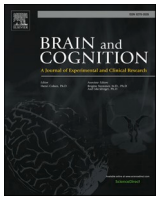
Brain and Cognition 196 (2026) 106443

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c



Full Length Article

Structural complexity of brain regions in mild cognitive impairment and Alzheimer's disease

Roni Tibon [a,b,*], Christopher R. Madan [a], Delshad Vaghari [c], Constantino Carlos Reyes-Aldasoro [b]

[a] School of Psychology, University Park, University of Nottingham, Nottingham NG7 2RD, UK
[b] Department of Computer Science, School of Science & Technology, City St George's, University of London, London EC1V 0HB, UK
[c] Department of Psychology, University of Cambridge, Cambridge CB2 3EB, UK

ARTICLE INFO

Keywords:
Alzheimer'sdisease
MildCognitiveImpairment
MagneticResonanceImaging(MRI)
FractalDimensionality
StructureComplexity
MachineLearning

ABSTRACT

Early detection of Alzheimer's disease (AD) is a major focus of current research efforts to guide early interventions. Subtle neural changes might be observed even before symptoms surface. We interrogated brain images obtained with Magnetic Resonance Imaging (MRI) from two large-scale dementia datasets (ADNI and BioFIND) to establish the utility of fractal dimensionality (FD)—an understudied measure that estimates the complexity of 3D structures (in this case, brain regions)—for AD detection. We show that FD measures are consistent across the two datasets, and can be used to detect group differences between patients and controls, as well as for individual-based classification. We further show that the contribution of specific brain regions to individual-based classification adheres to previous literature on the properties of the brain's memory network and how it relates to cognition. Taken together, the study offers novel and interpretable evidence for the utility of FD for the detection of AD.

1. Background

Alzheimer's disease (AD) is an age-related neurodegenerative disorder characterised by progressive dementia, from mild memory impairment to global cognitive dysfunction and eventually death (Goedert and Spillantini, 2006). Mild Cognitive Impairment (MCI) is a common early (prodromal) stage of AD, with a high probability of progression to dementia (Petersen, 2009). It had been suggested that MCI patients may already have subtle brain changes that are identifiable with neuroimaging (Cabral et al., 2015; Chincarini, 2011; Jack, 2010; Vaghari, 2022). Therefore, in recent decades, substantial research efforts are dedicated to characterising the abnormalities that can be observed in neuroimaging, in search for reliable biomarkers of MCI and AD (Frisoni et al., 2010; Frizzell, 2022; Rathore et al., 2017).

Fractal geometry analyses have been used widely throughout the neurosciences, showing utility in a range of spatial and temporal applications (Di Ieva, 2024). Here we particularly used Fractal dimensionality (FD) (Madan and Kensinger, 2018; Madan and Kensinger, 2016), a measure used to describe the complexity of brain shapes extracted from magnetic resonance imaging (MRI) (Krohn, 2026). A handful of studies have considered this measure in the context of MCI/AD (King et al., 2010; King, 2009; Ma, 2020; McDonough and Madan, 2021; Meregalli, 2022; Nicastro, 2020; Pantoni, 2019; Ruiz de Miras, 2017; Thompson, 1998; Ziukelis et al., 2022; Ni et al., 2024; Jiang et al., 2024; Di Ieva, 2024; Marzi, 2023), generally reporting some reduction in FD in MCI/AD patients vs. controls. However, most of these studies were conducted with relatively small samples, reducing the ability to decipher potential regional effects, and the analyses focused on group-level effects, and individual-based classification remained underexplored. In addition, in all but two studies (but those have not focused on MCI), analyses were restricted to a single sample thereby limiting the ability of the findings to generalise beyond specific settings. Finally, the discussion of cognitive functioning (and memory in particular) in the context of regional effects observed with FD remains limited.

Large-scale datasets offer additional opportunities to tackle these important issues. The Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, established in 2003, is an open dataset that includes MRI T1-weighted scans from N~500 participants (across multiple subcohorts and sessions) grouped into AD patients, MCI patients and healthy controls (HC) (Jack, 2008). A more recent resource is the

* Corresponding author at: School of Psychology, University Park, The University of Nottingham, Nottingham NG7 2RD, UK.
E-mail address: roni.tibon@nottingham.ac.uk (R. Tibon).

https://doi.org/10.1016/j.bandc.2026.106443
Received 16 February 2026; Received in revised form 29 April 2026; Accepted 14 May 2026
0278-2626/© 2026 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

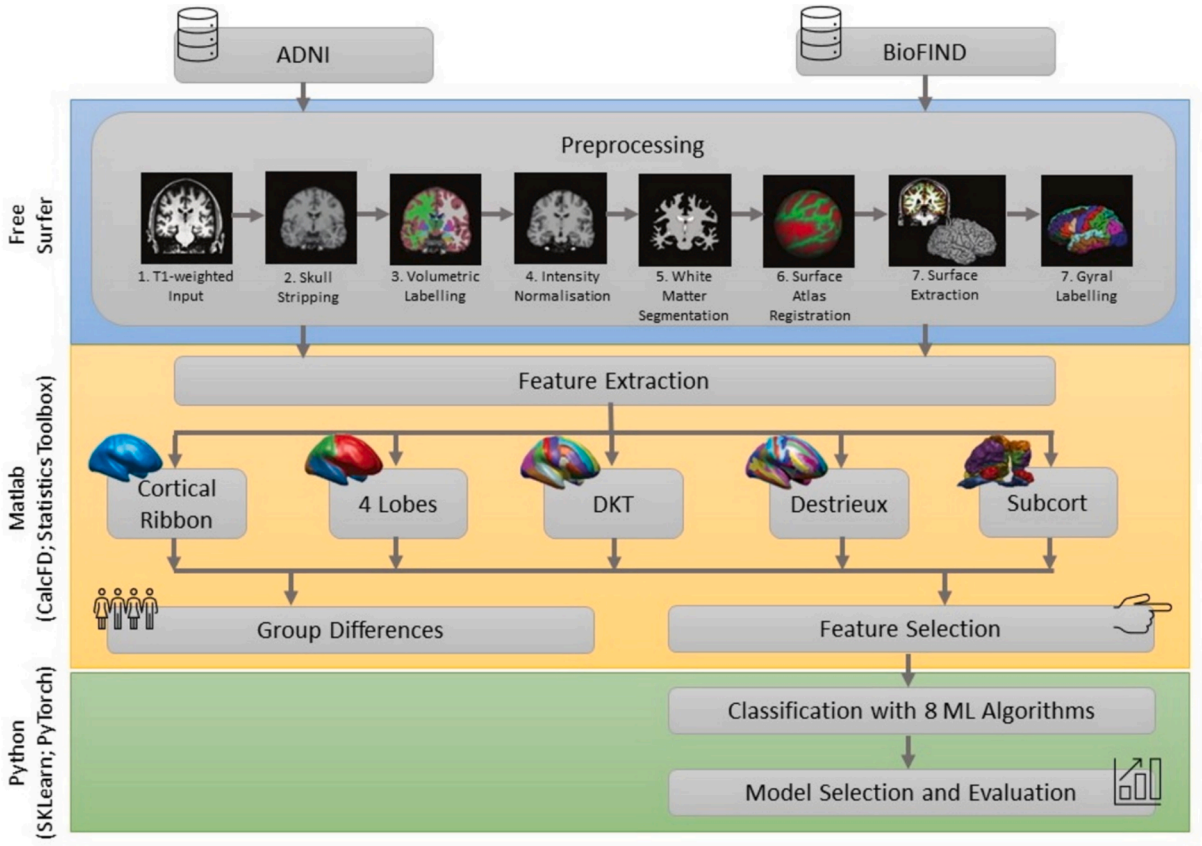


Fig. 1. Overview of approaches and methods. Abbreviations: *ADNI*: Alzheimer's Disease Neuroimaging Initiative; *CalcFD*: calculate fractal dimensionality; *SKLearn*: scikit-learn; *DKT*: Desikan-Killiany-Tourville; *Subcort:* subcortical; *ML*: machine learning.

Table 1
Estimates of group differences.

Parcellation	Statistic	P-value	Corrected p-value
BioFIND	*t*(305)		
Cortical ribbon	4.69	0<.001	0<.001**
4 lobes	4.38	0<.001	0<.001**
DKT	3.04	0=.003	0=.015*
Destrieux	3.08	0=.002	0=.01*
Subcortical	7.81	0<.001	0<.001**
ADNI	*F*(2,149)		
Cortical ribbon	5.31	0=.006	0=.03*
4 lobes	5.05	0=.008	0=.04*
DKT	5.64	0=.004	0=.02*
Destrieux	4.77	0=.01	0=.05
Subcortical	13.46	0<.001	0<.001**

For Bonferroni corrected p-values: $^{**}p < 0.01$; $^{*}p < 0.05$.

BioFIND dataset, which includes T1-weighted scans from N~320 participants grouped into MCI patients and HC (Vaghari, 2022). Whilst BioFIND was launched recently, and for the time being only yielded eight publications (Bruña, 2022; Vaghari et al., 2022; Javed, 2025; Martínez-Cañada, 2023; Liu et al., 2024; Gaubert, et al., 2025; Nugent et al., 2025; Estarellas et al., 2024), ADNI was extensively studied over the last decades, yielding, according to the project's website (https://adni.loni.usc.edu/), more than 3,000 published peer-reviewed articles. Despite this productivity, the enthusiasm might be tempered by the problematic consequences of repeatedly analysing the same data. Namely, extensive use of a single dataset enhances the risk of increasing false positives and limits the ability to generalise conclusions beyond that specific data (Madan, 2022; Tibon et al., 2022; Borchert, 2023). Indeed, it was recently found that studies that used an independent dataset for validation reported much lower accuracy than studies that validated models against held-out observations obtained within the same settings (Borchert, 2023; Klöppel, 2015; Bucholc, 2023). Recently, several studies were conducted to assess generalisability in unseen independent research datasets (De Carli, 2019; Qiu, 2020), demonstrating the importance of validation across datasets in identifying methodological issues relevant to the overall model performance.

