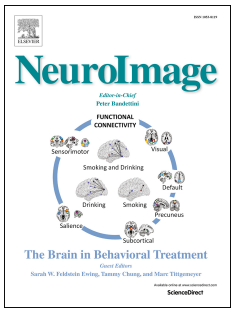

Evaluating the reliability of neurocognitive biomarkers of neurodegenerative diseases across countries: A machine learning approach

M. Belen Bachli, Lucas Sedeño, Jeremi K. Ochab, Olivier Piguet, Fiona Kumfor, Pablo Reyes, Teresa Torralva, María Roca, Juan Felipe Cardona, Cecilia Gonzalez Campo, Eduar Herrera, Andrea Slachevsky, Diana Matallana, Facundo Manes, Adolfo M. García, Agustín Ibáñez, Dante R. Chialvo

PII: S1053-8119(19)31047-X

DOI: https://doi.org/10.1016/j.neuroimage.2019.116456

Reference: YNIMG 116456

To appear in: *NeuroImage*

Received Date: 26 February 2019

Revised Date: 29 October 2019

Accepted Date: 9 December 2019

Please cite this article as: Bachli, M.B., Sedeño, L., Ochab, J.K., Piguet, O., Kumfor, F., Reyes, P., Torralva, T., Roca, Marí., Cardona, J.F., Campo, C.G., Herrera, E., Slachevsky, A., Matallana, D., Manes, F., García, A.M., Ibáñez, Agustí., Chialvo, D.R., Evaluating the reliability of neurocognitive biomarkers of neurodegenerative diseases across countries: A machine learning approach, *NeuroImage* (2020), doi: https://doi.org/10.1016/j.neuroimage.2019.116456.

Evaluating the reliability of neurocognitive biomarkers of neurodegenerative diseases across countries: A machine learning approach

M. Belen Bachli[a,#], Lucas Sedeño[b,c,#,], Jeremi K. Ochab[d,#], Olivier Piguet[e,f], Fiona Kumfor[e,f], Pablo Reyes[g,h], Teresa Torralva[b], María Roca[b], Juan Felipe Cardona[i], Cecilia Gonzalez Campo[b,c], Eduar Herrera[j], Andrea Slachevsky[k,l,m,n,o], Diana Matallana[p], Facundo Manes[b,c,e], Adolfo M. García[b,c,q], Agustín Ibáñez[b,c,e,r,s], and Dante R. Chialvo[a,c]*

[a] *Center for Complex Systems & Brain Sciences (CEMSC[3]), Escuela de Ciencia y Tecnologia (ECyT), Universidad Nacional de San Martín, 25 de Mayo 1169, San Martín, (1650), Buenos Aires, Argentina*

[b] *Institute of Cognitive and Translational Neuroscience (INCYyT), INECO Foundation, Favaloro University, Buenos Aires, Argentina*

[c] *Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Godoy Cruz 2290, Buenos Aires, Argentina*

[d] *Marian Smoluchowski Institute of Physics and Mark Kac Complex Systems Research Center Jagiellonian University, ul. Łojasiewicza 11, PL30–348 Kraków, Poland*

[e] *ARC Centre of Excellence in Cognition and its Disorders, Sydney, Australia*

[f] *The University of Sydney, Brain and Mind Centre and School of Psychology, Sydney, Australia*

[g] *Radiology, Hospital Universitario San Ignacio (HUSI), Bogotá, Colombia.*

[h] *Medical School, Physiology Sciences, Psychiatry and Mental Health Pontificia Universidad Javeriana (PUJ) – Centro de Memoria y Cognición Intellectus, Hospital Universitario San Ignacio (HUSI), Bogotá, Colombia.*

[i] *Departamento de Estudios Psicológicos, Universidad Icesi, Cali, Colombia*

[j] *Instituto de Psicología, Universidad del Valle, Cali, Colombia.*

[k] *Gerosciences Center for Brain Health and Metabolism, Santiago, Chile.*

[l] *.Neuropsychology and Clinical Neuroscience Laboratory (LANNEC), Physiopathology Department, ICBM, Neurosciences Department, East Neuroscience Department, Faculty of Medicine, University of Chile, Avenida Salvador 486, Providencia, Santiago, Chile*

[m] *Memory and Neuropsychiatric Clinic (CMYN) Neurology Department- Hospital del Salvador & University of Chile, Av. Salvador 364, Providencia, Santiago, Chile.*

[n] *Center for Advanced Research in Education (CIAE), University of Chile, 8330014, Santiago, Chile.*

[o] *Servicio de Neurología, Departamento de Medicina, Clínica Alemana-Universidad del Desarrollo, Chile.*

[p] *Medical School, Aging Institute, Psychiatry and Mental Health, Pontificia Universidad Javeriana (PUJ), Bogotá, Colombia.*

[q] *Faculty of Education, National University of Cuyo (UNCuyo), Sobremonte 74, C5500, Mendoza, Argentina.*

[r] *Universidad Autónoma del Caribe, Calle 90, No 46-112, C2754, Barranquilla, Colombia*

[s] *Center for Social and Cognitive Neuroscience (CSCN), School of Psychology, Universidad Adolfo Ibáñez, Diagonal Las Torres 2640, Santiago, Chile*

#First Authors

*Corresponding author: Lucas Sedeño (lucas.sedeno@gmail.com)

Declarations of interest:

None

Abbreviations:
- AD: Alzheimer Disease
- bvFTD: behavioral variant frontotemporal dementia
- ACE: Addenbrooke cognitive examination

- IFS: INECO Frontal Screening (IFS)
- HC: Healthy controls
- TR: repetition time
- VBM: voxel-based morphometry
- SPM12: Statistical Parametric Mapping software
- WM: white matter
- GM: grey matter
- CSF: cerebrospinal fluid
- ACC: anterior cingulate cortex
- Automated Anatomical Labeling (AAL)-Atlas: AAL-atlas

ABSTRACT

Accurate early diagnosis of neurodegenerative diseases represents a growing challenge for current clinical practice. Promisingly, current tools can be complemented by computational decision-support methods to objectively analyze multidimensional measures and increase diagnostic confidence. Yet, widespread application of these tools cannot be recommended unless they are proven to perform consistently and reproducibly across samples from different countries. We implemented machine-learning algorithms to evaluate the prediction power of neurocognitive biomarkers (behavioral and imaging measures) for classifying two neurodegenerative conditions – Alzheimer Disease (AD) and behavioral variant frontotemporal dementia (bvFTD)– across three different countries (>200 participants). We use machine-learning tools integrating multimodal measures such as cognitive scores (executive functions and cognitive screening) and brain atrophy volume (voxel based morphometry from fronto-temporo-insular regions in bvFTD, and temporo-parietal regions in AD) to identify the most relevant features in predicting the incidence of the diseases. In the Country-1 cohort, predictions of AD and bvFTD became maximally improved upon inclusion of cognitive screenings outcomes combined with atrophy levels. Multimodal training data from this cohort allowed predicting both AD and bvFTD in the other two novel datasets from other countries with high precision (> 90%), demonstrating the robustness of the approach as well as the differential specificity and reliability of behavioral and neural markers for each condition. In sum, this is the first study, across centers and countries, to validate the predictive power of cognitive signatures combined with atrophy levels for contrastive neurodegenerative conditions, validating a benchmark for future assessments of reliability and reproducibility.

Keywords: Alzheimer's disease, frontotemporal dementia, machine-learning, executive functions, voxel-based morphometry, classification.

1. INTRODUCTION

Neurodegenerative diseases are a world-wide epidemic [1, 2]. According to the World Alzheimer Report 2015 from the Alzheimer's Disease International, more than 130 million people above age 60 will be diagnosed with dementia in 2050 [3]. Accurate early diagnosis across different neurodegenerative conditions is important for establishing prognosis and accessing adequate treatment. Diagnosis and differential diagnosis represents a clinical challenge due to the complexity of neurodegenerative processes, which disturb the patients' brain structures and

functions, as well as their cognition and behavior [4, 5]. Indeed, current guidelines require the identification of a clinical phenotype and the recognition of specific patterns of atrophy or hypoperfusion in neuroimaging, which are frequently combined with neuropsychological evaluation[6, 7]. Accordingly, accurate and timely diagnosis depends on the clinical expertise to recognize the co-occurrence of clinical phenotype and neuroimaging data. This frequently leads to misdiagnosis in non-specialized dementia centers, even among the most common forms of dementia [8, 9]. This is especially true for developing countries, given their minimal mental-health infrastructure, the lack of regionally organized research, and reliance on non-local (mostly Anglo-Saxon) reference data [10]. The development of objective, automated, and multidimensional decision-support methods for dementia could critically enhance the current clinical toolkit by increasing diagnostic accuracy and confidence [11, 12]. Computational approaches prove most promising in this context, given their potential to detect consistent, reproducible markers across samples from different countries [13, 14]. Capitalizing on this novel approach, here we implement machine-learning algorithms to evaluate the performance of well-defined cognitive/behavioral and neural markers of neurodegeneration for classifying patients with Alzheimer's disease (AD) and behavioral variant frontotemporal dementia (bvFTD) across three different countries.

Neurodegenerative disorders are characterized by abnormalities at molecular, synaptic, neuroanatomical, network-level, cognitive, and behavioral levels [5, 15-17]. Although specific disturbances across these levels are generally associated with different types of dementia, patients present more heterogeneous profiles than would be expected according to diagnostic criteria [5, 15-17]. For example, executive function deficits, which are characteristic of bvFTD, may nevertheless be absent in some patients. Similarly, such deficits may be observed in AD [16]. This is also true for atrophy patterns. For example, a previous cluster-based analysis in bvFTD revealed four distinct patterns, with patients showing temporal-dominant, temporo-frontoparietal, frontal-dominant, or frontotemporal atrophy patterns [18]. Given this variability, computational decision-support methods have been proposed as a powerful novel approach due to their capacity to jointly assess relevant heterogeneous features [11, 19].

