
Article:
De Marco, M., Beltrachini, L., Biancardi, A. et al. (2 more authors) (2017) Machine-learning support to individual diagnosis of mild cognitive impairment using multimodal MRI and cognitive assessments. Alzheimer Disease & Associated Disorders, 31 (4). pp. 278-286. ISSN 0893-0341

https://doi.org/10.1097/WAD.0000000000000208

OPEN

Machine-learning Support to Individual Diagnosis of Mild Cognitive Impairment Using Multimodal MRI and Cognitive Assessments

Matteo De Marco, PhD, Leandro Beltrachini, PhD,†‡§ Alberto Biancardi, PhD,‖ Alejandro F. Frangi, PhD,† and Annalena Venneri, PhD*†*

Background: Understanding whether the cognitive profile of a patient indicates mild cognitive impairment (MCI) or performance levels within normality is often a clinical challenge. The use of resting-state functional magnetic resonance imaging (RS-fMRI) and machine learning may represent valid aids in clinical settings for the identification of MCI patients.

Methods: Machine-learning models were computed to test the classificatory accuracy of cognitive, volumetric [structural magnetic resonance imaging (sMRI)] and blood oxygen level dependent-connectivity (extracted from RS-fMRI) features, in single-modality and mixed classifiers.

Results: The best and most significant classifier was the RS-fMRI +Cognitive mixed classifier (94% accuracy), whereas the worst performing was the sMRI classifier (~80%). The mixed global (sMRI+RS-fMRI+Cognitive) had a slightly lower accuracy (~90%), although not statistically different from the mixed RS-fMRI+Cognitive classifier. The most important cognitive features were indices of declarative memory and semantic processing. The crucial volumetric feature was the hippocampus. The RS-fMRI features selected by the algorithms were heavily based on the connectivity of mediotemporal, left temporal, and other neocortical regions.

Conclusion: Feature selection was profoundly driven by statistical independence. Some features showed no between-group differences, or showed a trend in either direction. This indicates that clinically relevant brain alterations typical of MCI might be subtle and not inferable from group analysis.

Key Words: machine learning, magnetic resonance imaging, semantics, hippocampus, resting-state

(*Alzheimer Dis Assoc Disord* 2017;31:278–286)

Received for publication February 13, 2017; accepted July 3, 2017.

From the Departments of *Neuroscience; †Electronic and Electrical Engineering, Centre for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB); ‖Cardiovascular Science, University of Sheffield, Sheffield; ‡School of Physics and Astronomy; and §School of Psychology, Cardiff University Brain Research Imaging Centre (CUBRIC), Cardiff University, Cardiff, UK.

Supported by grant no. 42/RF-2010-2321718 from the Italian Ministry of Health to A.V. M.D.M. and A.B. were employed through funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 601055, VPH-DARE@IT to A.V. and A.F.F. This study was also partially supported by EPSRC, grant number, EP/M006328/1, OCEAN, to A.F.F. and A.V.

The authors declare no conflicts of interest.

Reprints: Annalena Venneri, PhD, Department of Neuroscience, Medical School, University of Sheffield, Beech Hill Road, Royal Hallamshire Hospital, N floor, Room N130, Sheffield, S10 2RX, UK (e-mail: a.venneri@sheffield.ac.uk).

Mild cognitive impairment (MCI) identifies adults who experience impairment in neuropsychological abilities, while retaining daily-life independence. The range of possible etiologies is heterogenous, with Alzheimer disease (AD) often being a prime suspect.[1] Nonpathologic processes of senescence, however, may also trigger a measurable decline in cognitive functioning,[2] and it is not uncommon that healthy adults complain of their declining cognitive abilities. This conceptual overlap is further complicated by additional factors. First, thresholds of impaired cognitive performance have been operationalized in many ways.[3] Second, variability in the choice of cognitive tests and their procedure of administration generates different diagnostic outputs.[4] Third, cross-cultural differences exist in test performance,[5] but this is rarely acknowledged. Fourth, raw neuropsychological scores may distribute skewly,[6] compromising the validity of the descriptors used to set the threshold of "normality". Fifth, high levels of education may mask the presence of cognitive impairment.[7]

Recently, revised versions of consensus guidelines have incorporated supporting evidence from neuromolecular imaging and cerebrospinal fluid biomarkers, for diagnosing MCI due to AD.[8] Despite the theoretical robustness of this approach, these techniques are not appropriate for characterizing AD burden in asymptomatic adults or patients with nonprogressive/nonpersistent MCI.[9] A more viable contribution is that of structural magnetic resonance imaging (sMRI) and resting-state functional magnetic resonance imaging (RS-fMRI). Both appear useful to describe patients diagnosed with clinically established AD,[10,11] and RS-fMRI in particular is increasingly receiving attention by researchers, as it seems to be sensitive to very early pathologic alterations.[12] Although significant reduction of regional functional connectivity in MCI has been reported in cross-sectional,[13] and longitudinal studies,[14] this evidence is the result of group-level inferential statistics, which is of limited utility for the clinical classification of single individuals. Multivariate and machine-learning techniques offer the opportunity to build data-driven classificatory models which can predict group membership of each participant based on MRI features. A number of recent studies have implemented these classificatory techniques to identify MCI patients using RS-fMRI as a single source of diagnostic information,[15–18] or in combination with sMRI.[19–21]

In this study we used machine-learning methods to carry out classifications of participants with a diagnosis of MCI based on features extracted from cognitive performance, sMRI, and RS-fMRI, with a series of single-type and mixed classifiers. No specific hypothesis was formulated in association with cognitive classifiers as the diagnostic status was heavily dependent on cognitive performance. We hypothesized that RS-fMRI-based classifiers would be superior to the

others (quantitative expectation), and that the selected features would yield important connection with neuropathologic models of abnormal aging (qualitative expectation). A major goal was to understand to what extent and in what way such methodology would be of aid in clinical settings.