The aim of the current study was to establish the utility of structural complexity measures for the detection of dementia, and suggest potential links to cognitive functioning. First, we used FD for the detection of group differences between MCI/AD patients and HC. To this end, we obtained T1-weighted MRI images from two large-scale datasets (BioFIND and ADNI), calculated structural complexity across 5 parcellation schemes (Madan and Kensinger, 2016), and performed group-level statistical analyses. We predicted that FD values will be lower in MCI/AD patients vs. healthy controls. Next, we identified brain regions in which FD values were predictive of class attribution and designated them as selected features used for classification of individual cases. Then, we evaluated to which extent the results remain valid and generalise beyond a specific dataset. To this end we assessed the correspondence between measures of structural complexity in BioFIND and ADNI using feature-wise correlations at the group level, and also evaluated the performance of the classifiers (trained on ADNI) when applied to the test set (BioFIND). Finally, we explored the importance of predictive features to allow additional delineation of the potential links between cognitive functioning and FD values.

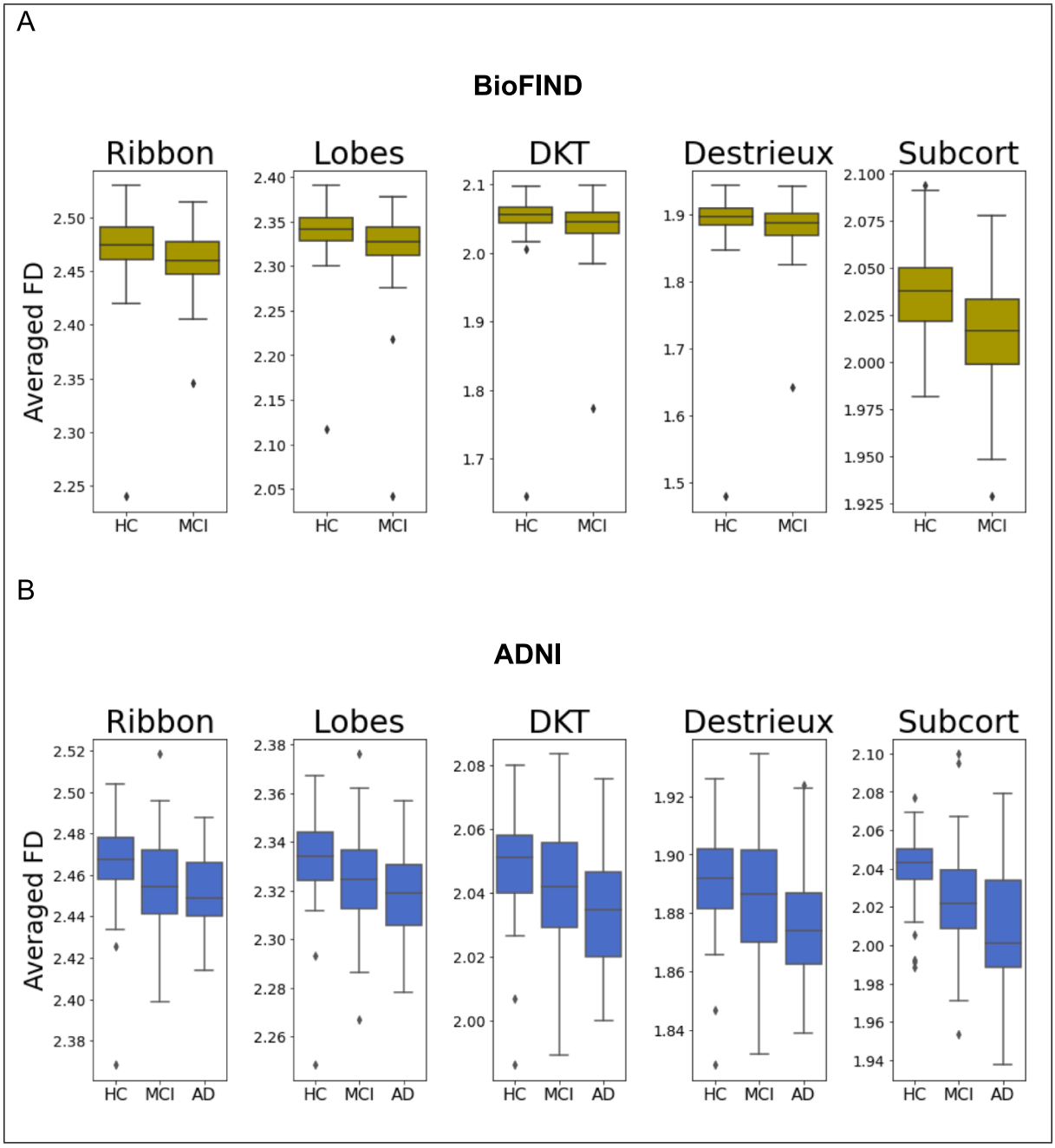


Fig. 2. Averaged FD values across each parcellation scheme for BioFIND data (A) and ADNI data (B). HC=Healthy controls; MCI = Mild cognitive impairment; AD = Alzheimer disease.

2. Methods

2.1. Datasets

2.1.1. ADNI

ADNI was launched in 2003 as an open dataset, with the primary goal of assessing whether biological markers (including MRI), and clinical and neuropsychological assessments, could be combined to measure the progression of MCI and AD. The acquisition protocol and the criteria used for the inclusion of participants are fully described in the ADNI protocol (Jack, 2008). In short, participants were 55–90 years old, had a study partner able to provide an independent evaluation of functioning, and spoke either English or Spanish. They were all willing and able to undergo all test procedures (including neuroimaging) and agreed to longitudinal follow ups.

Cognitive impairment in this sample was assessed using the Mini-Mental State Examination (MMSE)—a 30-point questionnaire commonly used in clinical and research settings to measure cognitive impairment (Folstein et al., 1975), and the Clinical Dementia Rating (CDR)—a global summary measure ranging between 0 and 3, designed to identify the overall severity of dementia (Morris, 1993). Throughout the entire ADNI dataset, Healthy Control (HC) participants had MMSE scores between 24 and 30 (inclusive), and a CDR of zero. They were non-depressed, non-MCI, and non-demented. MCI patients had MMSE scores between 24 and 30 (inclusive), a memory complaint, objective memory loss (as measured with the Wechsler Memory Scale Logical Memory II; Wechsler, 1945), a CDR of 0.5, and absence of significant levels of impairment in other domains. AD patients had MMSE scores between 20

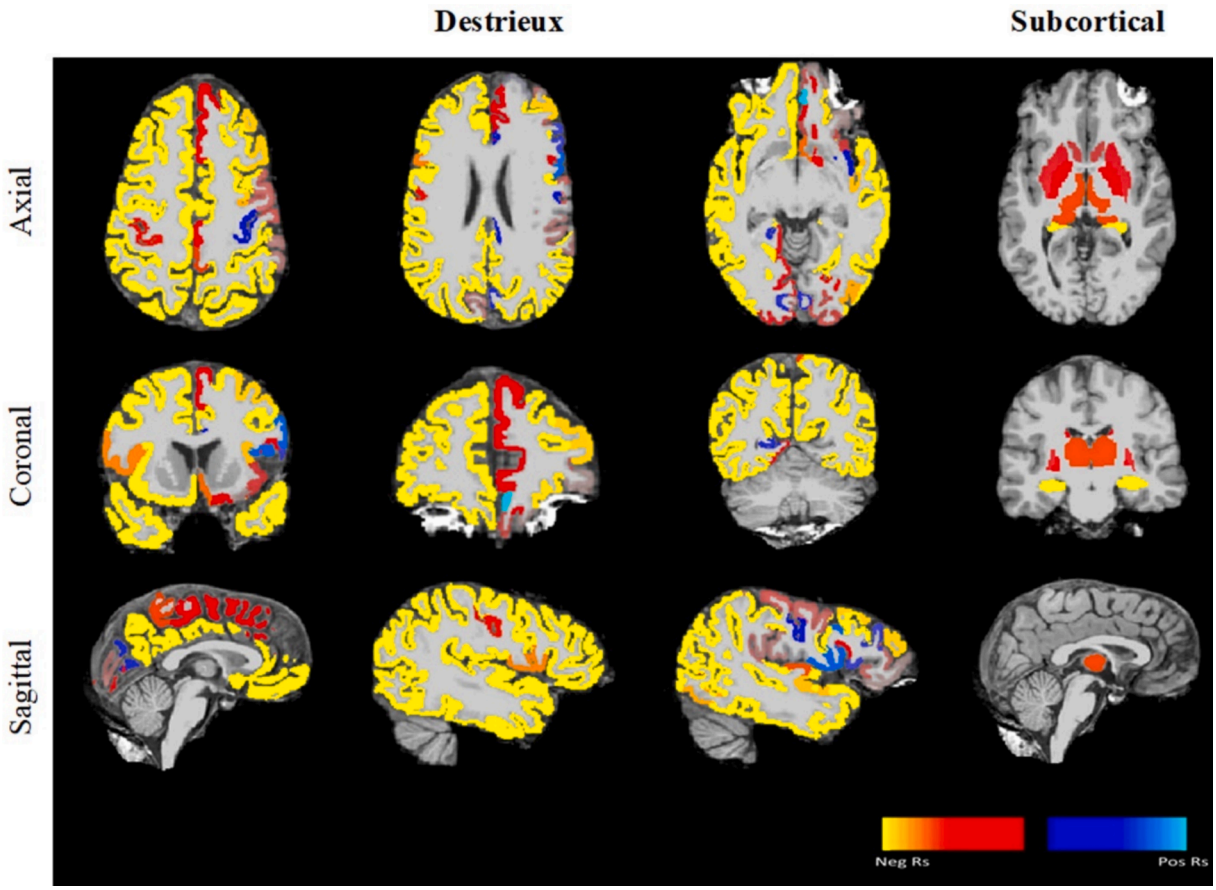


Fig. 3. Rs values for the destrieux (three left columns) and subcortical (right column) parcellation schemes, overlayed on a template mri and presented from an axial (top), coronal (middle), and sagittal (bottom) views. warm colours indicate negative values whereas cold colours indicate positive values. the greater the absolute value is, the better the ability of the region to accurately classify individuals into their respective groups. Note that due to the labelling of the groups (i.e., in the data matrix individuals in the control group were labelled as "0″ and individuals in the MCI group were labelled as "1"), negative values indicate greater structural complexity in healthy controls vs. MCI patients and vice versa. The values were adjusted to aid visualisation and therefore the exact numbers are arbitrary and not shown here.

and 26 (inclusive), CDR of 0.5 or 1.0, and met standard criteria for probable AD (McKhann, 1984).

