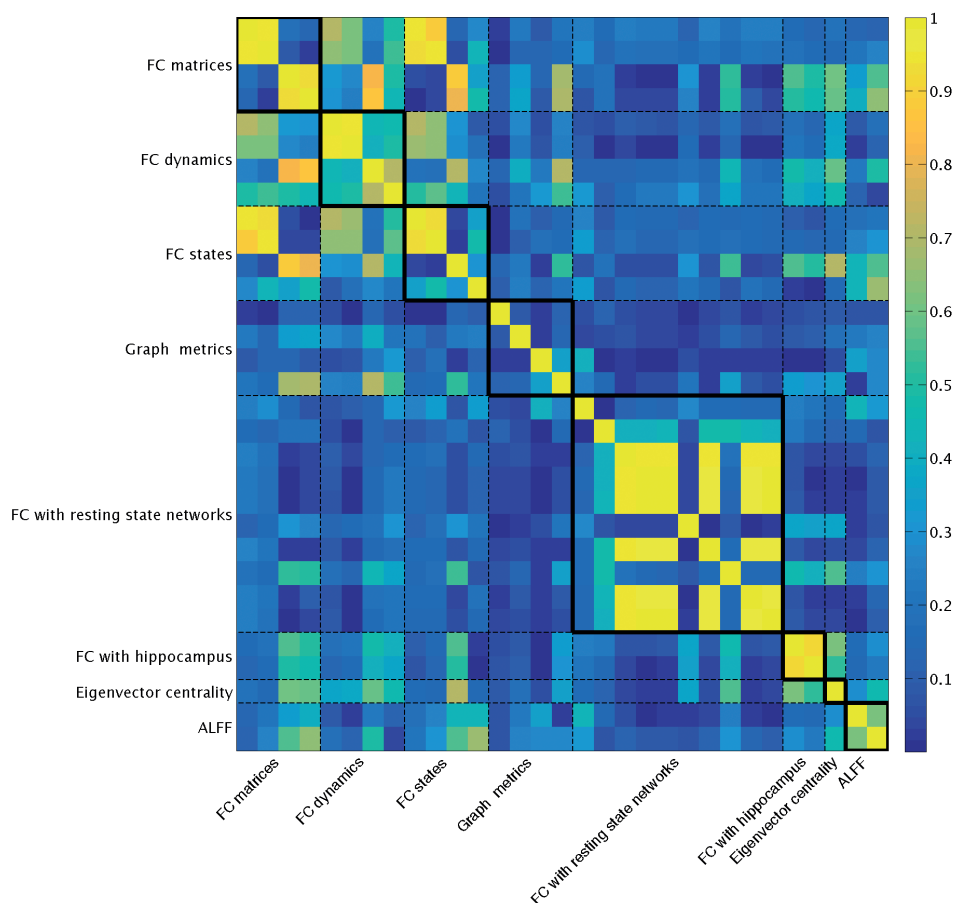


Figure S1. Correlations between the 31 resting state fMRI measures. For each measure, we ran a principal component analysis (PCA) and we cross correlated the component scores of all the 31 first components.