In the study of dementia, both cognitive and structural MRI data have arisen as potential candidates to establish effective, affordable, and massive markers of specific diseases [20-23]. Valuable information of the patients' cognitive status can be quickly obtained via cognitive screening instrument [24-26], such as the Addenbrooke cognitive examination (ACE) and the INECO Frontal Screening (IFS), which can be readily administered in general clinical settings. The ACE is a quick tool to evaluate general cognitive domains (e.g., memory, attention) with great sensitivity for AD and dementia in general [22, 24, 25, 27], while the IFS focuses on executive functions [23, 28-31] –a domain poorly evaluated by the ACE [27] – with the aim to identify characteristic deficits in bvFTD patients who have relative preservation of other cognitive domains [15]. These two widely used

screening tools [27, 28], in particular, have yielded good accuracy rates to discriminate dementia patients from healthy controls (with scores from 83 to 96%) [22, 23, 27, 28]. On the other hand, structural MRI is a non-invasive method, generally included in routine assessments of dementia, characterized by lower costs than other candidate biological biomarkers [32]. Of note, though anatomical images are only visually inspected for clinical evaluation, they offer automatically derivable metrics of atrophy, which can reveal subtle neural alterations untraceable to the naked eye [33]. Furthermore, several studies applying computational decision-methods on anatomical neuroimages have yielded good accuracy rates (from 80 to 100%) to discriminate AD and bvFTD patients from healthy controls [34-40].

Although this evidence underscores the potential role of cognitive screenings and neuroimaging as feasible and effective markers for dementia, several limitations undermine their reliability and potential generalizability. First, works testing the classification properties of the ACE and the IFS rely on simple statistical methods to establish cut-offs for maximizing true positives and minimizing false positives, but no machine-learning approaches have yet been applied to evaluate the performance of these screenings, alone or in combination with neural measures [22, 23, 27, 28]. Second, although a few neuroimaging studies have obtained high classification rates via sophisticated machine-learning approaches (such as support-vector machines), none of them has assessed the generalizability of their findings to new datasets [34-40], casting doubts on their potential overfitting confounds and widespread translational relevance. Third, no previous study has assessed the combined sensitivity of neuropsychological screenings and atrophy measures with computational-decision methods, despite the evidence underscoring the potential of this approach to tackle behavioral and anatomical heterogeneity [5, 15-17]. Finally, most previous research is based only on samples from one clinic (typically, from Anglo-Saxon populations), so that their results might not be robust against the variability introduced by cross-center differences in recording parameters, diagnostic criteria, and the patients' socio-demographic profiles.

In sum, computational-decision methods combining cognitive screenings and atrophy measures to detect dementia patient profiles prove promising, but key tests of their robustness remain to be determined. Here, we implemented machine-learning algorithms to evaluate the power of cognitive parameters and brain atrophy measures for classifying AD and bvFTD patients across three countries, including Latin-American and Anglo-Saxon samples. In addition, we applied a leave-two-out cross-validation approach to test the accuracy rates within a reference dataset (within-country analysis), and then performed a cross-country validation to further evaluate the generalization of our findings. We predicted that the ACE, given its target domains, would contribute more than the IFS for the classification of AD. The IFS, given its emphasis on frontal executive functions, was hypothesized to be more sensitive to bvFTD. Moreover, we expected that cognitive measures would generalize better than atrophy features, as the latter present higher variability due to

differences in MRI scanner and acquisition parameters among centers[41]. Finally, we anticipated that the combination of cognitive and atrophy features would yield higher classification scores for both within-country analysis and cross-country generalization.

2. MATERIALS AND METHODS

2.1. Participants

The data analyzed here partially belong to a previously reported multicenter protocol study [42], and comprised 202 participants from three countries. Fifty-seven patients fulfilling revised consensus criteria for probable bvFTD [43] and 29 patients who satisfied international criteria for AD [44] were recruited from three international clinics: INECO Foundation, Argentina (Country-1, C_1); San Ignacio University Hospital, Colombia (Country-2, C_2); and FRONTIER, the Frontotemporal Dementia Research Group, based in Sydney (Country-3, C_3); further details of the origin of each sample are in Table 1).

As described in previous reports [45-47] the clinical diagnosis in each center was established by a standard examination –involving extensive neurological, neuropsychiatric, and neuropsychological assessments–, and each case was discussed by a multidisciplinary clinical meeting of AD and bvFTD experts. All patients were in early/mild disease stages, and they did not fulfil criteria for specific psychiatric disorders. Patients presenting primarily with language deficits were excluded (further details about the clinical evaluation are reported in [42]). Each patient sample was matched on sex, age, and education with its own control group from the same scanning center (see Table 1). Healthy controls (HC, 116 in total) presented no history of psychiatric or neurological disease.

Participants (or their Person Responsible) provided signed informed consent in accordance with the Declaration of Helsinki. The study protocol was approved by the institutional Ethics Committee of each center.

Table 1: Summary of demographic data for each group.

	Country-1			
	C_1-FTD	**HC**	***F*-value**	***p*-value**
Age [years]	66.72 ±9.56	68.73 ±8.48	0.57	.452
Education [years]	15.05 ±2.97	15.86 ±2.92	0.85	.360
IFS	16.30 ±7.01	25.37 ±1.84	46.46	<.001

ACE	76.72 ±15.39	93.76 ±4.25	33.02	<.001
			Chi-square	***p*-value**
Gender [M/F]	F = 11 (16) M = 7 (13)	F = 21 (24) M = 9 (10)	0.40	.527
	C_1-AD	**HC**	***F*-value**	***p*-value**
Age [years]	75.37 ±8.72	71.54 ±6.00	2.57	.117
Education [years]	12.94 ±4.97	15.13 ±3.10	2.81	.102
IFS	16.59 ±4.46	25.07 ±1.87	64.34	<.001
ACE	69.50 ±14.11	93.90 ±4.28	58.89	<.001
			Chi-square	***p*-value**
Sex [M/F]	F = 13 (13) M = 3 (3)	F = 17 (18) M = 5 (6)	0.08	.766
Country-2				
	C_2-FTD	**HC**	***F*-value**	***p*-value**
Age [years]	66.55 ±9.37	61.18 ±7.74	2.72	.109
Education [years]	15.89 ±2.31	14.73 ±5.41	0.38	0.542
IFS	12.78 ±6.21	22.87 ±2.97	38.13	<.001
ACE	--	--	--	--
			Chi-square	***p*-value**
Sex [M/F]	F = 7 (12) M = 2 (4)	F = 11 (16) M = 11 (12)	2.03	.155
Country-3				
	C_3-FTD	**HC**	**F-value**	***p*-value**
Age [years]	64.90 ±9.44	69.50 ±6.43	1.88	.184
Education [years]	12.20 ±3.47	14.27 ±2.77	2.50	.128
IFS	--	--	--	--
ACE	75.36 ±14.99	96.25 ±2.42	22.73	<.001
			Chi-square	***p*-value**
Sex [M/F]	F = 4 (4) M = 7 (8)	F = 5 (7) M = 7 (8)	0.06	.794
	C_3-AD	**HC**	***F*-value**	***p*-value**
Age [years]	64.00 ±5.83	69.50 ±6.43	4.34	.051
Education [years]	12.80 ±2.89	14.27 ±2.77	1.47	.238
IFS	--	--	--	--
ACE	62.10 ±11.58	96.25 ±2.41	100.09	<.001
			Chi-square	***p*-value**

Sex [M/F]	F = 3 (5) M = 7 (8)	F = 7 (8) M = 5 (7)	1.76	.183

Table 1. Subject groups for the two diseases and the three countries. Because of occasional missing data, we indicate for each gender and country, first the number of subjects actually used in the current study and then in parentheses the total of recruited subjects. Note also that different cognitive tests were used depending on the country. Age, education, IFS and ACE scores are given with the mean± SD. C_1 = Country-1; C_2 = Country-2; C_3 = Country-3.

2.2. Cognitive assessment

Participants completed a general cognitive screening, the ACE [48], and an executive function brief cognitive test, IFS [47]. The IFS [47] is a sensitive tool to detect executive dysfunction in patients with dementia [19, 45, 49-57]. This test includes eight subtests that evaluate response inhibition and set shifting, abstraction skills, and working memory. The global score of the IFS (the sum of the subtests, with a maximum value of 30) was considered here, as in previous works [58]. The ACE [48] is a sensitive tool to detect early stages of dementia and, more particularly, to distinguish between AD and FTD patients [27]. This test evaluates orientation, attention, memory, verbal fluency, language, and visuospatial ability (with a maximum total score of 100).

2.3. Structural imaging

2.3.1. Image acquisition

We followed the guidelines from the Organization for Human Brain Mapping[59] to report the acquisition and preprocessing steps. Structural images were obtained from Country-1 participants through whole-brain T1-weighted spin echo sequences in a 1.5T Phillips Intera scanner, and were acquired parallel to the plane connecting the anterior and posterior commissures with the following parameters: matrix size=256×240, 120 slices, approx. 1x1x1 mm (1x0.97x0.97 mm); repetition time (TR)=7489 ms; echo time (TE)=3420 ms; flip angle=8°, acquisition time=7 minutes. Country-2 participants were scanned in a 3T Philips Achieva scanner. Whole-brain structural T1-rapid gradient-echo (MP RAGE) anatomical 3D scans were acquired with the following parameters: matrix size=256x256, 160 slices, 1x1x1 mm isotropic; TR=8521 ms; TE=4130 ms; flip angle=9º ms, acquisition time=8 minutes. In Country-3, whole-brain structural T1-weighted spin echo sequences were acquired through a 3T Philips MRI scanner with a standard head coil (matrix size=256x200, 256 slices, 1x1x1 mm isotropic; TR=5903 ms: TE=2660 ms; flip angle= 8º; acquisition time =7.42 minutes).

2.3.2. Voxel-based morphometry

Structural images from each center were analyzed via voxel-based morphometry (VBM) with the DARTEL Toolbox of the Statistical Parametric Mapping software (SPM12), following validated procedures[42]. Images were segmented into white matter (WM), grey matter (GM), and cerebrospinal fluid (CSF). Then, based on the segmented GM and WM images, we created a template from each center's complete data set with the "DARTEL (create template)" module. Next, the final template from the previous step was affine-registered into the MNI space with the "Normalize to MNI Space" module from DARTEL Tools, and then this transformation was applied to all segmented GM scans to translate them into standard space (images were modulated by Jacobian determinants). Finally, an isotropic Gaussian kernel of 12-mm full width at half maximum (FWHM) was applied to all images.