The current sample used ADNI1 Baseline 3T collection, including N = 152 participants (78 females) with defaced T1-weighted structural MRI images. In this sample, participants' age ranged between 55 and 89 (M = 74.87, SD = 6.99). The HC group included N = 47 participants (29 females) aged 70–86 (M = 75.06, SD = 3.89), the MCI group included N = 72 participants (27 females) aged 55–88 (M = 75.13, SD = 7.97), and the AD group included N = 33 participants (22 females), aged 57–89 (M = 74.03, SD = 8). Independent sample t-tests confirmed no significant age difference between the groups (HC vs. MCI: $t(117) = 0.005$, $p = 0.96$; HC vs. AD: $t(78) = 0.76$, $p = 0.45$; MCI vs. AD: $t(103) = 0.65$, $p = 0.52$). Nevertheless, Chi-squared tests for independence showed a significant difference in gender distribution in HC vs. MCI, $\chi^2(1) = 5.75$, $p = 0.016$, and in MCI vs. AD, $\chi^2(1) = 6.61$, $p = 0.01$, but not in HC vs. AD, $\chi^2(1) = 0.05$, $p = 0.83$, resulting from greater proportion of female participants in the HC and AD groups (62% and 67%, respectively) compared to the MCI group (38%). Importantly, when the two patient groups were grouped together (as was done for most subsequent analyses, see below) the analysis showed no differences in gender distribution for patients vs. HC, $\chi^2(1) = 2.37$, $p = 0.12$.

2.1.2. BioFIND

The BioFIND project was launched recently (Vaghari, 2022) as a multi-site, multi-participant MRI and magnetoencephalography (MEG) resting-state open dataset to study dementia. Data were obtained from two sites: the MRC Cognition & Brain Sciences Unit (CBU) in Cambridge, England, and the Centre for Biomedical Technology (CTB) in Madrid, Spain. MCI diagnosis was given by a clinician based on clinical and cognitive tests, self- and informant-report, and in the absence of full dementia or other obvious causes (e.g., psychiatric).

BioFIND includes data from N = 324 individuals. For the current study, we excluded data from participants with no (defaced) T1-weighted structural MRI images (N = 15) and for which the FreeSurfer's reconstruction process (detailed below) failed (N = 2). Thus, the current study included data from N = 307 individuals (151 females) aged 52–95 (M = 71.75, SD = 6.94). The HC group included N = 165 participants (83 females) aged 53–95 (M = 71.24, SD = 7.01), and the MCI group included N = 158 participants (68 females) aged 52–90 (M = 72.77, SD = 6.77). An independent sample *t*-test did not reveal a significant age difference between the groups, $t(305) = -1.93$, $p = 0.054$, nor did a Chi-squared test for independence showed any difference in gender distribution between the two groups, $\chi^2(1) = 0.095$, $p = 0.76$. AD patients were not included in this dataset.

2.1.3. Ethics Approval and consent

This study was conducted using secondary data analysis of open resources. Ethical approval and informed consent was obtained by the ADNI and BioFIND projects.

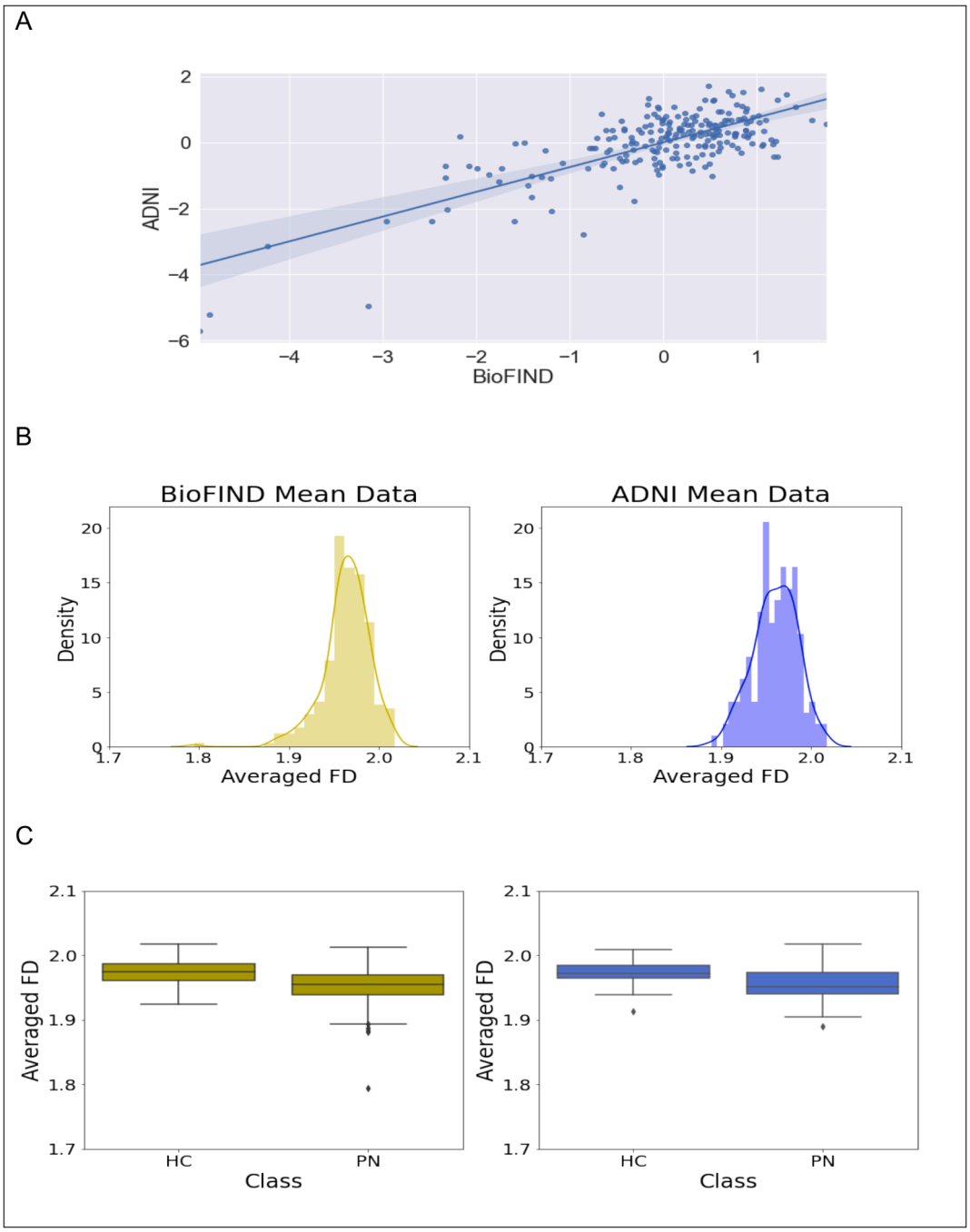


Fig. 4. Dataset characteristics. Panel A (top) shows the correlation between z-scored rs values in BioFIND (x-axis) and ADNI (y-axis). Each blue mark represents one brain region (i.e., one of 222 rs values). Solid blue line represents the linear fit for the data, whereas blue shades indicate the confidence boundaries. Panel B (middle) shows histograms for BioFIND (left) and ADNI (right) fractal dimensionality (FD) data, averaged across all features. Panel C (bottom) shows boxplots of averaged FD data in patients (PN) and healthy controls (HC) for each dataset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Analysis of T1-weighted MRI data

2.2.1. Preprocessing

An overview of the approaches and methods that were used is shown in Fig. 1. As shown in the figure, we used the same preprocessing pipeline for both datasets. Structural MRI data were preprocessed using FreeSurfer v.7.3.0 (https://surfer.nmr.mgh.harvard.edu), to automatically segment the volumetric datasets and parcellate the cortex from the T1-weighted images (Fischl, 2002). FreeSurfer's standard pipeline was used (i.e., recon-all), with no manual edits to the surface meshes. With this pipeline, a two-dimensional cortical surface is reconstructed from a three-dimensional volume. First, the skull was stripped from the anatomical image to generate a mask that only contains the brain. Then, the interface between the white matter and grey matter for both hemispheres was estimated. This initial boundary was refined and then used as a base from which recon-all extends feelers to search for the edge of the grey matter. Once this edge was reached, datasets representing the pial surface were created, an inflated derivative of these surfaces was inflated again into a sphere, normalised to a template image that contains an average of 40 subjects, and then re-shaped into an inflated

Table 2
Brain regions used for ML classification. Note that (1) regions that constitute part of the brain's memory network are indicated with an asterisk, and (2) the left-lateralisation pattern that was observed in cortical regions (right column). These are considered in the discussion.