METHODS

Participants

In total, 139 inhabitants of the Venetian lagoon, older than 50 years and still independent in their daily activities were considered for inclusion. Candidates were either outpatients referred to neurological examination by their general practitioner because of suspected cognitive decline, or adults willing to take part in research projects because of personal interest and/or subjective cognitive concerns. All underwent a comprehensive medical examination led by an experienced neurologist between May 2011 and November 2014. This was based on the anamnestic information, a neurological screening, a clinical MRI protocol (including diffusion-weighted, T1-weighted, T2-weighted, and Fluid Attenuation Inversion Recovery images) which was inspected by a senior neuroradiologist, and a battery of cognitive tests administered and interpreted by an experienced neuropsychologist. Upon application of exclusion criteria, participants were allocated to 1 of 2 diagnostic categories: healthy adult having no objective cognitive difficulties (control), or patient diagnosed with MCI (patient). Diagnoses of MCI were established by a consensus of opinions among clinicians and clinical follow-ups. Diagnostic exclusion criteria were as follows: a Mini Mental State Examination score <24, ongoing treatments (psychotropic medication, cholinesterase inhibitors, memantine, drugs for research purposes, or with toxic effects to internal organs); a significant disease at clinical level; history of transient ischemic attack; diagnosis of severe vascular pathology; baseline structural MRI revealing different diagnostic patterns from those expected in MCI; presence/diagnosis of uncontrolled seizures; peptic ulcer; cardiovascular disease; neuropathy with conduction difficulties; significant disabilities; proof of abnormal baseline levels of folates, vitamin B12, or thyroid-stimulating hormone. "Technical" exclusion criteria were as follows: >1 missing entry in the database of cognitive scores; presence of relevant signal artefacts or excessive in-scanner motion. On the basis of application of these criteria, 50 controls and 50 patients matched as closely as possible at a group level for age, education levels, and sex ratio were included. Demographic characteristics of the final sample are reported in Table 1. This study was approved by the Institutional Review Board of the IRCCS Fondazione Ospedale San Camillo (Venice, Italy), protocol number 11/09-version 2. Informed consent was obtained from all participants.

MRI and Cognitive Data Acquisition

The MRI protocol (1.5 T Philips Achieva), including structural and functional acquisitions, was completed in a single session. Participants were instructed to keep their eyes closed without falling asleep and remain as still as possible for the full duration of the examination. Turbo-field echo T1-weighted images were acquired with the following characteristics: voxel dimension 1.10×1.10×0.60 mm; Repetition Time 7.4 ms; Echo Time 3.4 ms; Field of View 250 mm; matrix size 256×256×124; flip angle 8 degrees.

TABLE 1. Demographic and Neuropsychological Characteristics (expressed as means and standard deviations in parentheses) of the Sample

Variables	Healthy	MCI	Group Difference	
Demographic Factor			$P\ U_{\text{Mann Whitney}}/\chi^2$	
Age (years)	69.54 (5.88)	73.86 (6.31)	<0.001	
Education (years)	10.94 (4.60)	10.70 (4.33)	0.840	
Sex (F/M)	31/19	25/25	0.227	
Neuropsychological Test			$P\ U_{\text{Mann Whitney}}$	$P\ F_{\text{Corrected}}$
Mini Mental State Examination	28.98 (1.32)	27.46 (1.92)	<0.001	<0.001
Raven Progressive Matrices	30.14 (4.62)	27.34 (5.77)	0.015	0.029
Digit Cancellation Test	53.52 (5.27)	48.16 (7.91)	<0.001	0.001
Stroop Test—time interference	23.70 (8.99)	35.77 (18.48)	<0.001	<0.001
Stroop Test—error interference	0.97 (2.81)	3.02 (5.79)	0.002	0.126
Letter Fluency Test	34.74 (12.81)	31.34 (11.08)	0.145	0.234
Category Fluency Test	41.36 (9.92)	30.18 (8.66)	<0.001	<0.001
Token Test	34.50 (1.79)	34.18 (1.90)	0.305	0.256
Similarities Test	20.80 (5.02)	19.78 (4.37)	0.175	0.467
Confrontational Naming Test	19.17 (1.46)	18.48 (1.72)	0.019	0.044
Digit Span Test—forward	6.08 (0.92)	5.74 (0.90)	0.026	0.026
Digit Span Test—backwards	4.30 (0.95)	3.72 (0.81)	0.002	0.001
Paired Associates Learning Test	13.36 (4.02)	9.54 (3.77)	<0.001	<0.001
Prose Memory Test—immediate recall	9.88 (3.68)	6.72 (3.64)	<0.001	<0.001
Prose Memory Test—delayed recall	13.10 (4.71)	7.32 (4.48)	<0.001	<0.001
Corsi Block Tapping Test	4.82 (0.87)	4.22 (0.79)	0.001	0.002
Visual Supraspan Test	20.70 (6.60)	13.20 (8.33)	<0.001	<0.001
Rey-Osterrieth Figure—copy	32.47 (3.65)	29.55 (6.18)	0.008	0.028
Rey-Osterrieth Figure—recall	15.98 (5.66)	8.45 (4.59)	<0.001	<0.001

Between-group differences in cognitive performance were analyzed both with Mann-Whitney tests as well as ANOVAs, correcting for age and years of education. A Bonferroni-corrected P threshold equal to 0.002 was adopted as the appropriate significance level. There were only 3 missing data points: 2 participants missing their Token Test score (1 control and 1 patient) and 1 participant (patient) missing their Paired Associates Learning Test score.

ANOVA indicates analysis of variance; F, female; M, male; MCI, mild cognitive impairment.

Sex is presented in units.

Echo-planar T2*-weighted volumes were instead registered at rest with the following settings: voxel dimensions 3.28×3.28×6.00 mm; Repetition Time 2 s; Echo Time 50 ms; Field of View 230 mm; flip angle 90 degrees. Two 120-volume runs were obtained, preceded by 20 s of dummy scans, set to allow the scanner to reach electromagnetic equilibrium.

A neuropsychological battery was designed for clinical purposes, with particular focus on those domains which are most sensitive to aging and early-stage neurodegeneration (Fig. 1).

MRI Data Preprocessing

T1-weighted images were processed with the FreeSurfer Image Analysis Suite (http://surfer.nmr.mgh.harvard.edu/) following standard segmentation and parcellation procedures. Morphologic indices were extracted from cortical and subcortical structures. RS-fMRI images were preprocessed using the Statistical Parametric Mapping 8 (Wellcome Trust Centre for Neuroimaging, London, UK) CONN toolbox,[22] in a Matlab R2012a environment (Mathworks Inc., UK). Images were realigned to estimate head-motion vectors, slice-timed to correct for intravolume temporal phasing-out, coregistered with their T1-weighted image, normalized with the echo planar imaging template, smoothed with a 6 mm full-width at half-maximum gaussian filter to minimize noise and residual anatomic discrepancies, partialized of the confounding signal coming from the top 5 orthogonal components estimated from the maps of white matter and cerebrospinal fluid (aCompCor procedure),[23] and band-pass filtered (0.008 to 0.09 Hz).

Feature Definition

A large number of candidate indices were defined from demographic/clinical variables and neurostructural/neurofunctional maps (Fig. 1). Basic demographic information and raw cognitive scores (extracted from clinical neuropsychological tests) were included in this list. Neuroanatomic volumetric indices were extracted from the segmentation and parcellation output. ROI-to-ROI (R2R) indices of functional connectivity were computed from RS-fMRI runs as part of the CONN pipelines. These ROIs were defined based on the anatomically automatic labeled atlas.[24] R2R indices identified aspects of connectivity among pairs of anatomically automatic labeled ROIs. To minimize potential selection bias, and in parallel optimize number of regions, the cerebellum was excluded from the model, as it is characterized by low presence of AD pathology,[25] and is usually considered a reference region in positron emission tomography-based studies. Primary sensorimotor areas were also excluded due to their prolonged preservation in AD.[26] Orbitofrontal and temporopolar regions subjected to signal dropout were excluded too to avoid miscalculations. In total, 2122 indices were extracted: demographics: 3, cognition: 19, sMRI: 84, RS-fMRI: 2016.