Following previous procedures [60], we used a mask resembling the characteristic atrophy pattern of AD and bvFTD, respectively, to extract the GM volume for each participant. Given that we aim to test cognitive and atrophy features with a cross-center validation strategy, we used this 'a priori' feature selection to avoid the potential bias of a data-driven approach, which might find optimum classification features for an specific dataset that do not necessarily enable high classification rates in another one [60]. As in a previous work, the general bvFTD atrophy mask was defined using the Automated Anatomical Labeling (AAL)-Atlas [61], and involves the main fronto-insulo-temporal areas of early degeneration, including [15, 43, 62, 63]: the anterior cingulate cortex (ACC), the orbitofrontal cortex, the gyrus rectus, the inferior frontal gyrus, the frontal middle gyrus, the amygdala, the basal ganglia (caudate nucleus, putamen and pallidum), the insular cortex, the hippocampus and parahippocampus (see Supp. Fig. 1). For AD, the general mask was also based on the AAL-atlas including the most common atrophy areas for this dementia such as the posterior cingulate cortex, the hippocampus, the parahippocampus, the amygdala, the angular gyrus, the precuneus, and the temporal superior and middle gyrus [18, 64] (see Supp. Fig. 1). Regions for both masks were selected bilaterally.

2.4. Classification analysis

To test the power of cognitive and brain atrophy features for classifying AD and bvFTD patients, we first implemented machine-learning algorithms within a reference dataset. For this analysis (from now on referred to as "within-country" approach), we selected the largest dataset available with full completion of cognitive screenings, namely, the one from Country-1. Then, we performed a cross-country classification analysis to further validate the generalization and prediction power of our findings with the within-country approach. In short, the classifier was trained with Country-1 subjects, and tested on participants from Country-2 or Country-3.

For both classification approaches, six features of interest were included in the analysis: IFS and ACE scores, the volume of atrophy, and demographic data including sex, age and number of years of formal education. Participants missing any of these parameters were excluded from

further study. The general processing and analysis steps (conducted separately for the two diseases) are presented in Figure 1. First the data was standardized by converting to z-scores, so that each feature in the control group had a zero mean and standard deviation of one (see Figure 1; the details of z-scoring depend on the chosen cross-validation scheme, and are described in Sections 2.4.2 and 2.4.3). This also transforms the distances between data points to be on the same scale. In this way, this step is preferred for some clustering and classification methods in order to improve their convergence and to avoid bias merely due to different feature ranges. The only categorical variable, sex, was not standardized but was given values ±1, since otherwise the values would be different in each group due to different proportions of females and males. Next, the hypothetically most informative features were inspected by graphing pairs of dimensions for Country-1 –the reference dataset (see 2.4.1 Clustering, below). This exploration was completed by principal component analysis (PCA), an exploratory technique that finds the most informative combination of features as measured by the explained variance of the data. We used MATLAB's default implementation of PCA with a singular value decomposition of feature correlation matrix. Indeed, the PCA showed that for Country-1 AD (C_1-AD) [vs Country-1 FTD (C_1-FTD)] the first principal component explains 46% (43%) of variance (of the standardized data), and 82% of this component (97%, calculated as the norm of the coefficient vector) comprises roughly equal shares of IFS, ACE, and atrophy scores. Due to the small number of features, dimensionality reduction was not needed, and so the clustering and classification analyses were not performed on the principal components, but only on the z-scored parameters.

2.4.1. Clustering

Clustering is an exploratory technique used when the actual groups (like disease and HC, here) are not known. The objective was to assess whether the subjects from the reference dataset (Country-1) could be clustered into two separate groups, without any knowledge about the meaning of the features. Such an exploration might provide additional insights into the distribution of the data, and help predict difficulties at the classification stage, and also show potential errors in the diagnostic label or the presence of potential sub-groups of participants. Given that the clustering analysis was conducted solely for exploratory purposes and that its results were not used in any form for the classification analyses (see sections 2.4.2 and 2.4.3), we used the classic k-means algorithm (see, e.g., [65]) implemented in Matlab, with k = 2 (see Supp. Data 1). To evaluate differences in the expected geometries, we use the Euclidean distance. Although there are many other more sophisticated clustering methods, k-means provided us with just enough baseline information (i.e., that the data can be reasonably divided into two meaningful groups) to proceed to classification.

2.4.2. Within-country classification

We used a default logistic regression classifier[66] implemented in Matlab, (see Supp. Data 1) to discriminate patients from HC for Country-1, our reference data-set. Logistic regression is a type of regression specifically designed to model probabilities –in this case, the probability of a subject belonging to a condition group (AD or bvFTD). Since probability is always limited to the interval from 0 to 1, predictions of linear regression are ill-defined, because they may lie outside of this interval. On the other hand, logistic regression uses a logistic function, which is a smooth step function whose values range exactly from 0 to 1, and is thus well fitted to the classification problems. This scheme was used, since the sample of this country was the largest one compared to the other ones. Also, only participants from this country have completed all the cognitive screening assessments (IFS and ACE). These characteristics are fundamental to train data both for classification within the country as well as for cross-country validation.

To evaluate the classifier's performance, we used a leave-two-out cross-validation that is computationally more demanding but allows a better sampling than either leave-one-out or 10-fold cross-validation [67, 68]. In each run, two participants were held out from the training set: one from the condition group (of size n_{cn}) and one from the control group (of size n_{cl}); the classifier was trained on the remaining $n_{\mathrm{cn}} + n_{\mathrm{cl}} - 2$ subjects and tested on the two hold-out subjects. In this cross-validation scheme, all *z*-scores were based on the mean and standard deviation of the $n_{\mathrm{cl}} - 1$ training controls. The idea is to use only what we know about the training set for classification. Any transformation of the test set must be performed using only that knowledge, including the values of training means and standard deviations.

The true/false positive/negative scores were accumulated over all runs. Due to the small size of the data set, the cross-validation was exhaustive, i.e., all $n_{\mathrm{cn}} \cdot n_{\mathrm{cl}}$ pairs of condition-control subjects were tested. From the accumulated scores, the sensitivity, $TP/(TP + FN)$, and specificity, $TN/(TN + FP)$, values were obtained.

To evaluate a classifier's performance the receiver-operating characteristic (ROC) curves [69] were calculated, see Figure 3. The curves depict sensitivity (true positive rate) plotted versus 1-specificity (false positive rate). They are calculated by shifting the probability threshold (a moving vertical decision line in the histograms of Figure 3 and 4) by means of which subjects are assigned to groups. The probability range was from 0 to 1 in steps of 0.05. The area under the ROC curve (AUC) and the accuracy classification rates were used as metrics of the classifier's performance. These were also used to evaluate the relative importance of a given feature for prediction results: first, the classification was performed on the whole feature set (denoted "All" in the ROC plots), and then, one by one, each of the six (or five for Country-2 and Country-3 data sets) features was omitted in the classification. If the resulting AUC or accuracy score decreased without a given feature, the feature can be considered relevant, conversely if these values

increased, the feature can be considered to introduce unwanted noise or correlations. The statistical significance of differences between the AUC involving all features and each of the AUC when one feature was removed, was estimated with the Mann-Whitney statistic[70] (*p*-values are expressed after Bonferroni correction). Typically this test would use as input $n_{cn}+n_{cl}$ classification scores, which in this case gives a small sample size (i.e. the expected confidence intervals and coverage probability in[71]). As a rule of thumb to obtain 90% power for distinguishing AUC 0.96 and 0.95 at 0.05 significance level one needs a sample size over 6000, which means in our case the tests for ROC differences are severely underpowered [40]. Since we perform the leave-two-out procedure, we decided to bootstrap the sample to 2·ncn·ncl scores (of the order of 1000). However, such a resampling leads to a stronger bias and produces underestimated p-values. We nevertheless provide these p-values in the Supplementary Data 2 to accompany AUC reported in Table 2.

Figure1

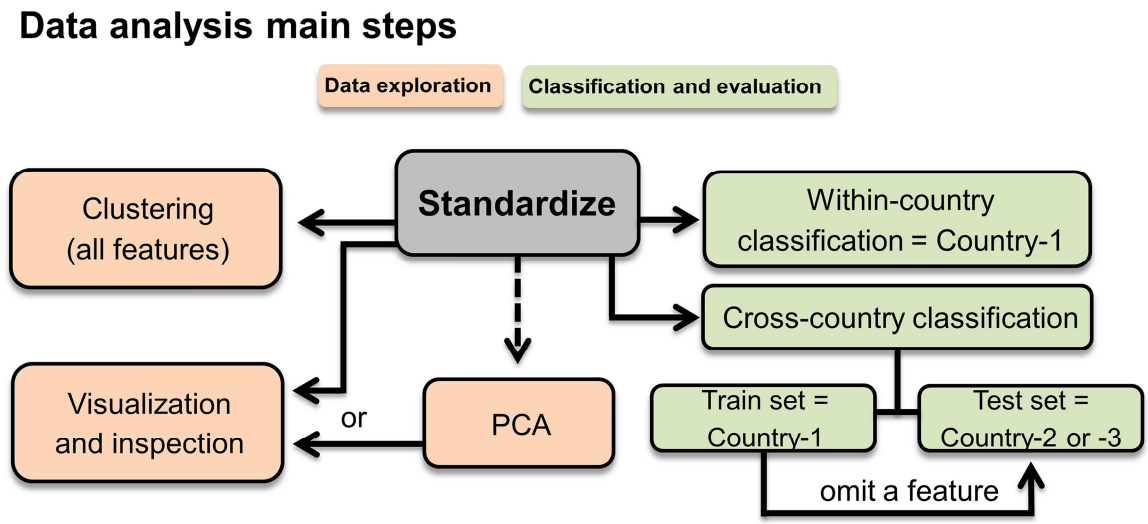

Figure 1. Standardize (gray box): data were standardized by converting them to z-scores, so that each feature in the control group had a zero mean and standard deviation of one. Data exploration (light orange boxes): these procedures were used only to explore and obtain knowledge about the behavior of the data. Clustering: we used a *k*-means algorithm with *k*=2 to separate groups in two clusters to explore data distribution, and evaluate the presence of potential sub-groups of participants (details in section 2.4.1). Visualization and inspection: the hypothetically most informative features (cognitive screenings and atrophy) were inspected by graphing pairs of dimensions for the reference dataset (Country-1) (details in section 2.4). Principal component analysis (PCA): We used MATLAB's default implementation of PCA to explore the most informative combination of features as measured by the explained variance of the data. Classification (light green boxes): Within-country classification: we implemented a logistic regression classifier with cognitive and brain atrophy features within the Country-1 dataset, given that it was the largest one with full completion of cognitive screenings. To evaluate the performance of this model, we used a leave-two-out cross-validation scheme. Cross-country classification: this was performed to further validate the generalization and prediction power of our findings. The logistic regression classifier was trained with Country-1 subjects and tested on participants from Country-2 or Country-3. Finally, to evaluate the relevance of each feature, after performing the classification with the whole feature set, the procedure was repeated but one-by-one each of the features was omitted in the classification (details in sections 2.4.2 and 2.4.3).