ID	FreeSurfer Label	Label Abbreviation	Label Description	Laterality
Destrieux				
*8	G_and_S_cingul-Mid-Post	MPosCgG/S	Middle–posterior part of the cingulate gyrus and sulcus	Left
*9	G_cingul-Post-dorsal	PosDCgG	Posterior–dorsal part of the cingulate gyrus	Left
18	G_insular_short	ShoInG	Short insular gyri	Left
21	G_oc-temp_lat-fusifor	FuG	Lateral occipito-temporal gyrus (fusiform gyrus, O4-T4)	Left
*23	G_oc-temp_med-Parahip	PaHipG	Parahippocampal gyrus, parahippocampal part of the medial occipito-temporal gyrus (T5)	Left
*25	G_pariet_inf-Angular	AngG	Angular gyrus	Left
*27	G_parietal_sup	SupPL	Superior parietal lobule (lateral part of P1)	Left
30	G_precuneus	PrCun	Precuneus (medial part of P1)	Left
32	G_subcallosal	SbCaG	Subcallosal area, subcallosal gyrus	Left
34	G_temp_sup-Lateral	SupTGLp	Lateral aspect of the superior temporal gyrus	Left
35	G_temp_sup-Plan_polar	PoPl	Polar plane of the superior temporal gyrus	Left
37	G_temporal_inf	InfTG	Inferior temporal gyrus (T3)	Left
*38	G_temporal_middle	MTG	Middle temporal gyrus (T2)	Left
44	Pole_temporal	TPo	Temporal pole	Left
47	S_cingul-Marginalis	CgSMarp	Marginal branch (or part) of the cingulate sulcus	Left
57	S_intrapariet_and_P_trans	IntPS/TrPS	Intraparietal sulcus (interparietal sulcus) and transverse parietal sulci	Left
61	S_oc-temp_lat	LOcTS	Lateral occipito-temporal sulcus	Left
62	S_oc-temp_med_and_Lingual	CoS/LinS	Medial occipito-temporal sulcus (collateral sulcus) and lingual sulcus	Left
68	S_postcentral	PosCS	Postcentral sulcus	Left
72	S_subparietal	SbPS	Subparietal sulcus	Left
73	S_temporal_inf	InfTS	Inferior temporal sulcus	Left
74	S_temporal_sup	SupTS	Superior temporal sulcus	Left
*23	G_oc-temp_med-Parahip	PaHipG	Parahippocampal gyrus, parahippocampal part of the medial occipito-temporal gyrus (T5)	Right
Subcortical				
1	thalamus			Bilateral
2	caudate			Bilateral
3	putamen			Bilateral
4	pallidum			Bilateral
*5	hippocampus			Bilateral
6	amygdala			Bilateral
7	accumbens			Bilateral

surface or a pial surface. Participants' individual surface maps were normalised to this template in order to allow the use of an atlas for the parcellation of the cortex into anatomically distinct regions.

2.2.2. Parcellation

Five parcellation schemes (see Fig. 1) were used for feature extraction based on FD: (1) entire cortical ribbon (one region; i.e., unparcellated); (2) each of the four lobes (four regions); (3) DKT atlas (62 regions) (Klein and Tourville, 2012); (4) Destrieux atlas (148 regions) (Destrieux et al., 2010); and (5) subcortical structures (7 regions). The DKT and Destrieux atlases are included as standard parcellation atlases within FreeSurfer. Though some of the Destrieux regions are small, limiting the utility in measuring FD, here we chose to use it as it is well-known and readily estimable; for an alternative approach see (Collantoni, 2020). The lobe parcellation was delineated by grouping together parcellated regions from the Destrieux atlas, as done in (Madan and Kensinger, 2018). Subcortical regions were defined using FreeSurfer's automatic subcortical segmentation of the brain (aseg atlas).

2.3. Calculating fractal dimensionality (FD)

To quantify the 'dimensionality' or complexity of the structures (=brain regions), the Minkowski–Bouligand (Mandelbrot, 1967) was calculated. For this calculation, the algorithm considers the 3D structure within a grid space, and the number of boxes that overlap with the edge of the structure are counted. Then, using another grid size (i.e., by changing the width of the box), the 'box-counting algorithm' is applied to determine the relationship between the grid size and number of counted boxes. The slope of this relationship in log–log space is the FD of the structure, corresponding to the equation:

$$FD = -\frac{\Delta \log_2(count)}{\Delta \log_2(size)}$$

where "count" is the number of counted boxes and "size" is the edge width of each box, and the negative sign ensures that FD is positive, as the number of occupied boxes increases when box size decreases. When the boxes overlapping with the edge and within the structure are both counted, the resulting slope represents the FD of the filled volume. To avoid potential biases due to the relative alignment of the grid space and the structure, we applied a dilation algorithm which uses a sliding grid space to calculate FD at each alignment (Madan and Kensinger, 2016). This was implemented using the calcFD Matlab's toolbox (https://cmadan.github.io/calcFD/) (Madan and Kensinger, 2016; Madan and Kensinger, 2017), which is designed to use files from the standard FreeSurfer analysis pipeline.

Using the above-described FD calculation, we extracted 10 derived datasets, one for each parcellation scheme and dataset. In each derived dataset, cases were labelled as AD (for ADNI's datasets), MCI, and HC, according to their original classification. The number of features in each of these derived datasets was equivalent to the number of regions defined by the scheme (for example, the dataset for the DKT atlas included 64 features).

2.4. Overall estimates of group differences

We estimated group differences in FD separately for each derived dataset using Matlab's Statistic Toolbox (MathWorks, Inc.). For ADNI's derived datasets, averaged FD was computed for each participant across all features and compared between groups using one-way ANOVA with group (HC, MCI, AD) as a between-subject factor. For BioFIND's derived datasets, averaged FD was compared between groups using a two-sample T-Test.

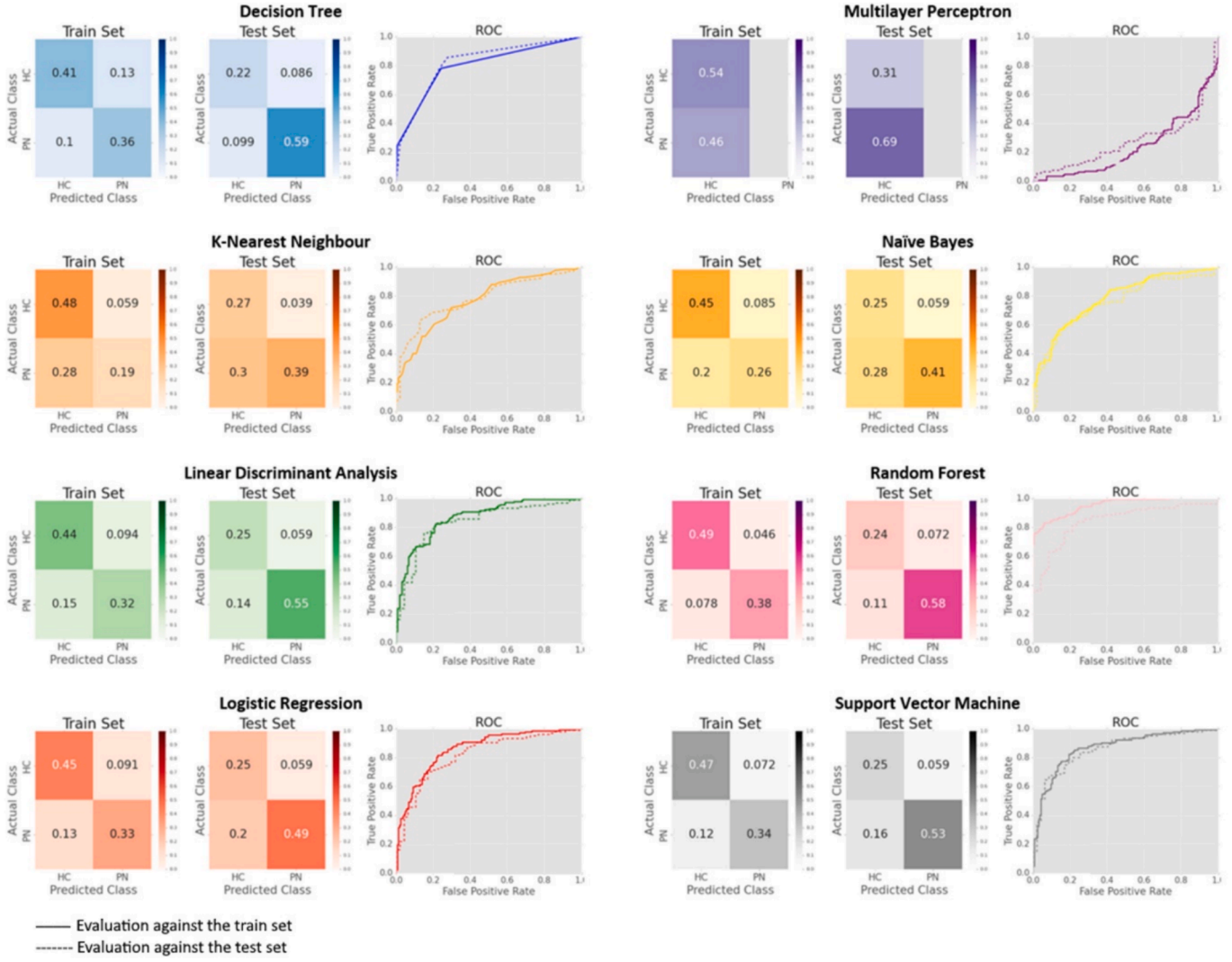


Fig. 5. Confusion matrices and ROCs for each algorithm following the model selection procedure, showing optimal performance for the Random Forest algorithm (see text for details). Evaluation was performed against the train set (left matrix for each algorithm; solid lines in ROCs) and against the test set (right matrix, dashed lines). PN: patients, HC: healthy controls.

2.5. Feature investigation and extraction

2.5.1. General linear model and structure coefficients

To identify brain regions in which FD most reliably predicts group attribution, we computed general linear models (GLMs), separately for each derived dataset. FD values in the various parcels were used as predictors, and class as the dependent variable. Because ADNI includes two classes of patients (AD and MCI), whereas BioFIND only includes one (MCI patients), ADNI data were collapsed across groups such that for both datasets, cases were classified as either patients (PN) or healthy controls (HC). We fitted the GLMs using Matlab's fitglm command with a binomial distribution (i.e., logistic regression, which is suitable for binary outcomes). This command outputs beta weights for the predictors, which can then be used for data description and subsequent predictions. However, when multicollinearity is present in the data, beta values are hard to interpret. Therefore, to support interpretation, we further computed structure coefficients (denoted r_s) by correlating the values of each predictor with the predicted score. Structure coefficients can be viewed as "loadings", indicating to what degree each variable contributes to the predictions made by the model. These coefficients can be used to guide interpretation and are particularly useful in the presence of multicollinearity (Sherry and Henson, 2005; Tibon, 2021; Tibon and Tsvetanov, 2022). Therefore, we used the computed structure coefficients for two main purposes, as detailed below: (1) to investigate the correspondence between the two datasets, and (2) to identify brain regions in which FD most reliably predicts group attribution.