Feature Selection

Two machine-learning algorithms were considered. These were the linear and quadratic Fisher discriminant analyses (LDA and QDA, respectively),[27] based on their proneness to being applicable to multiple research contexts, including small-sample scenarios.[28,29] Both classifications were modeled for each set of features. To pursue maximized classificatory accuracy, the classifier with higher accuracy was chosen each time. A feature-selection analysis was then run by testing the performance of the

A

Age	Years of Education	Gender

B

Mini Mental State Examination	*General cognitive screening*
Raven Progressive Matrices	*Visuospatial abstract reasoning*
Digit Cancellation Test	*Visuospatial exploration & speed of processing*
Stroop Test - time interference	*Inhibitory skills*
Stroop Test – error interference	*Inhibitory skills*
Letter Fluency Test	*Phonologically-cued lexical memory access*
Category Fluency Test	*Semantically-cued lexical memory access*
Token Test	*Verbal comprehension*
Similarities Test	*Verbal abstract reasoning*
Confrontational Naming Test	*Lexical memory access*
Digit Span Test - forward	*Verbal short-term memory*
Digit Span Test - backwards	*Working memory*
Paired Associates Learning Test	*Verbal learning*
Prose Memory Test – Immediate	*Verbal long-term memory*
Prose Memory Test – Delayed	*Verbal long-term memory*
Corsi Block Tapping Test	*Visuospatial short-term memory*
Visual Supraspan Test	*Visuospatial learning*
Rey-Osterrieth Figure – Copy	*Visuoconstructive skills*
Rey-Osterrieth Figure – Recall	*Visuospatial long-term memory*

C

FRONTAL	*PARIETAL*
Orbital Gyrus	Superior Parietal Lobule
Rectal Gyrus	Angular Gyrus
Frontomarginal Gyrus	Supramarginal Gyrus
Transverse Frontopolar Gyrus	Precuneus
Superior Frontal Gyrus	
Middle Frontal Gyrus	*OCCIPITAL*
Inferior Frontal Gyrus - Pars Opercularis	Superior Occipital Gyrus
Inferior Frontal Gyrus - Pars Triangularis	Middle Occipital Gyrus
Inferior Frontal Gyrus - Pars Orbitalis	Inferior Occipital Gyrus
Subcallosal Gyrus	Cuneus
Paracentral Lobule	Lingual Gyrus
TEMPORAL	*LIMBIC*
Long Insular Gyrus	Anterior Cingulate Cortex
Short Insular Gyri	Middle-Anterior Cingulate Cortex
Superior Temporal Gyrus - Lateral	Middle-Posterior Cingulate Cortex
Superior Temporal Gyrus - Planum Polare	Posterior-Dorsal Cingulate Cortex
Superior Temporal Gyrus - Planum Temporale	Posterior-Ventral Cingulate Cortex
Middle Temporal Gyrus	Parahippocampal Gyrus
Inferior Temporal Gyrus	*SUBCORTICAL*
Temporal Pole	Thalamus
Subcentral Gyrus	Caudate
Fusiform Gyrus	Putamen
	Globus Pallidus
	Hippocampus
	Amygdala

D

FRONTAL	*OCCIPITAL*
Medial Superior Frontal Gyrus	Superior Occipital Gyrus
Superior Frontal Gyrus	Middle Occipital Gyrus
Middle Frontal Gyrus	Inferior Occipital Gyrus
Inferior Frontal Gyrus - Pars Opercularis	Cuneus
Inferior Frontal Gyrus - Pars Triangularis	Lingual Gyrus
Supplementary Motor Cortex	
Paracentral Lobule	*LIMBIC*
TEMPORAL	Anterior Cingulate Cortex
Insula	Middle Cingulate Cortex
Superior Temporal Gyrus	Posterior Cingulate Cortex
Middle Temporal Gyrus	Parahippocampal Gyrus
Inferior Temporal Gyrus	
Fusiform Gyrus	*SUBCORTICAL*
	Thalamus
PARIETAL	Caudate
Superior Parietal Gyrus	Putamen
Inferior Parietal Gyrus	Globus Pallidus
Supramarginal Gyrus	Hippocampus
Angular Gyrus	Amygdala
Precuneus	

FIGURE 1. List of features and regions included in the study. Demographic features were included in the feature-selection process of all classifiers (A). Each cognitive test is listed together with the cognitive domain it relies on (B). Volumetric features did not include the cerebellum or nonassociative areas but did include regions normally subjected to artefacts during blood oxygen level dependent acquisitions (C). The 64 neocortical patches from which the blood oxygen level dependent signal was extracted were processed to calculate the 2016 resulting patterns of statistical association (D). Volumes and hemodynamic signal were extracted separately for each hemisphere. The exclusion of primary motor, primary sensory, and cerebellar areas allowed the feature-selection procedure to focus on the regions of the brain that are affected by Alzheimer pathology during the preclinical and prodromal stage of the disease, and during the phases of mild and moderate dementia. The regions retained by this methodological choice are involved in high-order processes of cognitive and behavioral function.