Figure 2

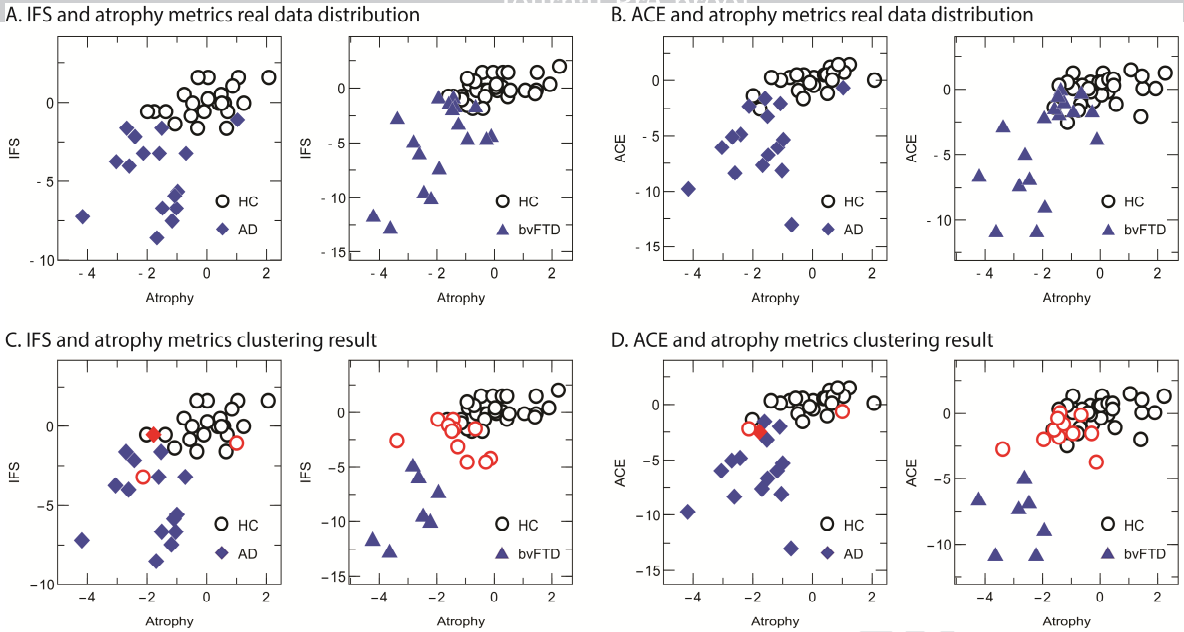

Figure 2. Atrophy measures and cognitive data distribution. A. Real distribution of IFS and atrophy data. The degree of brain atrophy and the IFS score are standardized in z-scores. **B. Real distribution of ACE and atrophy data.** The degree of brain atrophy and the IFS score are standardized in z-scores. **C. Clustering of IFS and atrophy results from panel A.** Red-white circles represent patients wrongly identified as controls, and red diamonds represent controls who were mistaken with patients. **D. Clustering of ACE and atrophy results from panel B.** Red-white circles represent patients wrongly identified as controls, and red diamonds represent controls who were mistaken with patients.

2.4.3. Cross-country classification

For cross-country classification, the classifier was trained on Country-1 subjects and tested on either Country-2 or Country-3, given that Country-1 was the largest one and included data from both the IFS and ACE for all the participants. As above, the *z*-scores were computed based on the mean and standard deviation of the control group of the training set (Country-1). Again, this is to avoid so called data leakage: if we intend to standardize the data based on the control group, then we must not use any knowledge from the test set. *Z*-scoring based on the control group of the test set already introduces some knowledge (via mean and standard deviation) about how the set is divided. We circumvented this issue by using only the training set for that purpose. Note that instead of z-scores the classifiers can also take as input PCA components – again, with PCA loadings computed only on the training set. Although the performance of those classifiers is comparable or only slightly better, for simplicity and interpretability of features, we report on the logistic regression classifier using z-scored features only (See Supp. Table 3 for results with PCA components).

Thanks to having separate training and test subjects, leave-two-out cross-validation was no longer necessary to estimate the out-of-sample error as above; however, it could be used to obtain an estimate on the classification uncertainties. To that purpose, we performed the same leave-two-out scheme as before, with two subjects being removed from the training set, with the crucial difference that the test set always stayed the same (i.e., it comprised all the Country-2 or Country-3 subjects). In other words, this scheme models a possible variance of the training set, but does not involve variance of the test set. While in the within-country case, in each run there were only two subjects classified and the results had to be accumulated to obtain true/false positive/negative scores, in the cross-country case the test set was big enough to allow computing the scores already in each run. Consequently, we calculated means and standard deviations of these scores over all cross-validation runs, thus obtaining a more robust estimate. The AUCs were then given error estimates as well, by shifting each point in ROC by respective standard deviation and calculating the areas under the maximally shifted curves. The errors are not symmetric, so for simplicity we report only on the larger of the two. The statistical significance of the differences between AUCs can already be approximately inferred from such error estimates. However, for comparability, the *p*-values were estimated following the same procedure as indicated in the previous section 2.4.2[70] (with the only difference that leave-one-out bootstrap was used to obtain comparable sample sizes: $n_{\text{train}} \cdot n_{\text{test}}$).

2.5. Research data for this article

Cognitive and neuroimaging final processed data from the patients of this study are available on the “Open Science Framework” repository under the following link: https://osf.io/ctjkv/

3. RESULTS

3.1. Clustering

In the case of k-means clustering with Euclidean distance for C_1-AD, the patients and control groups are clearly visible, and the obtained partition is approximately correct with IFS and atrophy measures, with only 3 mis-classifications (2 false negatives and 1 false positive) (Figure 2-C). For C_1-FTD, the partition is different from the real distribution with only the extreme atrophy and IFS score subjects clustered separately from the control group. The other 10 bvFTD individuals were clustered together with the control. A similar clustering partition was found regarding the ACE and atrophy features, in which the C_1-AD patients and controls present 3 misclassifications (2 false negatives and 1 false positive), while the C_1-FTD showed 10 misclassifications (10 false negatives). These results show that the unsupervised clustering techniques alone do not deal well with the borderline cases, especially given sparse data (as in the case of bvFTD patients).

Figure 3

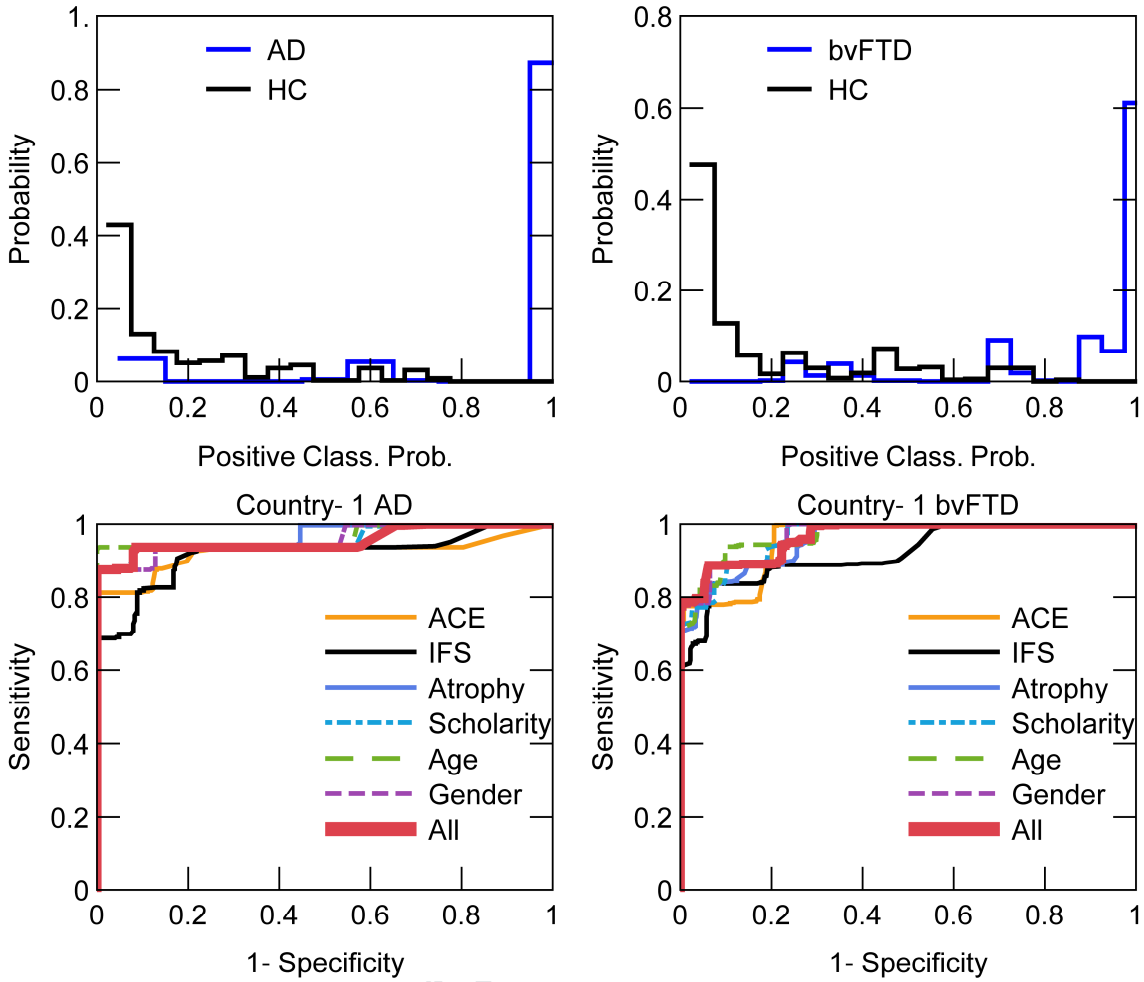

Figure 3. Within-country classification. The two top panels depict the histograms of the probability of belonging to the patient group, as revealed by logistic regression. The bottom panels correspond to ROC curves obtained for the groups' data from the first row. Different curves show the ROC calculation omitting the feature denoted in the legend on a one-by-one basis.

3.2. Within-country classification

The results of classification are illustrated in Figure 3 and also in Table 2. The histograms show the probability score (accumulated over all runs of cross-validation) provided by the logistic regression, which is the probability of assigning a subject to the dementia group. Visibly, the subjects from the control and dementia groups are well separated.