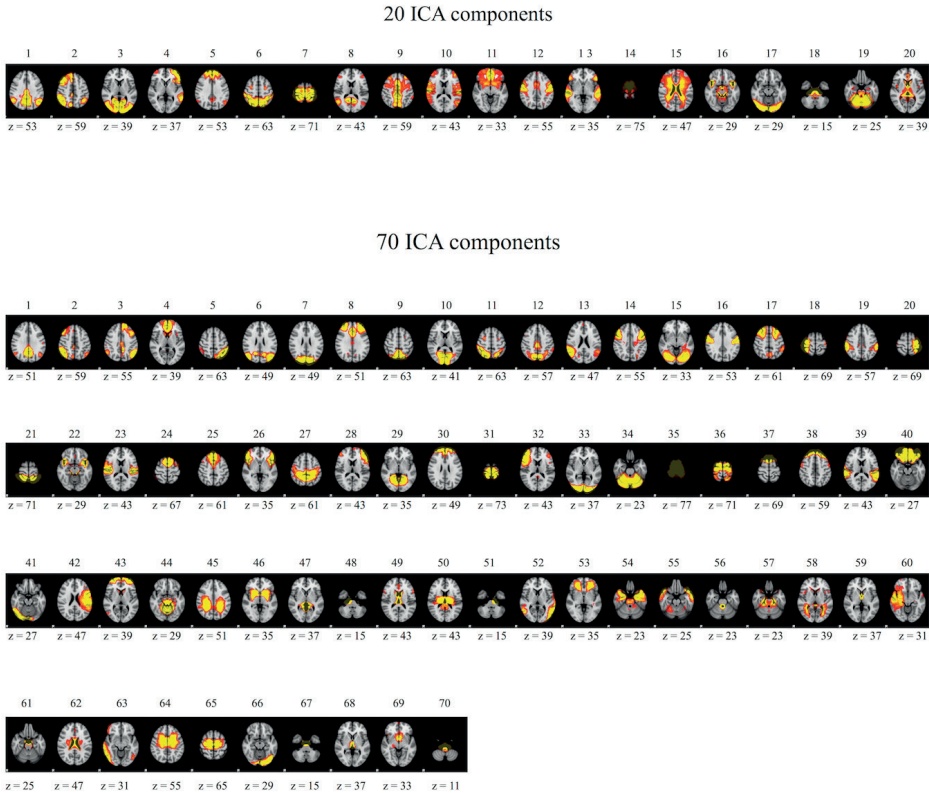


Figure S2. The 20 and 70 components extracted from the low and high dimensional independent component analysis (ICA).

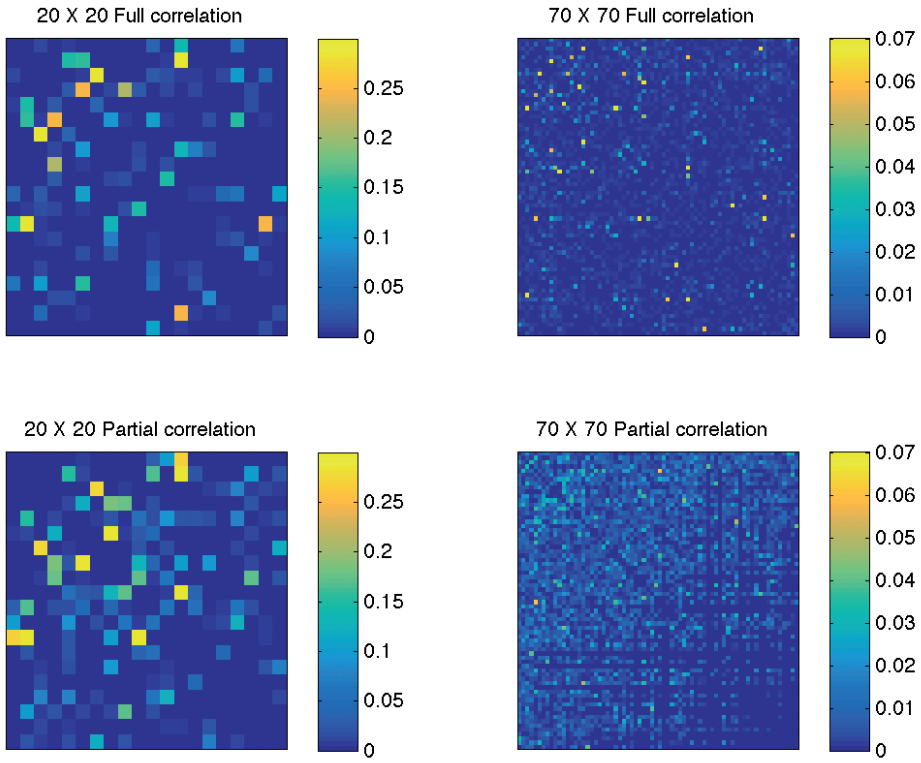


Figure S3. Mean beta values for the functional connectivity matrices. Beta values are averaged over the multiple cross-validation folds and multiple cross-validation repetitions.

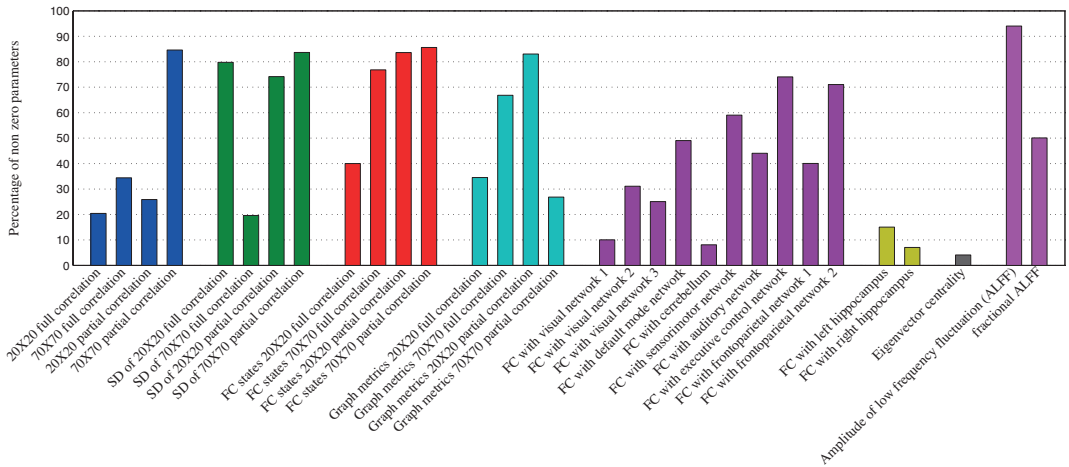


Figure S4. Mean percentage of non-zero parameters over all cross-validation folds and cross-validation repetitions for the 31 resting state fMRI measures.