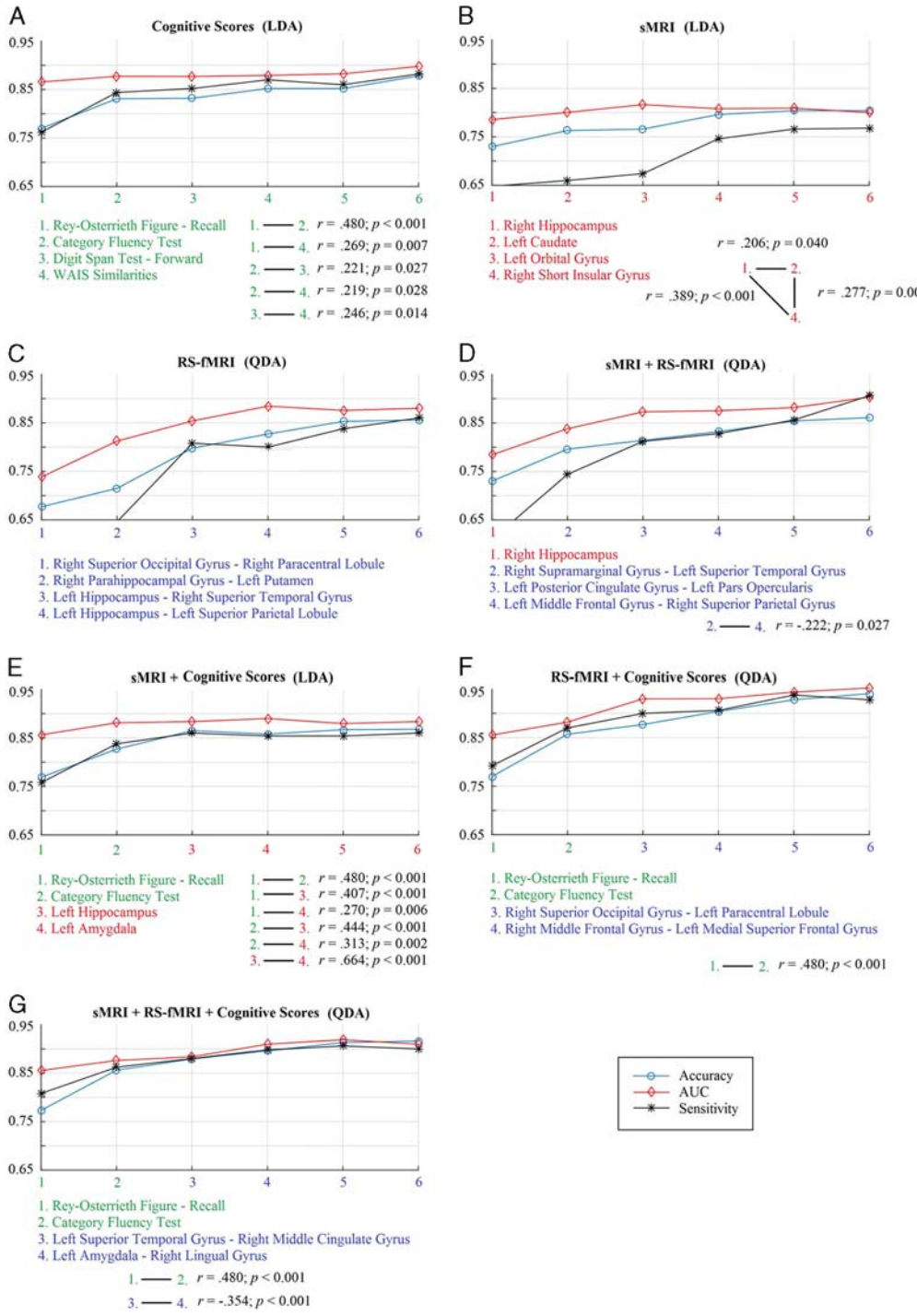

FIGURE 2. The Cognitive (A), sMRI (B), RS-fMRI (C), sMRI+RS-fMRI (D), sMRI+Cognitive (E), RS-fMRI+Cognitive (F), and global sMRI+RS-fMRI+Cognitive (G) classifiers. Accuracy levels are depicted together with measures of sensitivity and area under the curve (AUC). Cognitive, volumetric, and R2R features are indicated in green, red, and blue, respectively (please refer to the online version for color guidance). As the amount of classificatory accuracy decreases with the serial order of the index within the classifier, only the first 4 indices were examined in depth. Correlations among features are indicated below each classifier. AUC indicates area under the receiving-operator curve; LDA, linear Fisher discriminant analyses; QDA, quadratic Fisher discriminant analyze; RS-fMRI, resting-state functional magnetic resonance imaging; R2R, ROI-to-ROI; sMRI, structural magnetic resonance imaging. full color online

chosen classifier as a function of groups of indices. This was achieved via a cost function.[27] The complete data set was subdivided into training and testing subsets using a 10-fold Montecarlo cross-validation. The performance of each classifier was finally evaluated by computing accuracy, area under the receiver-operating-characteristic curve, and sensitivity.

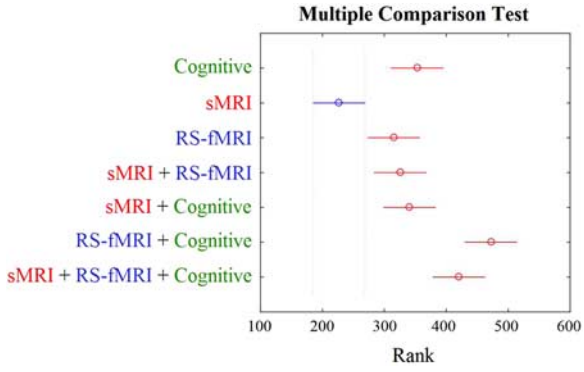

FIGURE 3. Between-classifier Kruskal-Wallis post hoc comparisons. RS-fMRI, resting-state functional magnetic resonance imaging; sMRI indicates structural magnetic resonance imaging (please refer to the online version for color guidance). full color online

Seven classifiers were tested: 3 basic "single-modality" (a, Cognitive; b, sMRI; c, RS-fMRI) and 4 "multiple-modality" classifiers (d, sMRI+RS-fMRI; e, sMRI+Cognitive; f, RS-fMRI +Cognitive; g, sMRI+RS-fMRI+Cognitive). Demographic features were included in all classificatory models. Bonferroni-corrected, post hoc Kruskal-Wallis statistics tested interclassifier differences in accuracy.[29]

RESULTS

The Cognitive classifier (Fig. 2A; LDA) was driven by a test of declarative memory (Rey-Osterrieth Figure—Delayed Recall), and a measure of semantic processing (Category Fluency test). These 2 were responsible for a classificatory accuracy of about 83%. Further tests improved the accuracy rate by an additional 5%. The first volumetric feature selected by the sMRI classifier (Fig. 2B; LDA) was the right hippocampus, followed by the left caudate and the left orbital gyrus. These 3 features approached a 77% accuracy, reaching 80% with additional indices. The RS-fMRI classifier (Fig. 2C; QDA) overstepped an 85% accuracy plateau after 5 indices. These were patterns of R2R connectivity widespread across various regions of the brain, but heavily hinging upon mediotemporal regions (3 of 5 indices). The mixed sMRI+RS-fMRI classifier (Fig. 2D; QDA) obtained performance levels equal to 85% accuracy after 5 indices. The volume of the right hippocampus was selected as the most accurate, followed by R2R connectivity of various associative (prefrontal, parietal, and temporal) cortices. As with the cognitive classifier, the remaining 3 mixed classifiers were reliant on declarative memory and semantic processing as the 2 leading features. In the sMRI+Cognitive classifier (Fig. 2E; LDA) these 2 indices reached an 83% accuracy, marginally improved by volumetric properties of the left mediotemporal complex. In the RS-fMRI+Cognitive classifier (Fig. 2F, QDA), and in the global sMRI+RS-fMRI+Cognitive classifier (Fig. 2G, QDA) the accuracy of the 2 tests reached an accuracy of over 85%, further improved by additional R2R indices. In the global classifier the accuracy was raised to 90% with the addition of indices characterizing left temporal connectivity. Conversely, in the RS-fMRI +Cognitive classifier the accuracy was further enhanced up to 94% with the contribution of 2 indices of widespread connectivity.