In Figure 3, the ROC curves shown refer to classifiers using all the available features or all but one (seven curves in total). The combination of features yielded high maximal classification accuracy rates for both AD (0.94) and bvFTD (0.91). Notably, in C_1-AD group, general performance decreased the most when the IFS parameter was removed, which highlights this measure as a crucial feature in predicting the population of AD in this cohort. In addition, all AUC results

presented significant differences compared to the one including all features, except for the one excluding gender (see Supp. Data 2). In the C_1-FTD group, classification accuracy values decreased the most upon removal of IFS scores and atrophy measures. Only this AUC and the one excluding ACE presented significant differences against the one combining all features (see Supp. Data 2). One must be cautious, however, that the result of removing a feature from the classifier depends on the correlations between the features. It is thus not surprising that the performance did not drop after removing atrophy in C_1-AD group, since it is highly linearly correlated with IFS, as visible in Figure 2-A. As a complementary analysis, we tested this same within-country approach but combining the data sets that share the same cognitive screening (i.e., Country-1 and Country-2, which share the IFS; and Country-1 and Country-3, which share the ACE), with the aim of increasing the sample size of our analyses. As shown in Table 2, accuracy scores were around 0.88 when all features were considered, and results also showed that cognitive screening scores and atrophy measures were relevant features for AD and bvFTD classification (see Supp. Fig. 2 and Supp. Data 3).

Table 2: Within-country and cross-country classification results (for all features and excluding one feature at a time from the overall model)

Group		All	IFS	ACE	Atrophy	Age	Gender	Scholarity
C_1-AD	AUC	0.956	0.918	0.924	0.967	0.964	0.958	0.959
	Acc	0.938	0.865	0.906	0.938	0.967	0.938	0.938
	1-Spec	1	0.824	1	1	1	1	1
	Sens	0.875	0.906	0.813	0.875	0.935	0.875	0.875
C_1-FTD	AUC	0.967	0.921	0.958	0.957	0.969	0.969	0.965
	Acc	0.912	0.885	0.895	0.885	0.920	0.912	0.893
	1-Spec	0.937	0.937	0.794	0.935	0.902	0.935	0.898
	Sens	0.877	0.833	0.996	0.835	0.939	0.889	0.887
C_2-FTD	AUC	0.935 ±0.022	0.800 ±0.035	--	0.919 ±0.026	0.949 ±0.020	0.935 ±0.019	0.925 ±0.022
	Acc	0.913 ±0.000	0.784 ±0.008	--	0.913 ±0.011	0.914 ±0.007	0.913 ±0.000	0.913 ±0.000
	1-Spec	1 ±0.000	0.999 ±0.007	--	0.997 ±0.016	1 ±0.000	1 ±0.000	1 ±0.000

	Sens	0.818 ±0.000	0.548 ±0.017	--	0.821 ±0.019	0.821 ±0.014	0.818 ±0.000	0.818 ±0.000
C_3-FTD	AUC	0.906 ±0.021	--	0.795 ±0.040	0.938 ±0.021	0.919 ±0.0151	0.908 ±0.018	0.882 ±0.028
	Acc	0.913 ±0.000	--	0.784 ±0.007	0.913 ±0.011	0.914 ±0.007	0.913 ±0.000	0.913 ±0.000
	1-Spec	1 ±0.000	--	0.999 ±0.007	0.997 ±0.016	1 ±0.000	1 ±0.000	1 ±0.000
	Sens	0.818 ±0.000	--	0.548 ±0.017	0.821 ±0.019	0.821 ±0.014	0.818 ±0.000	0.818 ±0.000
C_3-AD	AUC	1.0 ±0.000	--	0.936 ±0.060	1.0 ±0.000	1.0 ±0.000	1.0 ±0.000	1.0 ±0.000
	Acc	1.0 ±0.000	--	0.906 ±0.029	1.0 ±0.000	1.0 ±0.000	1.0 ±0.000	1.0 ±0.000
	1-Spec	1.0 ±0.000	--	0.847 ±0.031	1.0 ±0.000	1.0 ±0.000	1.0 ±0.000	1.0 ±0.000
	Sens	1.0 ±0.000	--	0.978 ±0.076	1.0 ±0.000	1.0 ±0.000	1.0 ±0.000	1.0 ±0.000

Table 2: Classification outcomes for each group and conditions (rows) computed for models using different features (columns). The column denoted "All" corresponds to results gathered from a model trained with all the features, while the others correspond to results obtained *excluding* the denoted feature. In all cases the model was trained with the Country-1 cohort, and prediction computed for the other countries. For cross-country classification the areas are provided with error estimates as described in Sec. III B 2. Auc = area under the ROC; Acc = maximal accuracy; 1-Spec = 1-specificity (for maximal accuracy); Sens = sensitivity (maximal accuracy)

3.3. Cross-country classification

As discussed above, it is uncertain how variable are the biomarkers of neurodegenerative diseases across countries. A way to test for that is to attempt out-of-sample predictions, i.e., train the classifier with data from one country and test it on the others. This case is presented in the Table 2 and Figure 4 (in the same format as used in Figure 3) where the results for Country-2 bvFTD (C_2-FTD) are in the left column, those from Country-3 bvFTD (C_3-FTD) and Country-3 AD (C_3-AD) in the middle and right columns, respectively. ROC curves showed that the combination of features yields high accuracy rates for C_2-FTD (0.91), for C_3-FTD (0.91), and for C_3-AD (1.00). Across countries, the cognitive screenings were the features that most contributed to the classification rates. Atrophy and demographic features showed lower accuracy results compared to cognitive measures. In bvFTD patients from Country-2 and -3, all the AUCs presented significant differences when features were excluded, compared to the AUC combining all of them (except for gender in Country-2) (see Supp. Data 2 and the error estimates in Table 2). Notwithstanding, for the

classification between controls and bvFTD from Country-2 and -3, atrophy data was the most relevant feature to correctly identify non-pathological cases (specificity outcome, 1-Spec). In the case of AD, the only AUC that presented significant differences was the one in which ACE was not included as a feature, given that the other models presented the same values as the AUC combining all the features (see Supp. Data 2). Overall, results showed that the training data obtained from the Country-1 cohort is able to predict classes with high scores for both bvFTD and AD in the other two countries, demonstrating the robustness of the approach and the reliability of the markers.

Figure 4

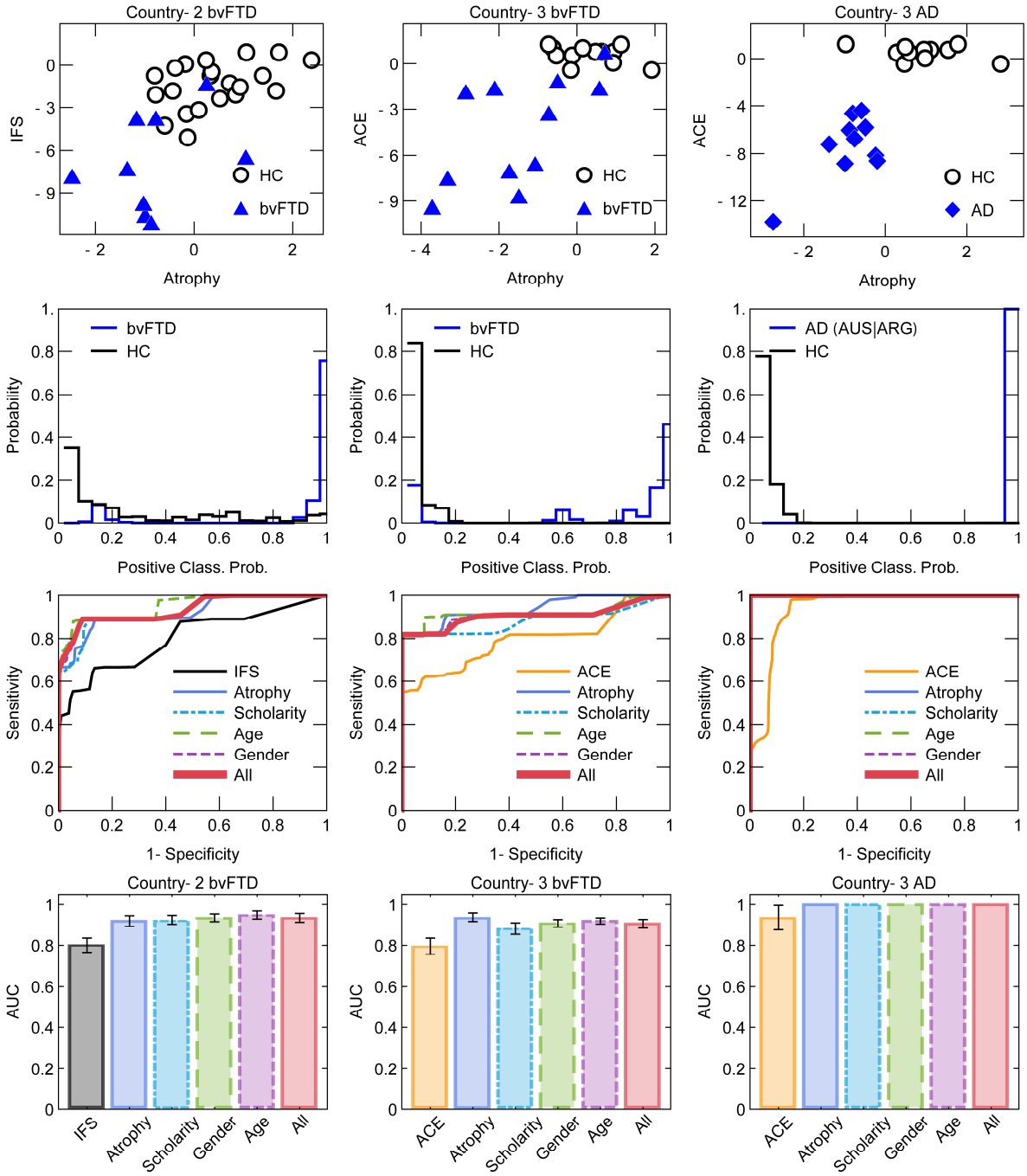

Figure 4. Cross-country classification. The three top panels depict the real data distribution in z-score values. The three middle panels show the histograms of the probability of belonging to the diseased group, as revealed by logistic regression. The graphs in the three bottom panels correspond to the ROC curves obtained for each condition, by considering all the features ("All") or by omitting the single features indicated in the legends. The fourth row illustrate the effect of removing a single feature from classification. Removing IFS or ACE affects results the most, which indicates their high informativeness in distinguishing FTD and AD from controls.