2.5.2. Correspondence between ADNI and BioFIND

We concatenated the structure coefficients across all parcellation schemes, such that for each dataset (i.e., ADNI and BioFIND) a vector with 222 values (i.e., one r_s for the cortical ribbon; 4 for the 4 lobes; 62 for the DKT atlas; and so on) was generated. To assess the correspondence between the two datasets, we then correlated the r_s vectors generated from ADNI and BioFIND data using Pearson correlation. A high correlation would indicate that the same brain regions are predictive of group attribution in both datasets, and would therefore suggest high correspondence between them.

2.5.3. Feature extraction based on structure coefficients' values

To decide which features should be used for machine learning (ML) classification, for each parcellation scheme, we identified features with significant r_s coefficients (following a Bonferroni correction across the number of parcels within the scheme). Only datasets derived from BioFIND were included in this procedure, in order to avoid any biases in the classification procedure and to allow ADNI to be used as an "unseen" dataset. That is, as described below, for classification purposes, ML

Table 3
Algorithms' scores on various metrics. Cells are colour coded such that green colours indicate good performance and red colours indicate poor performance.

	DT	KNN	LDA	LG	MLP	NB	RF	SVM
				Train Set (BioFIND)				
Acc	0.77	0.66	0.76	0.78	0.46	0.72	0.88	0.81
bACC	0.77	0.65	0.75	0.77	0.5	0.71	0.87	0.8
AUC	0.8	0.77	0.87	0.86	0.24	0.79	0.95	0.87
RecHC	0.78	0.4	0.68	0.71	1	0.57	0.83	0.74
PreHC	0.74	0.76	0.77	0.78	0.46	0.76	0.89	0.83
F1HC	0.76	0.53	0.72	0.75	0.63	0.65	0.86	0.78
RecPN	0.76	0.89	0.82	0.83	0	0.84	0.92	0.87
PrePN	0.8	0.63	0.75	0.77	0	0.69	0.96	0.79
F1PN	0.78	0.74	0.79	0.8	0	0.76	0.89	0.83
				Test Set (ADNI)				
Acc	0.82	0.66	0.8	0.74	0.69	0.66	0.82	0.78
bACC	0.79	0.72	0.8	0.76	0.5	0.7	0.8	0.79
AUC	0.82	0.78	0.84	0.82	0.28	0.75	0.84	0.86
RecHC	0.86	0.56	0.79	0.7	1	0.6	0.84	0.76
PreHC	0.87	0.91	0.9	0.89	0.69	0.88	0.89	0.9
F1HC	0.87	0.69	0.84	0.79	0.82	0.71	0.86	0.82
RecPN	0.72	0.87	0.81	0.81	0	0.81	0.77	0.81
PrePN	0.69	0.47	0.63	0.55	0	0.47	0.68	0.6
F1PN	0.71	0.61	0.71	0.66	0	0.6	0.72	0.69

Note: Machine learning algorithms are abbreviated as DT: Decision tree; KNN: K-nearest neighbour; LDA: Linear discriminant analysis; LG: Logistic regression; MLP: Multilayer perceptron; NB: Naïve Bayes; RF: Random forest; SVM: Support vector machines; Evaluation metrics are abbreviated as Acc: accuracy; bAcc: balanced accuracy; AUC: Area under the curve; RecHC / PN: Recall of healthy controls / patients; PreHC PN: Precision of healthy controls / patients; F1HC / PN: F1 for healthy controls / patients.

algorithms were trained on BioFIND data and tested on ADNI data. Therefore, any prior selection that includes information from ADNI or information about the relations between BioFIND and ADNI, can potentially bias the results. Following this procedure, features with non-significant r_s values were omitted.

Whereas the brain regions comprising the subcortical parcellation scheme are unique (i.e., are not present in other parcellation schemes), other schemes include different parcellations of the same cortical regions, and therefore partially overlap. To avoid double dipping, we employed a preliminary classification procedure using logistic regression, set to select one cortical scheme to be used in subsequent analyses. Namely, for each parcellation scheme within the BioFIND dataset, we included selected features (i.e., with significant r_s, as described above) as predictors in a logistic regression model (i.e., GLM model with binomial distribution) that classifies BioFIND's cases into PN and HC. We then used these models, trained with BioFIND data, to classify cases in the ADNI dataset. We computed accuracy scores as the proportion of correct classifications across both groups (true positives + true negatives), resulting in the following: Cortical ribbon: 0.63; 4 lobes: 0.66; DKT: 0.67; Destrieux: 0.69; and Subcortical (included for completeness): 0.74. Note that for this preliminary step only accuracy scores were taken into account. However, additional metrics were used for the evaluation of ML algorithms, as described later. Based on these accuracy scores, we decided to use features from the Destrieux parcellation scheme (30 features with significant r_s) and from the subcortical scheme (7 features with significant r_s) in any further ML classification. More information about these features is provided in the Results section below. Nevertheless, because group differences (for averaged FDs) were more robust when the DKT scheme was used, and to avoid potential bias in parcellation selection, the analyses were repeated with 32 features with significant r_s extracted from the DKT-derived dataset (see Supplementary Materials 3).

2.6. Classification algorithms

2.6.1. Training and evaluation

We used eight classification algorithms: Decision tree, K-nearest neighbour, Linear discriminant analysis, Logistic regression, Multilayer perceptron, Naïve Bayes, Random forest, and Support vector machine; described, e.g., in (Bishop, 2006; Pedregosa, 2011) to classify participants into their respective groups (i.e., PN, HC). All algorithms were implemented with scikit-learn v0.24.1 (Buitinck, 2013), aside from Multilayer perceptron algorithm which was implemented with PyTorch v1.8.0 (Paszke, 2019).

We designated BioFIND as the training set, due to the greater number of observations this set contains (307 vs 152 in ADNI), and better balance across groups (a 54% HC / 46% PN split in BioFIND vs. 31% HC / 69% PN split in ADNI). Therefore, each algorithm was initially trained

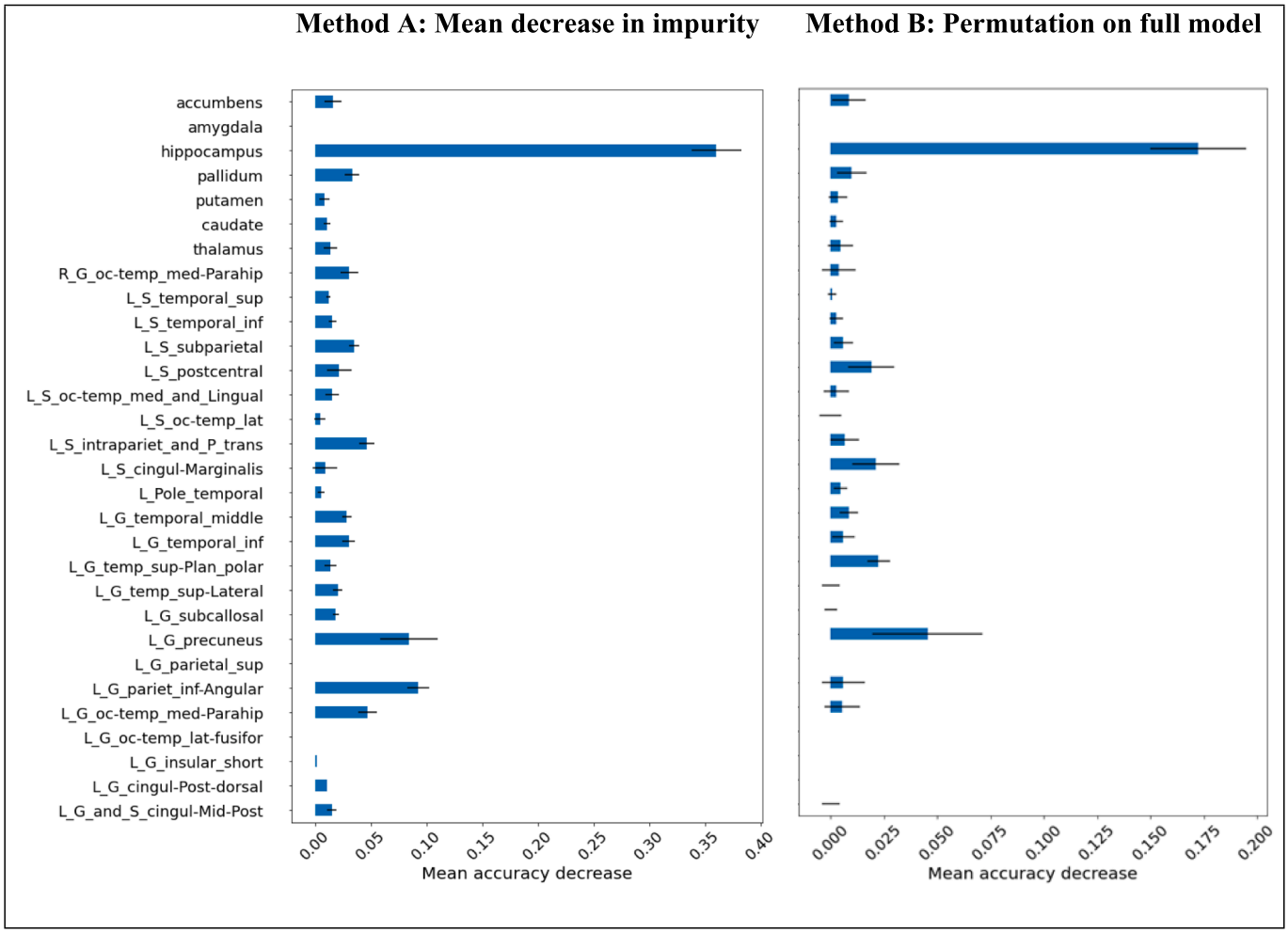


Fig. 6. Results of feature importance evaluation using the impurity decrease (left) and permutation (right) techniques. The blue bars are the feature importances of the forest, along with their inter-trees variability represented by the black error bars.

and tuned on data from BioFIND and then tested on data from ADNI. Accordingly, BioFIND data were used for training and validation in the process of model selection. We employed an exhaustive grid search to adjust the (hyper)parameters, separately for each algorithm, with a set of relevant parameters (see Supplementary Materials 1). To select the best model for each algorithm (model selection), models were evaluated using a stratified 10-fold cross-validation, in which the folds are selected so that the mean response value is approximately equal in all folds. Model scoring was determined based on accuracy, calculated as the proportion of correct classifications across both groups.