The comparison between classifiers revealed that the RS-fMRI+Cognitive classifier was, by far, the most accurate ensemble, accounting for a significantly more accurate classification than 5 of the other classifiers. Vice versa, the sMRI classifier was the least accurate, performing significantly worse than any other classifier (Fig. 3).

For each classifier, the performance of the less accurate classification methods (LDA or QDA) was associated with 2% to 3% less accuracy than the rates of those described above. Nonetheless, these were reliant on comparable sets of features (the recall of Rey-Osterrieth Complex Figure and Category Fluency, the volume of the right hippocampus, and the connectivity of mediotemporal regions).

Selected post hoc analyses were run to understand the clinical importance of these features driven by and in support of possible interpretational frameworks (Fig. 4). Within the cognitive classifier a significant difference existed

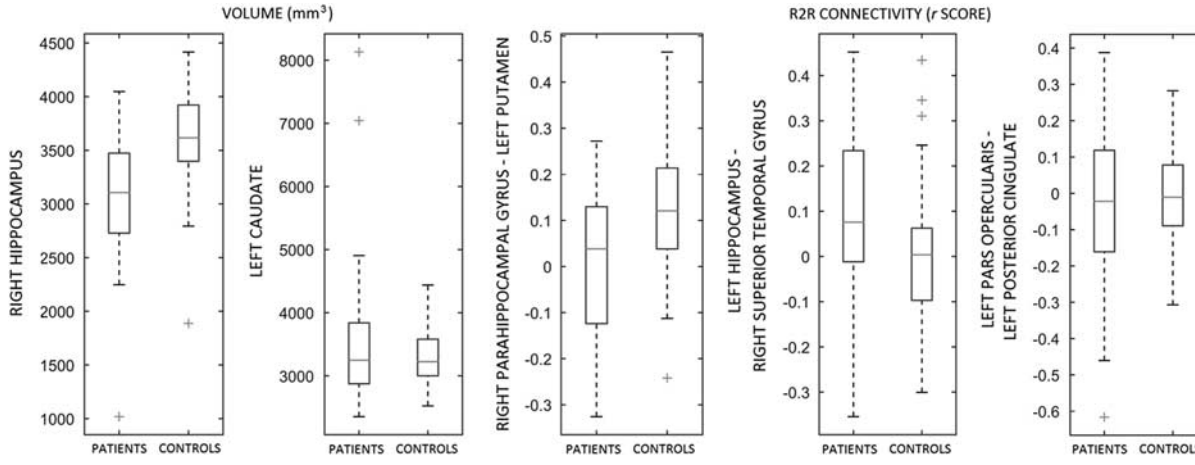

FIGURE 4. Selected post hoc analyses. Between-group comparisons (*t* test statistics) were run to explore the group-level differences of the main features included in the classifiers. MCI patients had significantly larger volumes in the right hippocampus but no difference in the left caudate nucleus. Moreover, patterns of connectivity showed a trend in either direction and, as exemplified by the association between the left posterior cingulate and the left pars opercularis, did only differ in the pattern of dispersion. MCI indicates mild cognitive impairment; R2R, ROI-to-ROI.

between the 2 diagnostic groups on the delayed recall scores on the Rey-Osterrieth Figure and on the Category Fluency Test (both $P < 0.001$). No significant difference was present; however, between the 2 groups on the 2 subsequent tests (Digit Span—forward and the similarities subtest of the Wechsler Adult Intelligence Scale). Moreover, a significant correlation was found between the delayed recall scores on the Rey-Osterrieth Figure and the delayed recall scores on the Prose Memory test (partial correlation correcting for age and education levels, $P = 0.000006$), and between the delayed recall scores on the Rey-Osterrieth Figure and the volume of the right hippocampus ($P = 0.00013$).

Within the structural classifier, a significant difference was found between groups solely for the volume of the right hippocampus, and the volumes of the 2 hippocampi and caudate nuclei were highly correlated ($P = 4.90e\text{-}30$ and 4.20e-52, respectively).

As most of the RS-fMRI R2R indices featured the connectivity of mediotemporal areas, their between-group directionality was explored. The association tended to be larger in controls for some of the features (eg, the second feature: right parahippocampal gyrus—left putamen R2R connectivity), and larger in patients for others (eg, the third feature: left hippocampus—right superior temporal gyrus R2R connectivity).

In the mixed sMRI+RS-fMRI classifier the first 2 R2R features were explored further. A 0.2 z difference from controls was seen for the first R2R (temporal-parietal) feature in patients. In contrast, the correlation between the posterior cingulate cortex and the left pars opercularis was close to 0 in both groups, but showed a larger dispersion in the patient group.

Finally, Pearson correlations were run to explore the association among the top features, within each classifier. Variable results were found, with cognitive and sMRI features showing significant correlations, and RS-fMRI indices tending instead to be statistically independent from each other (Fig. 2).

DISCUSSION

AD triggers a large number of alterations to brain structure, brain connectivity, and cognitive function. Partly, this is the result of a global process of decline which, homogenously, affects a large number of regions, circuital pathways, and cognitive domains (ie, global atrophy and ventricular enlargement, global loss of network connectivity and regional isolation, global cognitive decline). What looks like a general trend, however, can be broken down into separate processes. In AD, studies have highlighted that disease progression involves a number of separate routes. For instance, loss of posteromedial metabolism and atrophy in the mediotemporal complex seem to be driven by distinct mechanisms.[30] Similarly, changes in patterns of connectivity in crucial network pathways are governed by disease-specific, compensatory, and maladaptive mechanisms, which can induce decreases but also increases in the resulting phenotype.[31,32] The extrapolation of independent disease mechanisms can be helpful in clinical settings. For example, there are studies which highlight the importance of exploring mechanisms of both declarative and semantic memory for an early diagnosis of AD, as semantic processing is severely down-regulated in AD, but not significantly disrupted by the normal processes of aging.[33] On this note, the use of machine-learning algorithms for classification purposes is an excellent approach to clarify the diagnostic importance of features extracted from structural and functional neuroimaging. As commented below, however, the particularity of this approach lies in the elimination of any redundancy expressed by features significantly correlated with one another. The resulting combination of variables, therefore, captures distinct aspects of classification, and, thus, of disease.

RS-fMRI Improves Classification

A look at the quantitative aspects of classificatory performance reveals that the sMRI classifier was the least accurate. This indicates that morphometric biomarkers are not as effective as fMRI or cognitive features at detecting abnormalities in the presence of MCI. We argue that, as hippocampal and brain volumes are in fact also influenced by nonpathologic aging,[34] they are unsuitable to provide classificatory specificity.