4. DISCUSSION

This is the first work to validate the relevance of combined cognitive-behavioral assessment and neuroanatomical measures for identifying bvFTD and AD patients from controls, across countries, based on machine-learning algorithms. We obtained high classification rates (> 0.91) for both diseases in Country-1. More crucially, these results offered high predictive power (> 0.91) when used to classify new patient cohorts from other international centers using different MRI acquisition equipment. Therefore, despite further research is needed, our study strongly supports the implementation of computer-based methods combining cognitive screenings and anatomical information as a potential gold-standard for clinical neuroscience [12, 72].

As shown by the within-country analysis of Country-1, classification of bvFTD via combined measures (> 0.91) surpassed previous outcomes based on cognitive screenings (using simple statistical methods) [23, 27, 28] and anatomical neuroimaging features (relying on data-driven computational approaches) [35-39]. Similarly, discrimination of AD patients through combined measures was higher than or similar to previous results based solely on cognitive or atrophy measures [22, 24, 25, 27, 34, 73]. Moreover, our cross-country validation presented a large predictive accuracy power for both diseases (> 0.91 for bvFTD patients from Country-2 and -3, and 1.00 for AD patients from Country-3), highlighting the reliability of these markers for optimal classification of new patients. This robust generalization to independent data suggests that these measures are able to face the variability introduced by clinical assessments and MRI recordings from different centers, and that they might reflect universal properties and alterations of neurodegenerative conditions. This characteristic is critical to evaluate the potential role of a measure as an early biomarker for a disease [13, 14].

Furthermore, the latter results are particularly relevant, as previous research on dementia has yielded high detection and differentiation rates but limited generalization power [34-39, 73]. This is especially true for data-driven studies in which anatomical feature selection was based on the minimum number of areas providing optimal separation of samples for a specific dataset, which was later used for the validation process [34-39, 73]. Although this procedure can yield large accuracy rates, it does not necessarily enable the same performance for independent cohorts, given that features might be specific for the initial data [12]. To overcome this potential bias, we first selected hypothesis-driven cognitive and atrophy features [60] reported as hallmarks of bvFTD and AD in a country-unspecific fashion, and tested them in two independent samples. The high accuracy rates thus obtained extend previous MRI studies successfully using cross-center validation methods in Anglo-Saxon AD samples [21, 41, 60, 74-76]. Of note, to our knowledge, present results for this condition

(1.00 for AD from Country-3) surpass even the highest outcomes reported in the literature so far, emphasizing the relevance of combined neuropsychological and neuroanatomical methods.

Such a combined approach affords a more plausible model of the complex alterations found in dementia patients [4]. Neurodegenerative disorders are characterized by abnormalities at multiple levels –from molecular deficits to behavioral impairments[5, 15-17]. Also, although these abnormalities tend to present specific profiles according to different types of dementia, several works have shown a more heterogeneous scenario. In this way, despite that executive functions are a target for bvFTD [15], such deficits may prove subtle and they are nosologically unspecific –in fact, they are frequently observed in AD[16]. The same is true for other cognitive functions, such as memory skills, which is compromised in both AD and bvFTD[77, 78]. A similar scenario concerns atrophy patterns, as bvFTD patients might present anatomical alterations similar to those of AD (e.g., a posterior pattern comprising temporofrontoparietal regions), while AD may involve subtle frontal alterations overlapping with those of bvFTD [79, 80]. Although future research is needed, the integration of neuropsychological and neuroanatomical measures may prove critical to address this variability[11, 19] and provide useful insights for clinical settings. Our findings represent the first demonstration of the feasibility of this approach both within and across centers.

4.1. Contribution of cognitive screenings

To our knowledge, this is the first study showing the high reliability and predictive power of the ACE and IFS –two instruments greatly sensitive to AD and bvFTD, respectively [24-26]–, based on machine-learning methods with a cross-country validation. In the within-country analysis, the IFS was distinctively relevant for the classification of bvFTD patients than the ACE –classification rates decreased more upon exclusion of IFS results (Table 2). This was expected given that the IFS was specially designed to target executive function deficits (a domain poorly assessed by the ACE [28]), which are characteristically affected in bvFTD [15]. Moreover, previous studies have shown that relative to the ACE, the Mini-Mental State Examination, and even other frontal screenings such as the Frontal Assessment Battery, the IFS proves better to discriminate bvFTD patients from controls and other pathologies [23, 28, 57]. Regarding AD, although we expected a greater contribution from the ACE given its proven sensitivity for this condition [22, 24, 25, 27], it was again the IFS that afforded the greatest discriminatory contribution. This might partially reflect the heterogeneity in cognitive and atrophy profiles of dementias, as discussed above [5, 15-17]. Indeed, AD patients can also present with deficits in executive function[16] (even in early stages and previous to global cognition alterations [81, 82]), as well as frontal atrophy [79, 80] (which is more common with disease progression[79]). Furthermore, the IFS has systematically differentiated and discriminated AD patients from healthy controls [28-31, 57]. Thus, though unexpected, the relevance of this instrument for identifying AD patients aligns with previous evidence. Moreover, the classification rates obtained when the ACE

was considered with other features but not the IFS were high (0.86), and similar to previous studies that tested its (isolated) discriminative power [22, 24, 25, 27].

Despite the partial missing data of these cognitive measures in the two independent cohorts, the cross-country validation approach showed that classification of bvFTD from Country-2 was mainly driven by the IFS, whereas classification of both bvFTD and AD from Country-3 was mainly informed by the ACE. These findings constitute the first demonstration of the reliability and predictive power of these cognitive screenings to novel and unseen data from socio-culturally diverse contexts. Since cognitive screenings are standardized instruments, they rely on predefined procedures, norms, and scoring rules that help reducing bias and discrepancies in administration and interpretation [27, 28]. This may explain the consistency of our results with both tools across countries. Additionally, these instruments were specially designed to target specific cognitive domains affected in each disease and provide useful information in clinical settings. Moreover, their psychometric properties have been further evaluated and validated in several works [22, 24, 25, 27, 28]. Finally, these screenings have yielded large differences between dementia patients and controls from different origins (including both Anglo-Saxon and Latin America participants) [27, 29, 30, 83], which underscores their reliability and consistency in the face of socio-cultural diversity.

Briefly, our findings validated for the first time the application of the ACE and IFS as robust and reliable markers to discriminate both bvFTD and AD patients from healthy controls based on machine-learning algorithms. Their combination allows covering, in a very short time, a large number of cognitive domains, even despite the complexity and variability found across dementia subtypes.

4.2. Contribution of anatomical metrics

The contribution of brain atrophy measures to patient discrimination proved inconsistent. Although they were not as informative as cognitive screenings in the within-country analyses, classification of bvFTD reached its peak upon their inclusion as a feature (0.92, when age was not considered). The same was true for these patients from Country-2 in the cross-country analysis, in which the combination of atrophy values and IFS scores yielded the highest discrimination accuracy (0.91). However, these data proved mostly irrelevant in every other analysis, especially for identifying AD patients in the within-country analysis and both pathological groups in the cross-country analyses. In all of these, exclusion of atrophy values did not affect the discrimination of patients from controls. In the case of the within-country analysis, this may be due to the large association found between atrophy values and the cognitive screening scores in AD (see Figure 2), which is in line with previous findings in these patients [84, 85]. Thus, given their co-linear association, the novel

information afforded by atrophy values might, in some cases, prove marginal for classification compared to the cognitive tools.

Even though this explanation might also apply to the cross-country findings, inter-center variability in MRI equipment and acquisition parameters needs to be considered, too. Between the training (Country-1) and the testing (Country-2, and -3) samples, there are several differences regarding scanner models of the equipment (Philip Intera in Country-1, and Philip Achieva in Country -2 and -3), magnetic field intensity (1.5 Tesla for Country-1, and 3 Tesla for Country-2 and -3), and the parameters used in each center for the 3D T1 sequence. As previously shown, these differences may affect the consistency of MRI sequences across centers for classification analysis [41] – however, there are studies showing that variability across centers is relatively low and comparable [42, 86]. Thus, the higher variance of neuroimaging data compared to cognitive screenings may have undermined its predictive power for independent and unseen cohorts. Hence, the standardized nature of cognitive instruments [27, 28] may thus represent a clear advantage for cross-country validation protocols. Yet, despite the mixed neuroanatomical results, these features yielded discrimination values similar to previous reports based on MRI images when cognitive values were not included in the cross-country analysis [21, 41, 60, 74-76]. Moreover, the contribution of atrophy results was highlighted by specificity outcomes, which showed that its removal affects these values the most, especially for bvFTD patients from the cross-country analysis. Therefore, this highlights the relevance of neuroanatomical features for discriminating dementia patients, underscoring the inclusion of neuroimaging automatized methods as potential complementary tools for clinical settings.

4.3. Relevance of multimodal machine-learning approach

Although further studies comparing our machine-learning approach with other data-driven and automatic strategies are needed, our findings represent a potential milestone regarding the clinical implementation of machine-learning algorithms. Currently, timely detection of bvFTD and AD involves several challenges: varying levels of expertise and training from clinicians, non-systematic confirmation from clinical routine-MRI via visual inspection [87-89], variability of clinical and atrophy patterns, a certain degree of subjective interpretation and evaluation of signs and symptoms [10, 90], and strong variability of these factors across countries and centers. Against this framework, our approach underscores the reliability and predictive power of cognitive screenings and quantitative anatomical measures. Their combination yielded high classification rates for both conditions (bvFTD and AD). In addition, these measures showed great generalization power, indicating that they were able to precisely identify whether a new and unseen participant belongs to a given pathological group. Given the complexity and multi-level nature of alterations on the neurodegenerative process [5, 15-17], it is not expected that only one type of biomarker could be

enough to highly discriminate patients [60, 91]. Our findings support this view, as they testify to the relevance of combining cognitive screening and atrophy measures for the discrimination of these dementias.

On the other hand, these features also showed consistency to face within- and cross-country variability. The socio-cultural heterogeneity of our Latin American and Anglo-Saxon participants was further marked by divergences in equipment and acquisition parameters. Thus, our high classification results suggest that cognitive screening and atrophy measures are robust against the variability that characterized individualized clinical assessments. This is an essential characteristic of a potential early biomarker [13, 14], as it might reflect sensitivity to potentially universal properties of each condition.