To compare how well each algorithm performs, the selected ("best") trained model for each algorithm was then evaluated against the entire training set (BioFIND data) and, more importantly, against the unseen test set (ADNI data). The evaluation was mainly based on accuracy and AUC, that is, the two-dimensional area underneath the receiver operating characteristic (ROC) curve which shows the performance of the classification model at all classification thresholds. Nevertheless, the following additional metrics were also considered: (1) Recall, defined as the number of true positives (TP) divided by the total number of TP and false negatives (FN), i.e., $\frac{TP}{(TP+FN)}$; (2) Precision, defined as the number of TP divided by the total number of TP and false positives (FP), i.e., $\frac{TP}{(TP+FP)}$; (3) F1 score, i.e., the harmonic mean of precision and recall, i.e., $2 \times (\frac{Precision \times Recall}{Precision+Recall})$, which takes both metrics into account, giving them equal weights; and (4) balanced accuracy score, defined as the average of recall obtained on each class. These measures were calculated separately for each group. In addition, confusion matrices, summarising the performance of each algorithm, and ROC curves were plotted and evaluated.

2.6.2. *Further investigation*

As further detailed in the Results section, based on performance, Random Forest (RF) was identified as the most successful algorithm and was therefore investigated further. This algorithm can be used to evaluate the importance of features in a classification task (Murphy, 2022), and so two methods, commonly used for this purpose, were employed here. The first is based on impurity decrease (also referred to as "gini importance"). With this method, importances are computed as the mean and standard deviation of accumulation of the impurity decrease within each tree. However, impurity-based feature importances can be misleading for high cardinality features (that is, features that can contain many unique values, as in the current case). The second common method—permutation feature importance—addresses this issue. With this technique, importance is defined as the decrease in a model score when a single feature value is randomly shuffled. This shuffling breaks the relationship between the feature and the target for the selected feature. Therefore, any drop in the model score is indicative to what extent the model depends on the feature that was shuffled. One advantage of this technique is that it does not depend on specific assumptions (i.e., model agnostic). Additionally, it can be calculated many times with different permutations of the feature thereby allowing more reliable estimates.

2.7. Data and code availability

Raw ADNI data is available via ADNI's project: https://adni.loni.usc.edu/, and raw BioFIND data is available via DPUK's portal: https://www.dementiasplatform.uk/. Preprocessed (derived) data used for group analyses, ML classification, trained models, and all code used throughout the study, are available through: https://github.com/ronitibon/FDAD.

3. Results

3.1. Overall estimates of group differences

t/F-statistics, p-values, and Bonferroni corrected p-values (across all parcellation schemes) are shown in Table 1. Boxplots depicting averaged FDs for each derived dataset are shown in Fig. 2. As expected, an independent sample *t*-test, held separately for each BioFIND's derived dataset, revealed lower averaged FD values for MCI patients vs. HC. Similarly, for ADNI's derived datasets, one-way ANOVA revealed a linear trend whereby averaged FD values for AD patients < MCI patients < HC.

As shown in Fig. 2, each dataset included a few outliers. To confirm that these outliers did not drive the results, we removed them by excluding cases with averaged FD of more than three scaled median absolute deviation (MAD) from the median of each group and repeated the analyses again. The outcomes of these additional analyses (included in Supplementary Materials 2) showed that the removal of these outliers had very little influence and did not change any of the key results. Therefore, these cases were included in all subsequent analyses.

3.2. Feature investigation and extraction

3.2.1. General linear model and structure coefficients

To visualise the contribution of different brain regions to the prediction of class attribution (Fig. 3), we replaced values in a 3D brain volume of an exemplar participant ('Bert') with r_s values obtained from BioFIND data, and overlayed on a structural brainmask MRI image of that participant. This was done for the Destrieux and subcortical parcellation schemes, subsequently used for ML classification (see below).

As shown in Fig. 3, negative r_s values were observed in most brain regions within these selected parcellation schemes, adhering, as expected, to lower FD values for MCI patients compared to HC. Within the Destrieux parcellation scheme, some positive values were also observed (for example, in the right central sulcus, in the right opercular and triangular parts of the inferior frontal gyrus, and in the right cuneus). However, as outlined below, these values were not significant and therefore were not included in subsequent ML classification. Otherwise, negative values were observed throughout the cortex, although the effect was somewhat left lateralised, with stronger (i.e., more negative) values observed in the left hemisphere. Within the subcortical parcellation, all r_s values were negative, and the strongest (i.e., most negative) value was observed in the hippocampi followed by the amygdalae. This scheme was implemented bilaterally, and therefore lateralisation patterns cannot be inferred. Note that due to the arbitrary scale, direct comparisons can only be made between different regions within the same parcellation scheme, and not across schemes.

3.2.2. Correspondence between ADNI and BioFIND

The relations between r_s obtained from BioFIND and ADNI are shown in Fig. 4. As shown in the figure, and confirmed by a Pearson correlation test, the datasets were highly correlated, $r = 0.72$, $p < 0.001$. Nevertheless, one-sample Kolmogorov-Smirnov test ran separately for each dataset, indicated that the data are not normally distributed (for both BioFIND and ADNI: $h = 1$, $p < 0.001$), rendering the use of a Pearson coefficient suboptimal. Therefore, we repeated the analysis using a Spearman coefficient instead. This analysis also revealed a highly significant (though more modest) correlation between the datasets, rho = 0.49, $p < 0.001$. Taken together, these results suggest that similar brain regions can be used to predict class attribution across the two datasets based on FD values.

3.2.3. Feature extraction based on structure coefficients' values

A list of brain regions used for ML classification based on the procedure described in the methods section above, is shown in Table 2. Altogether, we extracted 30 features, consisting of 23 FD values obtained via the Destrieux parcellation scheme, and 7 FD values from the subcortical parcellation scheme. Thus, for each of the two datasets, a matrix with 31 columns was created: 30 features and an additional column for the target variable (class), labelled as either 0 (HC) or 1 (PN). As described above, BioFIND data included 307 cases (142 HC and 165 PN), whereas ADNI data included 152 cases (47 HC and 105 PN). These data, summarised in Fig. 4, were used for ML classification.

3.3. Machine learning classification

As described above, ML algorithms were trained and validated on BioFIND's data, and then tested on ADNI's data, and the results of this procedure are described in the main text below. Additional analyses, reported in the supplementary materials, included the following: First, as mentioned in the Methods section, because group differences were more robust when the DKT scheme (instead of the Destrieux parcellation scheme) was used, and to avoid potential biases due to dependencies in the selection of the parcellation scheme, the analyses were repeated with 32 features extracted from the DKT derived dataset. These results, that highly converged with those obtained with the Destrieux parcellation scheme, are reported in Supplementary Materials 3. Second, models were initially trained on a scaled version of the data. However, when reran on unscaled data, performance was highly similar and in some cases slightly improved. As the overall pattern of results was unchanged, we report the non-scaled analyses in the main text to preserve interpretability of the original feature values, and provide results with scaled data in Supplementary Materials 4.

3.3.1. Model selection

Accuracy was validated across 10 folds, as outlined in the methods section. Cross-validation accuracy for the various algorithms, following a grid search with various combinations of (hyper)parameters (see above), is shown in Supplementary Materials 1. Overall, aside from the MLP algorithm, all algorithms were able to provide some accurate predictions (≥66%). For some algorithms, accuracy rates varied significantly as a function of the specific set of parameters (e.g., for SVM), whereas for others accuracy was more stable across models (e.g., LG). This suggests that some algorithms are more influenced by specific parameter combinations and should therefore be tuned with cautious. The highest accuracies were achieved by LDA, LG, and RF, for which the best models yielded classification accuracy greater than 73%.

3.3.2. Algorithm comparison

For validation of the models against the training set, the best model for each algorithm was fit on the entire training set (i.e., BioFIND data). For validation against the test set, the best model for each algorithm was fit on the unseen ADNI data. Confusion matrices for the validation against the training and testing sets, as well as the respective ROC curves, are shown in Fig. 5. Additional metrics—accuracy, balanced accuracy, AUC, recall, precision, and F1 scores for the control and patient groups—were also calculated and are shown in Table 3.

As can be seen in Table 3, the RF algorithm outperformed all other algorithms (in all metrics) for the train set. This algorithm also performed well on the test set, although in this case, other algorithms (namely, DT, LDA, and SVM) performed comparably well. Therefore, the RF algorithm was explored further in order to gain additional insights into the results.

3.3.3. Further investigation

The evaluation of feature importance using the impurity decrease technique and the permutation technique is shown in Fig. 6. For both techniques, the hippocampus was designated as the most important feature for the model, followed by the precuneus. The inferior parietal lobe (angular gyrus) also showed relatively high values of impurity decrease, but this decrease was not significant across inter-trees variability, and was not replicated for permutation-based importances. Furthermore, the polar plane of the superior temporal gyrus, the marginal branch of the cingulate sulcus, and the postcentral sulcus, were identified as relatively important features based on permutation-based importances, but not based on impurity decrease.

Notably, the two importance measures showed partial divergence. As noted above, Impurity decrease can be biased towards features with higher cardinality or variance and may therefore overestimate their contribution. In contrast, permutation importance is model-agnostic and reflects the impact of each feature on overall predictive performance, and is therefore often preferred. However, in the presence of correlated predictors—as is expected in brain data—permutation importance may underestimate the contribution of individual features if their information is redundant and can be compensated for by other regions. Accordingly, while both methods can provide informative results, regions identified by only one method should be interpreted with caution, whereas regions consistently identified by both methods (i.e., hippocampus and precuneus) represent more robust contributors to model performance.