Classifiers based on cognitive features performed very well. This is necessarily due to the fact that the standard of truth (ie, "patient" or "control") was heavily based on the presence of cognitive impairment measured with cognitive tests.

The most accurate classifications were obtained when RS-fMRI features were included in the feature-selection process. The performance of the RS-fMRI classifier did in fact not differ from that of the Cognitive classifier. In addition, RS-fMRI features improved classification of both sMRI and cognitive features. One possible reason behind such good performance may be the large number (2016) of available RS-fMRI features. This should be seen as an advantage enabled by RS-fMRI modalities (rather than a methodological imbalance), as RS-fMRI offers the opportunity of exploring properties of the blood oxygen level dependent signal which are not absolute (ie, related only to a specific voxel or ROI), but relative (ie, reflective of the relationship between 2 voxels or ROIs). These dynamic characteristics are profoundly associated with the basic processes of brain functioning, as task performance is supported by the interactive coactivation/codeactivation of multiple structures.

Each Classifier as Informant of Distinct Mechanisms

A closer, qualitative look at each classifier allows the clarification of: (1) how useful machine-learning algorithms are to extract classificatory information, and (2) how this method helps the understanding of the various types of mechanisms which may separate patients from controls.

As for the Cognitive classifier, the first feature was a measure of declarative memory (the delayed recall of the Rey-Osterrieth Figure), a domain well known to be severely affected in AD. Although cognitive assessment featured a second measure of long-term declarative memory (the delayed recall of the Prose Memory test), this variable was not chosen as part of the classifier. We argue that the significant correlation found between the 2 memory tests translates into comparable classificatory accuracies, hence the non-necessity of including both. In contrast, the performance on a measure of semantic processing (the Category Fluency test) accounted for an exclusive and relevant amount of variability. Declining semantic processing is one of the major features of various forms of neurodegeneration, and occurs as a result of compromised circuits sustained by regions that are anatomically distinct from those in support of declarative memory.[35] By relying on the same argument, we speculate that the global classifier did not include both the performance on the delayed recall of the Rey-Osterrieth Figure and the volume of the right hippocampus (the "top"

cognitive and MRI-based features, respectively) because of a conceptual association between the 2 variables.[36]

The sMRI classifier was heavily reliant on the right hippocampus in our sample, whereas the left hippocampus, presumably because of a very high interhemispheric correlation coefficient, was not included. The second volumetric feature was the left caudate nucleus (presumably contributed by both caudate nuclei, given the large intrahemispheric correlation). Although the volume of the right hippocampus was significantly smaller in the group of patient, no significant between-group difference emerged for the left caudate. It is interesting to note how features with no between-group differences may yield classificatory relevance. We argue that there might be structures subjected to minor morphometric changes, which, however, are more distinctively related to cognitive impairment than any more extensive morphometric dysregulation located elsewhere. On this note, studies on human and primate brains show that neuronal and synaptic densities are not homogenous across the entire cortex.[37,38] Small group differences in a region with high cell density or sustaining a crucial function might have profound biological implications. Dopaminergic neurons represent an example of this mechanism in Parkinson disease, as they are a minimal portion of the total number of nervous cells, but they serve paramount purposes. In this respect evidence does shows that the caudate manifests volumetric shrinkage in AD,[39] whereas this does not occur in healthy aging.[34] These findings show that the caudate alterations seen in patients, albeit not reaching statistical significance in any specific direction, seem to be independent from mediotemporal modifications, yet conceptually relevant for the diagnosis of MCI.

The RS-fMRI classifier was profoundly based on the connectivity of the left and right hippocampal formation. The first feature represented the R2R pathway accounting for the single largest portion of variability in our sample. The subsequent 4 features all entailed independent aspects of mediotemporal connectivity. Since the earliest histopathologic descriptions, AD has been described as a disease that causes a computational isolation of the hippocampus.[40] Loss of hippocampal and parahippocampal connectivity would be the in vivo equivalent of this process. In addition, one of the R2R features showed a trend toward the opposite direction, with patients having increased hippocampal-temporal connectivity. In line with the evidence of increased hippocampal metabolism shown during the MCI stage,[32] we hypothesize that up-regulated connectivity in patients may be the result of neuroplastic modifications triggered by the early stages of hippocampal disconnection, and that the RS-fMRI classifier is suitable to capture disease mechanisms as well as neuroplastic responses. These latter would in all likelihood not be recordable by morphometric acquisitions, which reflect instead gross anatomy, well known to be more resistant to neuroplastic alterations.

We then included a mixed sMRI+RS-fMRI classifier to understand whether the sole information extracted from an MRI protocol could be exploited clinically. Hippocampal volumes were confirmed as the most informative feature. Decreased connectivity (a 0.2 average drop in the strength of the correlation coefficient) between temporal and parietal region improved this classification. Interestingly, for the third feature (posterior cingulate to Broca area), the *r* coefficient was close to 0 in both groups (indicating no association). In this case, the 2 groups differed in the dispersion levels, suggesting that the informative aspect for this pathway might be the presence of an association (regardless of the directionality) in a pathway where an association would normally not exist.

The mixed sMRI+Cognitive classifier was constructed based on the combination of features that are usually at disposal of the clinician (a cognitive assessment and an anatomic brain scan). The results are perfectly in line with the typical pattern of clinical features that drives a diagnosis of early-stage neurodegeneration, as the selected features are measures of declarative memory and mediotemporal volumes.

The RS-fMRI+Cognitive classifier was the top-performing one. When the analysis of declarative memory is flanked by measures of connectivity, the classification approaches optimal levels (accuracy ~94%) and outperforms the support provided instead by sMRI. The superior performance of this classifier might reflect the qualitatively different disruption caused by AD neurodegeneration on brain function, leading often to compensatory change in controls and maladaptative alteration in the early stage of neurodegeneration.[31]

Finally, the outcome of the global classifier confirmed that the characterization of cognitive profiles (presence of declarative memory and semantic processing deficits) was by far the most accurate predictive formula for classifying patients. R2R features contributed to improving the accuracy by highlighting the role played by various aspects of the limbic system, and temporooccipital areas.