Regarding the machine-learning algorithm applied, we used a very simple but powerful one (namely, logistic regression) to test the validity of the cognitive screenings and MRI information, which yielded high discrimination rates (>90) that were consistent and reliable in the context of a cross-center validation approach. We did not attempt any regularization procedure (we used the default parameters of the model) because we did not face any overfitting problem. Moreover, given our findings, we did not perform a direct comparison with other machine-learning algorithms because using a simple and fast one with its default parameters helps to promote its generalization and highlight its potential scalability. Further studies might compare the performance of different algorithms, and test their generalization power but considering a trade-off between the cost and benefit of each model.

Biomarkers should be low-cost, affordable, and massively applicable[13, 14], which is especially relevant for developing countries given their minimal mental health infrastructure, and lack of standardized diagnostic procedures[10]. Against this background, cognitive screenings emerge as potential candidates given that they are cost-effective, quick (they are completed in 10-15 minutes), easy to implement and learn for clinicians, and, hence, broadly applicable in primary care levels [24, 25, 27, 28]. Moreover, the application of similar digital version of these tasks, such as the Cambridge Neuropsychological Test Automated Battery (CANTAB, https://www.cambridgecognition.com/cantab) [92-95] (offering an automatized platform for administration and scoring) would allow for more efficient and faster transferring of clinical data to a machine-learning model already implemented, leading to a comprehensive report with the results of the model. Also, structural MRI is a non-invasive method, usually included as a routine exam for dementias, and it proves less time-consuming than other neuroimaging modalities (such as positron emission tomography, functional, and diffusion-weighted MR imaging [21, 37, 60, 91, 96]). Although its availability in developing countries is limited compared to high-income countries[10], the implementation of quantitative analysis of MRI data can be beneficial for patients who have access

to a more complete medical coverage, and especially for those whose cognitive screenings and clinical evaluation yield inconclusive results.

Finally, our study showed the potential translational relevance of automatic image quantification methods that are sensitive to subtle brain alterations which escape the naked eye or even traditional univariate methods [33]. In addition, our results also underscore the potential clinical implementation of computerized decision-support approaches (such as machine-learning algorithms) given that they allow characterizations at the individual level, which could be useful for diagnosis and treatment decisions [33].

4.4 Limitations and future directions

First, each sample had a moderate size; yet, similar (and smaller) sizes have been used in previous works [24, 42, 73], and the consistency of our results suggests that they were not biased by power issues. Second, the patients' diagnosis was based on clinical evaluations without pathological/genetic confirmation. However, this approach is similar to previous studies [35-40, 73] yielding compatible results. Moreover, the research centers from this work are specialized in the diagnosis, treatment and study of dementia, and they followed validated protocols and diagnostic guidelines (combining clinical information, neuroimaging data, and neuropsychological assessments). Third, from a technical viewpoint, our hypothesis-driven approach for estimating atrophy metrics could miss relevant information that is outside the predefined mask –especially compared to a whole-brain data driven approach. Yet, atrophy regions were selected based on robust evidence[15, 43, 62-64, 97, 98], and our procedure avoids bias regarding feature selection and allows testing the generalization power of atrophy levels in independent cohorts [60]. Future studies should compare the performance of our machine-learning pipeline with one employing a data-driven feature selection strategy. Fourth, given the absence of AD patients from Country-2, we were not able to perform a cross-validation analysis between dementia subtypes. However, given that our study was based on cognitive screenings and main atrophy areas of AD and bvFTD, our goal was to test their generalization power to discriminate patients from healthy controls given its feasibility to be implemented in different contexts (for example, both high- or low-income countries). Nevertheless, future research should test whether this and similar features could be used to discriminate different dementia subtypes based on a cross-center validation approach. In this sense, regarding the potential clinical application of our approach, future studies should: (i) check for inter-relations between classification results and each patient's functional severity, progression, and response to rehabilitation therapy; (ii) include functional connectivity measures, which have been proposed as potential biomarkers for dementia [99, 100]; (iii) evaluate whether digital cognitive tasks (such as the CANTAB) also afford robust and reliable markers that generalize to new and unseen data given their advantages over traditional pen-and-paper screening tasks

(automatization of administration and scoring); and (iv) be tested in pre-symptomatic patients to search for markers in prodromal disease stages, and also in the comparison between different subtypes of dementias.

5. CONCLUSION

Our study is the first to use machine-learning algorithms to show the high classification rates (> 0.91) obtained from the combination of cognitive screenings and quantitative neuroanatomical measures for identifying bvFTD and AD patients across three countries. Moreover, our results presented a robust generalization power (> 0.91), validated with two independent samples from different countries, which underscores the reliability of these measures to new, unseen data from heterogeneous contexts. Therefore, although further research is needed, our work supports the implementation of computer-based methods combining these measures as a potential affordable and complementary tool with clinical value for individual diagnosis and treatment decisions.

Acknowledgments

JKO thanks Valeria Pattacini from the Office of International Relations of Universidad de San Martín, UNSAM, (Argentina) for facilitating his visit and the UNSAM's hospitality. The authors are grateful to Marcin Ochab for valuable discussions. We thank the participants and their families for being involved in this research.

Funding sources

This work was supported by the Jagellonian University-UNSAM Cooperation Agreement, as well as the CEUNIM-INCYT-CEMSC^3 Collaboration Agreement. JKO was supported by the Grant DEC-2015/17/D/ST2/03492 of the National Science Centre (Poland). DRC was supported in part by CONICET (Argentina) and Escuela de Ciencia y Tecnología, UNSAM. AI is supported by grants from CONICET; CONICYT/FONDECYT Regular (1170010); FONDAP 15150012; the Inter-American Development Bank (IDB); PICT, Grant/ Award Number: 2017-1818 and 2017-1820; the INECO Foundation, and by the National Institute On Aging of the National Institutes of Health under Award Number R01AG057234. PR and DM are supported by COLCIENCIAS grant 697-2014. JF is supported by COLCIENCIAS grant 110674455314. This work was also supported in part by funding to Forefront, a collaborative research group specialized in the study of frontotemporal dementia and motor neurone disease, from the National Health and Medical Research Council (NHMRC) of Australia program grant (APP1037746) and the Australian Research Council (ARC) Centre of Excellence in Cognition and its Disorders Memory Program (CE110001021). FK is supported by an NHMRC-ARC Dementia Research Development

Fellowship (APP1097026). OP is supported by an NHMRC Senior Research Fellowship (APP1103258).

REFERENCES

1. Shah H, Albanese E, Duggan C, et al. Research priorities to reduce the global burden of dementia by 2025. The Lancet Neurology. 2016;15(12):1285-94.
2. Shaw LM, Korecka M, Clark CM, et al. Biomarkers of neurodegeneration for diagnosis and monitoring therapeutics. Nature reviews Drug discovery. 2007;6(4):295-303.
3. International AsD. World Alzheimer report 2015: the global impact of dementia.2015.
4. Oxtoby NP, Alexander DC, Euro Pc. Imaging plus X: multimodal models of neurodegenerative disease. Current opinion in neurology. 2017;30(4):371-9.
5. Palop JJ, Chin J, Mucke L. A network dysfunction perspective on neurodegenerative diseases. Nature. 2006;443(7113):768-73.
6. Forman MS, Farmer J, Johnson JK, et al. Frontotemporal dementia: clinicopathological correlations. Annals of neurology. 2006;59(6):952-62.
7. Tong T, Ledig C, Guerrero R, et al. Five-class differential diagnostics of neurodegenerative diseases using random undersampling boosting. NeuroImage Clinical. 2017;15:613-24.
8. Johnson JK, Head E, Kim R, et al. Clinical and pathological evidence for a frontal variant of Alzheimer disease. Archives of neurology. 1999;56(10):1233-9.
9. Padovani A, Premi E, Pilotto A, et al. Overlap between frontotemporal dementia and Alzheimer's disease: cerebrospinal fluid pattern and neuroimaging study. Journal of Alzheimer's disease : JAD. 2013;36(1):49-55.
10. Parra MA, Baez S, Allegri R, et al. Dementia in Latin America: Assessing the present and envisioning the future. Neurology. 2018;90(5):222-31.
11. Arbabshirani MR, Plis S, Sui J, Calhoun VD. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. NeuroImage. 2017;145(Pt B):137-65.
12. Huys QJ, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. Nature neuroscience. 2016;19(3):404-13.
13. Henley SM, Bates GP, Tabrizi SJ. Biomarkers for neurodegenerative diseases. Current opinion in neurology. 2005;18(6):698-705.
14. Humpel C. Identifying and validating biomarkers for Alzheimer's disease. Trends in biotechnology. 2011;29(1):26-32.
15. Piguet O, Hornberger M, Mioshi E, Hodges JR. Behavioural-variant frontotemporal dementia: diagnosis, clinical staging, and management. The Lancet Neurology. 2011;10(2):162-72.
16. Seelaar H, Rohrer JD, Pijnenburg YA, et al. Clinical, genetic and pathological heterogeneity of frontotemporal dementia: a review. Journal of neurology, neurosurgery, and psychiatry. 2011;82(5):476-86.
17. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2011;7(3):280-92.
18. Whitwell JL, Przybelski SA, Weigand SD, et al. Distinct anatomical subtypes of the behavioural variant of frontotemporal dementia: a cluster analysis study. Brain : a journal of neurology. 2009;132(Pt 11):2932-46.
19. Dottori M, Sedeno L, Martorell Caro M, et al. Towards affordable biomarkers of frontotemporal dementia: A classification study via network's information sharing. Scientific reports. 2017;7(1):3822.
20. Fox NC, Schott JM. Imaging cerebral atrophy: normal ageing to Alzheimer's disease. Lancet. 2004;363(9406):392-4.
21. Kloppel S, Stonnington CM, Chu C, et al. Automatic classification of MR scans in Alzheimer's disease. Brain : a journal of neurology. 2008;131(Pt 3):681-9.