4. Discussion

The aim of the current study was to establish FD values obtained from T1-wighted MRI images as useful measures for the detection of AD. We initially used FD to identify group differences between MCI/AD patients and healthy controls. Based on previous work (Meregalli, 2022; Ziukelis et al., 2022) we predicted that patients would exhibit reduced structural complexity, as measured by FD values, compared to controls. This hypothesis was confirmed: for both datasets, and for all parcellation schemes, averaged FD values (across all parcels) were significantly lower in patients than in healthy controls. Moreover, in the ADNI datasets, that includes three groups (i.e., HC, MCI, and AD), we observed a linear trend whereby the averaged FD values in HC was greater than in MCI, which was in turn greater than in AD. This pattern verifies MCI as a prodromal stage of AD, during which some brain changes may already be observed, but to a lesser extent.

To investigate whether FD can be used for classification at the individual's level, 8 ML algorithms were trained and evaluated on features extracted from the BioFIND dataset, and then tested on parallel features extracted from ADNI. When evaluated against the training set, most models (7 out of 8) achieved accurate classification, at least to some extent (accuracy range: 0.66–0.88; AUC range: 0.77–0.95). These models were also successful when tested against the unseen test set (ADNI data; accuracy range: 0.66–0.82; AUC range: 0.75–0.86). The only algorithm that was unable to discriminate between cases in the PN and HC groups was MLP, presumably due to the small number of cases relative to the number of features. Amongst successful algorithms, random forest (RF) received the highest scores in all metrics when evaluated against the train set. For this algorithm, accuracy score was 0.88, AUC was 0.95, and scores in all other metrics were > 0.83. This algorithm was also successful (though to a somewhat lesser degree) when evaluated against the unseen test set (accuracy $= 0.82$, AUC $=$ 0.84).

Finally, we evaluated to which extent the results remain valid and generalise beyond specific datasets. Feature-wise correlations performed at the group level indicated that similar brain regions can be used to predict class attribution across the two datasets based on their structural complexity. In addition, as mentioned above, ML classifiers trained on BioFIND data were able to correctly classify cases from the unseen ADNI dataset. Taken together, these results support good correspondence between these two datasets. This is of particular importance given that ADNI data had been widely (if not overly) used in the field, but often with limited generalisation when findings are applied to other datasets (De Carli, 2019; Qiu, 2020). The current results suggest that the measures used in the current study might be more stable across changes in settings, and therefore might be particularly useful for aiding diagnostic procedures in clinical settings.

Interestingly, some algorithms performed better when tested on unseen data than when evaluated against the training data. For example, DT's overall accuracy scores were 0.77 for the training set but 0.82 for the test set (AUC $= 0.8$ for train vs. 0.82 for test). Although at a first glance obtaining better performance when evaluated against unseen data might seem surprising, this pattern is reasonable given the particularities of the current datasets. Namely, the training data (extracted from BioFIND), only included MCI patients in the patient group, whereas the test data (extracted from ADNI) included both MCI and AD patients. The current data support the notion that in some circumstances, AD can be thought of as an exaggerated version of MCI, rendering classification of AD against HC as an easier task. Therefore, the same features can be used to establish a diagnosis, but even more accurately so. Further elaboration of these exact circumstances, as well as full characterization of the utility of our approach for early detection, potentially at stages preceding any cognitive decline, are yet to be pursued, and will greatly benefit from exploitation of longitudinal cohorts (such as those available via ADNI).

Investigation of the contribution of specific brain regions revealed that some regions had greater contribution to accurate classification than others (see Fig. 3 and Table 2), implying two interesting patterns. First, regions with significant contribution were mostly left lateralised. Namely, amongst the cortical regions that were selected for ML classification (based on significant structure coefficients), 22/23 were left-lateralised. Notably, many previous studies (Habib et al., 2003) showed that mnemonic functions in the human brain are often lateralised to the left. This pattern adheres to the cognitive profile that characterises MCI and AD. More specifically, AD is characterised predominantly by a stark decline in episodic memory, that is, the ability to recollect events from one's past (Dening and Sandilyan, 2015). Therefore, the left-lateralisation of regions that contribute to the classification of MCI/AD patients is well aligned with the cognitive profile associated with these conditions, mainly expressed as mnemonic deficits, known to be supported by regions that reside on the left side of the brain.

A second interesting pattern concerns the contribution of specific brain regions. For several decades, regions within the medial temporal lobe—the hippocampus in particular, together with the adjacent areas including the entorhinal, perirhinal, and parahippocampal cortices—had been known as "the memory centre of the brain" and were identified as critical regions for learning and memory (Burgess et al., 2002; Eichenbaum et al., 2007; Pribram, 2012). A large body of research showed that lesions to these regions often result in severe memory deficits (Pribram, 2012; Milner et al., 1998), and that the hippocampus is especially vulnerable to damage at early stages of AD (Mu and Gage, 2011). More recently, research further identified a "core-recollection network", that is, a network of brain regions that is applicable in various episodic memory tasks, regardless the nature of the recollected content. In addition to the hippocampus, this network includes the angular gyrus, the medial prefrontal cortex, retrosplenial/posterior cingulate cortex, precuneus, and middle temporal gyrus (King et al., 2015; Ritchey and Cooper, 2020; Thakral et al., 2017). Importantly, many of these regions (including the hippocampus, parahippocampus, angular gyrus, precuneus, cingulate cortex, and middle temporal gyrus) were identified in the current study as regions in which FD significantly contributes to classification. Moreover, investigation of the features contributing to the RF algorithm designated the hippocampus as the most important feature for the model. These, once again, reinforce the link between the

measures that were used in the current study and the profile of neurocognitive deficits observed in AD/MCI.

Despite its promising results, some important limitations of the current study should be considered. First, although the datasets used here are relatively 'large-scaled' within the domain of medical neuroimaging, in the context of ML the number of cases is highly limited. Even though performance of ML algorithms was overall promising, their accuracy falls far below any threshold that would be viable for aiding decisions in clinical settings. More data is required to improve training and to allow for more complex models, that can pick up additional subtleties in the data when making predictions. Such data are available, for example, via additional ADNI's cohorts (including longitudinal data mentioned above), as well as via additional datasets not considered here (e.g., OASIS (Marcus, 2007)). While these additional explorations go beyond the scope of the current study, we hope that our promising results will motivate further research broadening the utility of this approach, whilst also recognising (and potentially addressing) any limitations. Another limitation concerns some imbalances in demographic data. Specifically, in the ADNI dataset gender was not equally distributed across groups. Given that FD values were also associated with gender, with reduced FD for females compared to males (in BioFIND: $t(306) = 6.56$, $p < 0.001$; in ADNI: $t(150) = 2.54$, $p = 0.01$; calculated for the averaged value across all selected features), it is possible that group differences between patients and controls are confounded by gender. However, this is unlikely to pose a concern for two reasons. First, as reported above (in the 'participants' section), when assessed on the concatenated data (i.e., collapsed across MCI and AD patients so that both groups are contrasted against HC), gender was evenly distributed across patients and controls. These concatenated, demographically balanced data were used for the majority of the analyses. Second, the imbalanced gender distribution in ADNI takes the form of greater proportion of females in the HC group (62%) compared to the MCI group (38%). Given the relative lower FD values in females (vs. male) and in MCI patients (vs HC), any pattern driven by a potential gender confound would attenuate the results rather than exaggerate them. In other words, the current results were obtained *despite* this potential confound rather than *because* of it; a notion that greatly alleviates this concern.

To conclude, we offer novel evidence for the utility of FD—a relatively understudied measure—in identifying group differences associated with MCI/AD. The study also provides promising direction for individual classification based on this measure. Although the translation of this work to clinical settings would require additional steps, and in particular, improving classification accuracy by additional training and tuning of the models, the correspondence that was identified here between two datasets collected in different settings points to generalisability of the current approach. Ultimately, we want to aid early diagnosis of neurodegenerative conditions in the clinical settings in which they are normally assessed. The current study entails a step in this direction.

Funding resources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Roni Tibon: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Christopher R. Madan:** Writing – review & editing, Writing – original draft, Software, Methodology. **Delshad Vaghari:** Writing – review & editing, Writing – original draft, Validation, Methodology. **Constantino Carlos Reyes-Aldasoro:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandc.2026.106443.

Data availability

Raw data available via ADNI's project: https://adni.loni.usc.edu/, and DPUK's portal: https://www.dementiasplatform.uk/. Derived data are available through: https://github.com/ronitibon/FDAD.