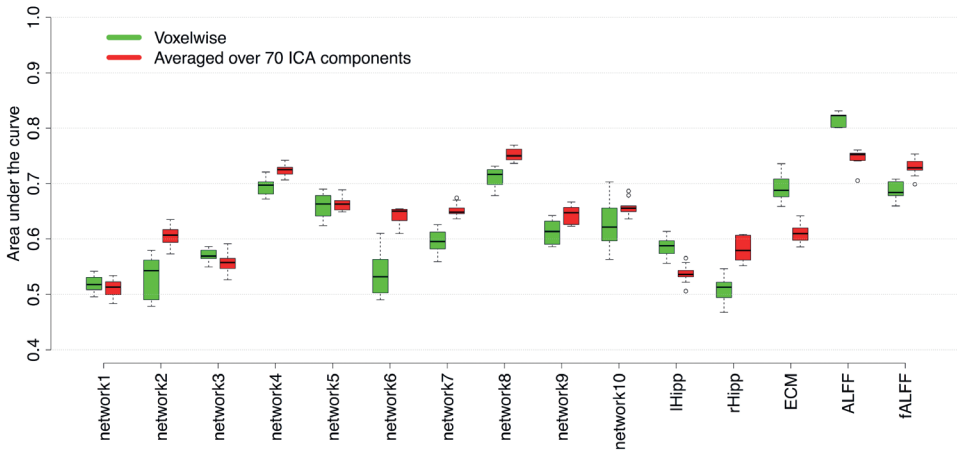


Figure S5. The effect of averaging over regions. The original classification results (green) and the results when the voxel-wise data is averaged over the 70 ICA components (red). The boxplots represent the different cross-validation repetitions.

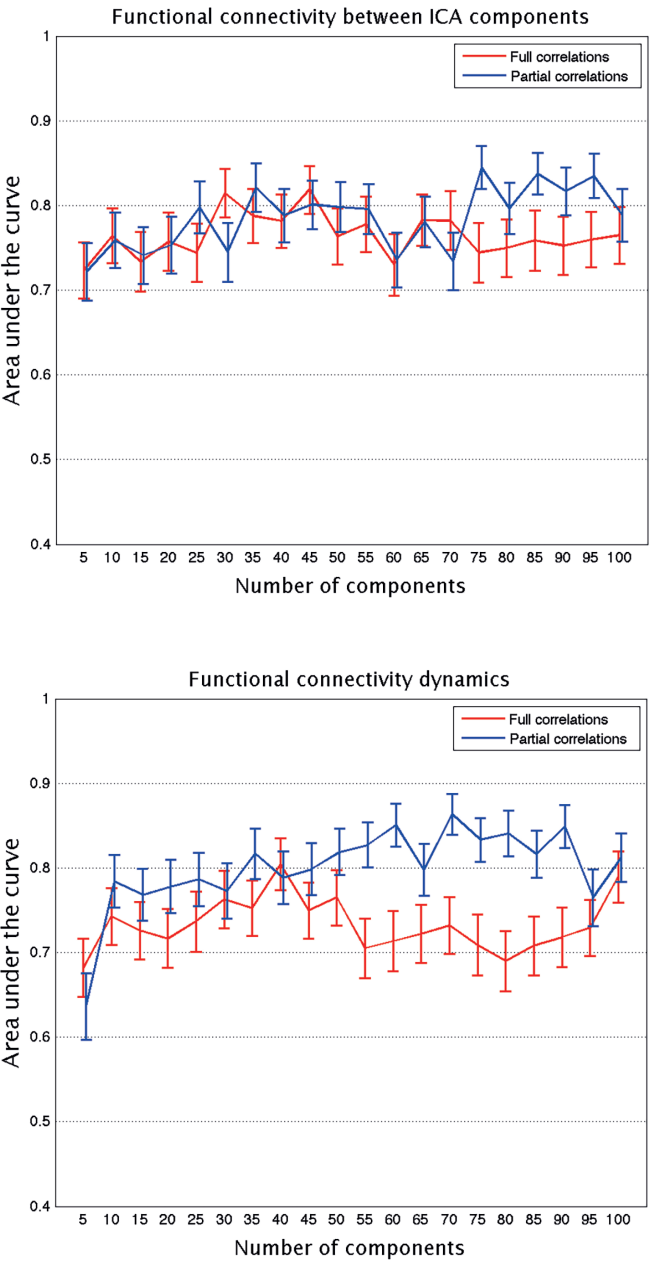


Figure S6. AUC values for the functional connectivity between ICA components (top), and functional connectivity dynamics (bottom) for a range of numbers of ICA components. The error bars represent one standard error above and below the AUC values.