Limitations

Despite the protection toward bias offered by a data-driven approach and a sample of comparable or larger size than that of other studies,[16–20] the outcome is still the result of feature and algorithm definition. Although we selected "standard" cognitive tests and segmentation/parcellation atlases, and 2 basic machine-learning algorithms, we cannot rule out the possibility that other methodological choices might have yielded slightly different patterns of findings. This, however, would not undermine the core findings and interpretations. Moreover, the sets of cognitive, neuroanatomic, and neurofunctional variables are qualitatively different from one another, for example, in their number, in the presence of a numerical ceiling, or in their directionality (as patients may show either decreased or increased RS-fMRI connectivity, but only an impoverishment of cognition and brain structure, see Table 2 for the most distinctive anatomic and R2R). Inevitably, feature selection will be affected by these different properties. As a consequence, comparisons of classifiers will be meaningful as far as quantitative performance is concerned, but any analysis focusing on confronting different types of features has to be interpreted with caution. Post hoc inter-feature correlations are in line with the presence of such qualitative differences, as, for instance, most cognitive features (fewer in number) were mutually correlated, determining a certain degree of collinearity, whereas RS-fMRI indices (many in number) were unrelated with one another.

Clinical Usefulness of Machine-learning Methods

In conclusion, these findings indicate that RS-fMRI R2R connectivity improves diagnostic classification of patients with MCI, and outperforms the accuracy of sMRI, which was profoundly reliant on the importance of hippocampal volumes. A careful look at each classifier revealed that machine-learning approaches, by circumventing

TABLE 2. Distinctive Neuroanatomic and Neurofunctional Characteristics (expressed as means and standard deviations in parentheses) of the 2 Diagnostic Groups

Neuroimaging Variables	Healthy	MCI	P $F_{Corrected}$
sMRI Index—significant difference: controls > patients			
Left hippocampus	3508.65 (397.22)	3038.34 (490.47)	<0.001
Right hippocampus	3598.68 (425.13)	3092.91 (554.34)	<0.001
Left parahippocampal gyrus	3288.38 (476.96)	2908.46 (589.22)	0.004
Right lingual gyrus	4041.22 (563.28)	3646.56 (555.56)	0.002
RS-fMRI Index—significant difference: controls > patients			
Right middle frontal gyrus—left insula	0.0450 (0.17)	−0.0599 (0.16)	0.002
Right inferior frontal gyrus, pars triangularis—right thalamus	0.1534 (0.18)	0.0413 (0.16)	0.003
Left supplementary motor cortex—left globus pallidus	0.2330 (0.12)	0.1221 (0.14)	<0.001
Left posterior cingulate cortex—left inferior temporal gyrus	0.2131 (0.16)	0.0928 (0.18)	0.002
Left parahippocampal gyrus—right paracentral lobule	0.0681 (0.14)	−0.0522 (0.17)	0.001
Right parahippocampal gyrus—left putamen	0.1159 (0.14)	0.0148 (0.15)	0.004
Left cuneus—left paracentral lobule	0.1043 (0.19)	−0.0123 (0.22)	0.005
Left cuneus—right putamen	0.0226 (0.17)	−0.0722 (0.17)	0.004
Right cuneus—right paracentral lobule	0.1628 (0.18)	0.0217 (0.22)	0.003
Right cuneus—right putamen	0.0741 (0.16)	−0.0495 (0.20)	<0.001
Left occipital superior gyrus—right putamen	0.0254 (0.17)	−0.0874 (0.17)	0.002
Right fusiform gyrus—left supramarginal gyrus	0.0425 (0.17)	−0.0452 (0.17)	0.001
Right fusiform gyrus—left superior temporal gyrus	0.0815 (0.19)	−0.0370 (0.18)	0.004
Right fusiform gyrus—right superior temporal gyrus	0.1791 (0.22)	0.0423 (0.18)	<0.001
Right fusiform gyrus—right inferior frontal gyrus, pars opercularis	−0.0269 (0.16)	−0.1087 (0.17)	0.003
Right fusiform gyrus—right inferior frontal gyrus, pars triangularis	−0.0028 (0.15)	−0.0920 (0.20)	0.005
Left precuneus—left middle temporal gyrus	0.3054 (0.17)	0.1787 (0.20)	0.001
Left precuneus—left inferior temporal Gyrus	0.1434 (0.18)	0.0508 (0.20)	0.002
Right caudate—left middle temporal Gyrus	0.0244 (0.15)	−0.0770 (0.16)	0.004
Left putamen—left thalamus	0.2350 (0.16)	0.1365 (0.17)	0.004
Left putamen—right middle temporal Gyrus	0.0859 (0.15)	0.0001 (0.13)	0.003
RS-fMRI Index—significant difference: patients > controls			
Left middle frontal gyrus—right superior medial frontal gyrus	0.1316 (0.23)	0.2526 (0.21)	0.005
Right middle frontal gyrus—left superior medial frontal gyrus	−0.0267 (0.25)	0.1617 (0.24)	<0.001
Right middle frontal gyrus—right superior medial frontal gyrus	0.2337 (0.27)	0.4066 (0.22)	0.001
Left inferior temporal gyrus—right middle frontal gyrus	0.0029 (0.19)	0.1301 (0.21)	0.001
Left inferior temporal gyrus—right inferior frontal gyrus, pars triangularis	0.0064 (0.19)	0.1210 (0.21)	0.004
Left hippocampus—left superior temporal gyrus	0.0362 (0.17)	0.1791 (0.16)	<0.001

Between-group differences in sMRI and RS-fMRI indices were analyzed with ANOVAs, correcting for age. As a Bonferroni correction was judged too conservative for such a large number of statistical comparisons (n = 2100), a still relatively strict $P = 0.005$ was used. Of the entire set of indices, only 4 sMRI and 27 RS-fMRI indices survived this threshold and were reported in the table.

ANOVA indicates analysis of variance; MCI, mild cognitive impairment; RS-fMRI, resting-state functional magnetic resonance imaging; sMRI, structural magnetic resonance imaging.

feature-to-feature statistical redundancy, generate classifiers in which each feature accounts for an independent portion of classificatory accuracy, presumably in reflection of separate disease mechanisms. These might manifest as decrease/increase in R2R correlation (and these differences are often very small and not significant), or in the presence of a correlation between 2 otherwise uncorrelated areas. In addition, between-group volumetric differences do not seem to scale to a common denominator, as minimal differences in specific structures might be more relevant than larger differences elsewhere. These alterations might represent an important source of clinical information and have to be further explored to be implemented in neurological settings. The nature of these findings suggest that clinically relevant alterations seen in brain function of MCI patients might be quite subtle and not potentially inferable from group-based analyses.

ACKNOWLEDGMENTS

The authors thank Francesca Meneghello, Cristina Pilosio, Jessica Rigon, and all personnel in the MRI unit of the IRCCS Fondazione Ospedale San Camillo in Venice, Italy, for their various contribution to patient assessment and imaging acquisitions.