References

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.
Borchert, R. J., et al. (2023). Artificial intelligence for diagnostic and prognostic neuroimaging in dementia: A systematic review. *Alzheimer's & Dementia, 19*, 5885–5904.
Bruña, R., et al. (2022). Modified MRI Anonymization (De-Facing) for improved MEG Coregistration. *Bioengineering, 9*, 591.
Bucholc, M., et al. (2023). Artificial intelligence for dementia research methods optimization. *Alzheimer's & Dementia, 19*, 5934–5951.
L. Buitinck et al. API design for machine learning software: Experiences from the scikit-learn project. Preprint at 2013.
Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The Human Hippocampus and Spatial and Episodic memory. *Neuron, 35*, 625–641.
Cabral, C., Morgado, P. M., Campos Costa, D., & Silveira, M. (2015). Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages. *Computers in Biology and Medicine, 58*, 101–109.
Chincarini, A., et al. (2011). Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. *NeuroImage, 58*, 469–480.
Collantoni, E., et al. (2020). Cortical Complexity in Anorexia Nervosa: A Fractal Dimension Analysis. *Journal of Clinical Medicine, 9*, 833.
De Carli, F., et al. (2019). Accuracy and generalization capability of an automatic method for the detection of typical brain hypometabolism in prodromal Alzheimer disease. *European Journal of Nuclear Medicine and Molecular Imaging, 46*, 334–347.
Dening, T., & Sandilyan, M. B. (2015). Dementia: Definitions and types. *Nursing Standard, 29*, 37–42.
Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage, 53*, 1–15.
Di Ieva, A. (Ed.). (2024). *The fractal geometry of the brain* (2nd ed.). Springer. https://doi.org/10.1007/978-3-031-47606-1.
Eichenbaum, H., Yonelinas, A. R., & Ranganath, C. (2007). The Medial Temporal Lobe and Recognition memory. *Annual Review of Neuroscience, 30*, 123–152.
Estarellas, M., Huntley, J., & Bor, D. (2024). Neural markers of reduced arousal and consciousness in mild cognitive impairment. *International Journal of Geriatric Psychiatry, 39*, e6112.
Fischl, B., et al. (2002). Whole Brain Segmentation: Automated labeling of Neuroanatomical Structures in the Human Brain. *Neuron, 33*, 341–355.
Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). 'Mini-mental state'. a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189–198.
Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P., & Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nature Reviews. Neurology, 6*, 67–77.
Frizzell, T. O., et al. (2022). Artificial intelligence in brain MRI analysis of Alzheimer's disease over the past 12 years: A systematic review. *Ageing Research Reviews, 77*, Article 101614.
Gaubert, S. *et al.* Exploring the neuromagnetic signatures of cognitive decline from mild cognitive impairment to Alzheimer's disease dementia. *eBioMedicine* **114**, (2025).
Goedert, M., & Spillantini, M. G. (2006). *A Century of Alzheimer's Disease. Science, 314*, 777–781.
Habib, R., Nyberg, L., & Tulving, E. (2003). Hemispheric asymmetries of memory: The HERA model revisited. *Trends in Cognitive Sciences, 7*, 241–245.
Jack, C. R., Jr., et al. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging, 27*, 685–691.
Jack, C. R., et al. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology, 9*, 119–128.
Javed, E., et al. (2025). A Shift Toward Supercritical Brain Dynamics Predicts Alzheimer's Disease Progression. *The Journal of Neuroscience, 45*.
Jiang, S., Yang, S., Deng, K., Jiang, R., & Xue, Y. (2024). Machine learning models for diagnosing Alzheimer's disease using brain cortical complexity. *Frontiers in Aging Neuroscience, 16*.

King, R. D., et al. (2009). Characterization of Atrophic changes in the Cerebral Cortex using Fractal Dimensional Analysis. *Brain Imaging and Behavior, 3*, 154–166.
King, R. D., Brown, B., Hwang, M., Jeon, T., & George, A. T. (2010). Fractal dimension analysis of the cortical ribbon in mild Alzheimer's disease. *NeuroImage, 53*, 471–479.
King, D. R., de Chastelaine, M., Elward, R. L., Wang, T. H., & Rugg, M. D. (2015). Recollection-Related increases in Functional Connectivity Predict Individual differences in memory Accuracy. *The Journal of Neuroscience, 35*, 1763–1772.
Klein, A., & Tourville, J. (2012). 101 Labeled Brain Images and a Consistent Human Cortical labeling Protocol. *Frontiers in Neuroscience, 6*.
Klöppel, S., et al. (2015). Applying Automated MR-Based Diagnostic Methods to the memory Clinic: A prospective Study. *Journal of Alzheimer's Disease, 47*, 939–954.
Krohn, S., et al. (2026). Fractal analysis of brain shape formation predicts age and genetic similarity in human newborns. *Nature Neuroscience, 29*, 171–185.
Liu, Y., Wang, L., Ning, X., Gao, Y., & Wang, D. (2024). Enhancing early Alzheimer's disease classification accuracy through the fusion of sMRI and rsMEG data: A deep learning approach. *Frontiers in Neuroscience, 18*.
Ma, Z., et al. (2020). Identifying Mild Cognitive Impairment with Random Forest by Integrating Multiple MRI Morphological Metrics. *Journal of Alzheimer's Disease, 73*, 991–1002.
Madan, C. R. (2022). Scan once, Analyse many: Using Large Open-Access Neuroimaging Datasets to Understand the Brain. *Neuroinform, 20*, 109–137.
Madan, C. R., & Kensinger, E. A. (2016). Cortical complexity as a measure of age-related brain atrophy. *NeuroImage, 134*, 617–629.
Madan, C. R., & Kensinger, E. A. (2017). Age-related differences in the structural complexity of subcortical and ventricular structures. *Neurobiology of Aging, 50*, 87–95.
Madan, C. R., & Kensinger, E. A. (2018). Predicting age from cortical structure across the lifespan. *European Journal of Neuroscience, 47*, 399–416.
Mandelbrot, B. (1967). How Long is the Coast of Britain? *Statistical Self-Similarity and Fractional Dimension. Science, 156*, 636–638.
Marcus, D. S., et al. (2007). Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented older adults. *Journal of Cognitive Neuroscience, 19*, 1498–1507.
Martínez-Cañada, P., et al. (2023). Combining aperiodic 1/f slopes and brain simulation: An EEG/MEG proxy marker of excitation/inhibition imbalance in Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 15*, Article e12477.
Marzi, C., et al. (2023). Fractal dimension of the cortical gray matter outweighs other brain MRI features as a predictor of transition to dementia in patients with mild cognitive impairment and leukoaraiosis. *Frontiers in Human Neuroscience, 17*.
McDonough, I. M., & Madan, C. R. (2021). Structural complexity is negatively associated with brain activity: A novel multimodal test of compensation theories of aging. *Neurobiology of Aging, 98*, 185–196.
McKhann, G., et al. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology, 34*, 939.
Meregalli, V., et al. (2022). Cortical complexity estimation using fractal dimension: A systematic review of the literature on clinical and nonclinical samples. *European Journal of Neuroscience, 55*, 1547–1583.
Milner, B., Squire, L. R., & Kandel, E. R. (1998). Cognitive Neuroscience and the Study of memory. *Neuron, 3*, 445–468.
Morris, J. C. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology, 43*, 2412–2414.
Mu, Y., & Gage, F. H. (2011). Adult hippocampal neurogenesis and its role in Alzheimer's disease. *Molecular Neurodegeneration, 6*, 85.
Murphy, K. P. (2022). *Probabilistic Machine Learning: An Introduction*. Cambridge, Massachusetts: MIT Press.
Nicastro, N., et al. (2020). Cortical Complexity analyses and their Cognitive Correlate in Alzheimer's Disease and Frontotemporal Dementia. *Journal of Alzheimer's Disease, 76*, 331–340.
Ni, H., Xue, J., Qin, J., & Zhang, Y. (2024). Accurate identification of individuals with subjective cognitive decline using 3D regional fractal dimensions on structural magnetic resonance imaging. *Computer Methods and Programs in Biomedicine, 254*, Article 108281.
Nugent, A. C., Namyst, A. M., Carver, F. W., Thompson, P. M. & Stout, J. D. Multi-site Harmonization for Magnetoencephalography Spectral Power Data. *Imaging Neuroscience* https://doi.org/10.1162/IMAG.a.1099 (2025) doi:10.1162/IMAG.a.1099.
Pantoni, L., et al. (2019). Fractal dimension of cerebral white matter: A consistent feature for prediction of the cognitive performance in patients with small vessel disease and mild cognitive impairment. *NeuroImage: Clinical, 24*, Article 101990.
Paszke, A., et al. (2019). *PyTorch: An Imperative Style, High-Performance Deep Learning Library. in Advances in Neural Information Processing Systems* (vol. 32). (Curran Associates Inc.
Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. https://hal.inria.fr/hal-00650905.
Petersen, R. C. (2009). Early Diagnosis of Alzheimer's Disease: Is MCI too late? *Current Alzheimer research, 6*, 324.
Pribram, K. (2012). *Biology of memory*. Elsevier.
Qiu, S., et al. (2020). Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. *Brain, 143*, 1920–1933.
Rathore, S., Habes, M., Iftikhar, M. A., Shacklett, A., & Davatzikos, C. (2017). A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *NeuroImage, 155*, 530–548.
Ritchey, M., & Cooper, R. A. (2020). Deconstructing the Posterior Medial Episodic Network. *Trends in Cognitive Sciences, 24*, 451–465.
Ruiz de Miras, J., et al. (2017). Complexity analysis of cortical surface detects changes in future Alzheimer's disease converters. *Human Brain Mapping, 38*, 5905–5918.
Sherry, A., & Henson, R. K. (2005). Conducting and Interpreting Canonical Correlation Analysis in Personality Research: A User-Friendly Primer. *Journal of Personality Assessment, 84*, 37–48.
Thakral, P. P., Wang, T. H., & Rugg, M. D. (2017). Decoding the content of recollection within the core recollection network and beyond. *Cortex, 91*, 101–113.
Thompson, P. M., et al. (1998). Cortical variability and asymmetry in normal aging and Alzheimer's disease. *Cerebral Cortex, 8*, 492–509.
Tibon, R., et al. (2021). Transient neural network dynamics in cognitive ageing. *Neurobiology of Aging, 105*, 217–228.
Tibon, R., & Tsvetanov, K. A. (2022). The "Neural Shift" of sleep Quality and Cognitive Aging: A Resting-State MEG Study of Transient Neural Dynamics. *Frontiers in Aging Neuroscience, 13*, Article 746236.
Tibon, R., Geerligs, L., & Campbell, K. (2022). Bridging the big (data) gap: Levels of control in small- and large-scale cognitive neuroscience research. *Trends in Neurosciences, 45*, 507–516.
Vaghari, D., et al. (2022). A multi-site, multi-participant magnetoencephalography resting-state dataset to study dementia: The BioFIND dataset. *NeuroImage, 258*, Article 119344.
Vaghari, D., Kabir, E., & Henson, R. N. (2022). Late combination shows that MEG adds to MRI in classifying MCI versus controls. *NeuroImage, 252*, Article 119054.
Wechsler, D. (1945). A Standardized memory Scale for Clinical Use. *The Journal of Psychology, 19*, 87–95.
Ziukelis, E. T., Mak, E., Dounavi, M.-E., Su, L., & T O'Brien, J. (2022). Fractal dimension of the brain in neurodegenerative disease and dementia: A systematic review. *Ageing Research Reviews, 79*, Article 101651.