REFERENCES

1. Fischer P, Jungwirth S, Zehetmayer S, et al. Conversion from subtypes of mild cognitive impairment to Alzheimer dementia. *Neurology*. 2007;68:288–291.
2. Hanninen T, Koivisto K, Reinikainen KJ, et al. Prevalence of ageing-associated cognitive decline in an elderly population. *Age Ageing*. 1996;25:201–205.
3. Schinka JA, Loewenstein DA, Raj A, et al. Defining mild cognitive impairment: impact of varying decision criteria on neuropsychological diagnostic frequencies and correlates. *Am J Geriatr Psychiatry*. 2010;18:684–691.
4. Bondi MW, Edmonds EC, Jak AJ, et al. Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *J Alzheimers Dis*. 2014;42:275–289.
5. Helms-Lorenz M, Van De Vijver FJR, Poortinga YH. Cross-cultural differences in cognitive performance and Spearman's hypothesis: g or c? *Intelligence*. 2003;31:9–29.
6. O'Connell ME, Tuokko H, Kadlec H. Demographic corrections appear to compromise classification accuracy for severely

skewed cognitive tests. *J Clin Exp Neuropsychol*. 2011;33: 422–431.
7. Lenehan ME, Summers MJ, Saunders NL, et al. Relationship between education and age-related cognitive decline: a review of recent research. *Psychogeriatrics*. 2015;15:154–162.
8. Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7:270–279.
9. Johnson KA, Minoshima S, Bohnen NI, et al. Appropriate use criteria for amyloid PET: a report of the Amyloid Imaging Task Force, the Society of Nuclear Medicine and Molecular Imaging, and the Alzheimer's Association. *Alzheimers Dement*. 2013;9:e1–e16.
10. Ferreira LK, Diniz BS, Forlenza OV, et al. Neurostructural predictors of Alzheimer's disease: a meta-analysis of VBM studies. *Neurobiol Aging*. 2011;32:1733–1741.
11. Prvulovic D, Bokde ALW, Faltraco F, et al. Functional magnetic resonance imaging as a dynamic candidate biomarker for Alzheimer's disease. *Prog Neurobiol*. 2011;95:557–569.
12. Sheline YI, Raichle ME. Resting state functional connectivity in preclinical Alzheimer's disease. *Biol Psychiatry*. 2013;74:340–347.
13. Agosta F, Pievani M, Geroldi C, et al. Resting state fMRI in Alzheimer's disease: beyond the default mode network. *Neurobiol Aging*. 2012;33:1564–1578.
14. Bai F, Watson DR, Shi Y, et al. Specifically progressive deficits of brain functional marker in amnestic type mild cognitive impairment. *PLoS One*. 2011;6:e24271.
15. Khazaee A, Ebrahimzadeh A, Babajani-Feremi A. Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease. *Brain Imaging Behav*. 2016;10:799–817.
16. Jie B, Shen D, Zhang D. Brain connectivity hyper-network for MCI classification. *Med Image Comput Comput Assist Interv*. 2014;17:724–732.
17. Challis E, Hurley P, Serra L, et al. Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI. *Neuroimage*. 2015;112:232–243.
18. Jie B, Zhang D, Wee CY, et al. Topological graph kernel on multiple thresholded functional connectivity networks for mild cognitive impairment classification. *Hum Brain Mapp*. 2014;35:2876–2897.
19. Wee CY, Yap PT, Zhang D, et al. Identification of MCI individuals using structural and functional connectivity networks. *Neuroimage*. 2012;59:2045–2056.
20. Kim JH, Lee JH. Integration of structural and functional magnetic resonance imaging improves mild cognitive impairment detection. *Magn Reson Imaging*. 2013;31:718–732.
21. Beltrachini L, De Marco M, Taylor ZA, et al. Integration of cognitive tests and resting state fMRI for the individual identification of mild cognitive impairment. *Curr Alzheimer Res*. 2015;12:592–603.
22. Whitfield-Gabrieli S, Nieto-Castanon A. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect*. 2012;2:125–141.
23. Behzadi Y, Restom K, Liau J, et al. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage*. 2007;37:90–101.
24. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*. 2002;15:273–289.
25. Klunk WE, Nordberg A, Engler H, et al. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Ann Neurol*. 2004;55:306–319.
26. Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol*. 1991;82:239–259.
27. Okun O. *Feature Selection and Ensemble Methods for Bioinformatics: Algorithmic Classification and Implementations*. Hershey, PA: Medical Information Science Reference; 1987.
28. Wolpert DH. The lack of a priori distinctions between learning algorithms. *Neural Comput*. 1996;8:1341–1390.
29. Braga-Neto U. Fads and fallacies in the name of small-sample microarray classification—a highlight of misunderstanding and erroneous usage in the applications of genomic signal processing. *IEEE Signal Process*. 2007;24:91–99.
30. Villain N, Fouquet M, Baron JC, et al. Sequential relationships between grey matter and white matter atrophy and brain metabolic abnormalities in early Alzheimer's disease. *Brain*. 2010;133:3301–3314.
31. Gardini S, Venneri A, Sambataro F, et al. Increased functional connectivity in the default mode network in mild cognitive impairment: a maladaptive compensatory mechanism associated with poor semantic memory performance. *J Alzheimers Dis*. 2015;45:457–470.
32. Putcha D, Brickhouse M, O'Keefe K, et al. Hippocampal hyperactivation associated with cortical thinning in Alzheimer's disease signature regions in non-demented elderly adults. *J Neurosci*. 2011;31:17680–17688.
33. Venneri A, Mitolo M, De Marco M. Paradigm shift: semantic memory decline as a biomarker of preclinical Alzheimer's disease. *Biomark Med*. 2016;10:5–8.
34. Fjell AM, Walhovd KB, Fennema-Notestine C, et al. One-year brain atrophy evident in healthy aging. *J Neurosci*. 2009;29: 15223–15231.
35. Barbeau EJ, Didic M, Joubert S, et al. Extent and neural basis of semantic memory impairment in mild cognitive impairment. *J Alzheimers Dis*. 2012;28:823–837.
36. Persson J, Söderlund H. Hippocampal hemispheric and long-axis differentiation of stimulus content during episodic memory encoding and retrieval: an activation likelihood estimation meta-analysis. *Hippocampus*. 2015;25:1614–1631.
37. Collins CE, Airey DC, Young NA, et al. Neuron densities vary across and within cortical areas in primates. *Proc Natl Acad Sci U S A*. 2010;107:15927–15932.
38. Jacobs B, Schall M, Prather M, et al. Regional dendritic and spine variation in human cerebral cortex: a quantitative Golgi study. *Cereb Cortex*. 2001;11:558–571.
39. Jiji S, Smitha KA, Gupta AK, et al. Segmentation and volumetric analysis of the caudate nucleus in Alzheimer's disease. *Eur J Radiol*. 2013;82:1525–1530.
40. Hyman BT, Van Hoesen GW, Damasio AR, et al. Alzheimer's disease: cell-specific pathology isolates the hippocampal formation. *Science*. 1984;225:1168–1170.