22. Larner AJ, Mitchell AJ. A meta-analysis of the accuracy of the Addenbrooke's Cognitive Examination (ACE) and the Addenbrooke's Cognitive Examination-Revised (ACE-R) in the detection of dementia. International psychogeriatrics. 2014;26(4):555-63.
23. Moreira HS, Costa AS, Castro SL, et al. Assessing Executive Dysfunction in Neurodegenerative Disorders: A Critical Review of Brief Neuropsychological Tools. Frontiers in aging neuroscience. 2017;9:369.
24. Crawford S, Whitnall L, Robertson J, Evans JJ. A systematic review of the accuracy and clinical utility of the Addenbrooke's Cognitive Examination and the Addenbrooke's Cognitive Examination-Revised in the diagnosis of dementia. International journal of geriatric psychiatry. 2012;27(7):659-69.
25. Galton CJ, Erzinclioglu S, Sahakian BJ, et al. A comparison of the Addenbrooke's Cognitive Examination (ACE), conventional neuropsychological assessment, and simple MRI-based medial temporal lobe evaluation in the early diagnosis of Alzheimer's disease. Cognitive and behavioral neurology : official journal of the Society for Behavioral and Cognitive Neurology. 2005;18(3):144-50.
26. Velayudhan L, Ryu SH, Raczek M, et al. Review of brief cognitive tests for patients with suspected dementia. International psychogeriatrics. 2014;26(8):1247-62.
27. Hsieh S, Schubert S, Hoon C, et al. Validation of the Addenbrooke's Cognitive Examination III in frontotemporal dementia and Alzheimer's disease. Dementia and geriatric cognitive disorders. 2013;36(3-4):242-50.
28. Torralva T, Roca M, Gleichgerrcht E, et al. INECO Frontal Screening (IFS): a brief, sensitive, and specific tool to assess executive functions in dementia. Journal of the International Neuropsychological Society : JINS. 2009;15(5):777-86.
29. Bahia VS, Cecchini MA, Cassimiro L, et al. The Accuracy of INECO Frontal Screening in the Diagnosis of Executive Dysfunction in Frontotemporal Dementia and Alzheimer Disease. Alzheimer disease and associated disorders. 2018.
30. Custodio N, Herrera-Perez E, Lira D, et al. Evaluation of the INECO Frontal Screening and the Frontal Assessment Battery in Peruvian patients with Alzheimer's disease and behavioral variant Frontotemporal dementia. eNeurologicalSci. 2016;5:25-9.
31. Moreira HS, Lima CF, Vicente SG. Examining Executive Dysfunction with the Institute of Cognitive Neurology (INECO) Frontal Screening (IFS): normative values from a healthy sample and clinical utility in Alzheimer's disease. Journal of Alzheimer's disease : JAD. 2014;42(1):261-73.
32. Mueller SG, Schuff N, Weiner MW. Evaluation of treatment effects in Alzheimer's and other neurodegenerative diseases by MRI and MRS. NMR in biomedicine. 2006;19(6):655-68.
33. Orru G, Pettersson-Yeo W, Marquand AF, et al. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neuroscience and biobehavioral reviews. 2012;36(4):1140-52.
34. Zheng C, Xia Y, Pan Y, Chen J. Automated identification of dementia using medical imaging: a survey from a pattern classification perspective. Brain informatics. 2016;3(1):17-27.
35. Bron EE, Smits M, Papma JM, et al. Multiparametric computer-aided differential diagnosis of Alzheimer's disease and frontotemporal dementia using structural and advanced MRI. European radiology. 2017;27(8):3372-82.
36. Zhang Y, Schuff N, Camacho M, et al. MRI markers for mild cognitive impairment: comparisons between white matter integrity and gray matter volume measurements. PloS one. 2013;8(6):e66367.
37. Dukart J, Mueller K, Horstmann A, et al. Combined evaluation of FDG-PET and MRI improves detection and differentiation of dementia. PLoS One. 2011;6(3):e18111.
38. Kuceyeski A, Zhang Y, Raj A. Linking white matter integrity loss to associated cortical regions using structural connectivity information in Alzheimer's disease and fronto-temporal dementia: the Loss in Connectivity (LoCo) score. NeuroImage. 2012;61(4):1311-23.
39. Tahmasian M, Shao J, Meng C, et al. Based on the network degeneration hypothesis: separating individual patients with different neurodegenerative syndromes in a preliminary hybrid PET/MR study. Journal of Nuclear Medicine. 2016;57(3):410-5.
40. Zhou J, Greicius MD, Gennatas ED, et al. Divergent network connectivity changes in behavioural variant frontotemporal dementia and Alzheimer's disease. Brain : a journal of neurology. 2010;133(5):1352-67.

41. Abdulkadir A, Mortamet B, Vemuri P, et al. Effects of hardware heterogeneity on the performance of SVM Alzheimer's disease classifier. NeuroImage. 2011;58(3):785-92.
42. Sedeño L, Piguet O, Abrevaya S, a, et al. Tackling variability: A multicenter study to provide a gold-standard network approach for frontotemporal dementia. Human Brain Mapping. 2017.
43. Rascovsky K, Hodges JR, Knopman D, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain : a journal of neurology. 2011;134(Pt 9):2456-77.
44. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2011;7(3):263-9.
45. Baez S, Couto B, Torralva T, et al. Comparing moral judgments of patients with frontotemporal dementia and frontal stroke. JAMA neurology. 2014;71(9):1172-6.
46. Piguet O, Petersen A, Yin Ka Lam B, et al. Eating and hypothalamus changes in behavioral-variant frontotemporal dementia. Annals of neurology. 2011;69(2):312-9.
47. Torralva T, Roca M, Gleichgerrcht E, et al. A neuropsychological battery to detect specific executive and social cognitive impairments in early frontotemporal dementia. Brain : a journal of neurology. 2009;132(Pt 5):1299-309.
48. Mathuranath PS, Nestor PJ, Berrios GE, et al. A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. Neurology. 2000;55(11):1613-20.
49. Baez S, Kanske P, Matallana D, et al. Integration of Intention and Outcome for Moral Judgment in Frontotemporal Dementia: Brain Structural Signatures. Neuro-degenerative diseases. 2016;16(3-4):206-17.
50. Baez S, Manes F, Huepe D, et al. Primary empathy deficits in frontotemporal dementia. Frontiers in aging neuroscience. 2014;6:262.
51. Baez S, Morales JP, Slachevsky A, et al. Orbitofrontal and limbic signatures of empathic concern and intentional harm in the behavioral variant frontotemporal dementia. Cortex; a journal devoted to the study of the nervous system and behavior. 2016;75:20-32.
52. Baez S, Pinasco C, Roca M, et al. Brain structural correlates of executive and social cognition profiles in behavioral variant frontotemporal dementia and elderly bipolar disorder. Neuropsychologia. 2017.
53. Melloni M, Billeke P, Baez S, et al. Your perspective and my benefit: multiple lesion models of self-other integration strategies during social bargaining. Brain : a journal of neurology. 2016;139(11):3022-40.
54. Santamaria-Garcia H, Baez S, Reyes P, et al. A lesion model of envy and Schadenfreude: legal, deservingness and moral dimensions as revealed by neurodegeneration. Brain : a journal of neurology. 2017;140(12):3357-77.
55. Santamaria-Garcia H, Reyes P, Garcia A, et al. First Symptoms and Neurocognitive Correlates of Behavioral Variant Frontotemporal Dementia. Journal of Alzheimer's disease : JAD. 2016;54(3):957-70.
56. Sedeno L, Couto B, Garcia-Cordero I, et al. Brain Network Organization and Social Executive Performance in Frontotemporal Dementia. Journal of the International Neuropsychological Society : JINS. 2016;22(2):250-62.
57. Gleichgerrcht E, Roca M, Manes F, Torralva T. Comparing the clinical usefulness of the Institute of Cognitive Neurology (INECO) Frontal Screening (IFS) and the Frontal Assessment Battery (FAB) in frontotemporal dementia. Journal of clinical and experimental neuropsychology. 2011;33(9):997-1004.
58. Ibanez A, Cetkovich M, Petroni A, et al. The neural basis of decision-making and reward processing in adults with euthymic bipolar disorder or attention-deficit/hyperactivity disorder (ADHD). PLoS One. 2012;7(5):e37306.
59. Nichols TE, Das S, Eickhoff SB, et al. Best practices in data analysis and sharing in neuroimaging using MRI. Nature neuroscience. 2017;20(3):299-303.
60. Dukart J, Mueller K, Barthel H, et al. Meta-analysis based SVM classification enables accurate detection of Alzheimer's disease across different clinical centers using FDG-PET and MRI. Psychiatry research. 2013;212(3):230-6.

61. 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(1):273-89.
62. Ibanez A, Manes F. Contextual social cognition and the behavioral variant of frontotemporal dementia. Neurology. 2012;78(17):1354-62.
63. Schroeter ML, Raczka K, Neumann J, Yves von Cramon D. Towards a nosology for frontotemporal lobar degenerations-a meta-analysis involving 267 subjects. NeuroImage. 2007;36(3):497-510.
64. Du AT, Schuff N, Kramer JH, et al. Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia. Brain : a journal of neurology. 2007;130(Pt 4):1159-66.
65. Duda RO, Hart PE, Stork DG. Pattern classification. Wiley, editor2001.
66. Hastie T, Tibshirani R, Friedman J. An Introduction to Statistical Learning. 2nd ed: Springer-Verlag New York; 2009.
67. Kearns M, Ron D. Algorithmic stability and sanity-check bounds for leave-one-out cross-validation. Neural computation. 1999;11(6):1427-53.
68. Wang X, Ren P, Mapstone M, et al. Identify a shared neural circuit linking multiple neuropsychiatric symptoms with Alzheimer's pathology. Brain imaging and behavior. 2017.
69. Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters. 2006;27:861-74.
70. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837-45.
71. Gengsheng Q, Hotilovac L. Comparison of non-parametric confidence intervals for the area under the ROC curve of a continuous-scale diagnostic test. Statistical methods in medical research. 2008;17(2):207-21.
72. Cohen JD, Daw N, Engelhardt B, et al. Computational approaches to fMRI analysis. Nature neuroscience. 2017;20(3):304-13.
73. Salvatore C, Battista P, Castiglioni I. Frontiers for the Early Diagnosis of AD by Means of MRI Brain Imaging and Support Vector Machines. Current Alzheimer research. 2016;13(5):509-33.
74. Gerardin E, Chetelat G, Chupin M, et al. Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. NeuroImage. 2009;47(4):1476-86.
75. Varol E, Gaonkar B, Erus G, et al. Feature Ranking Based Nested Support Vector Machine Ensemble for Medical Image Classification. Proceedings IEEE International Symposium on Biomedical Imaging. 2012:146-9.
76. Yang W, Lui RL, Gao JH, et al. Independent component analysis-based classification of Alzheimer's disease MRI data. Journal of Alzheimer's disease : JAD. 2011;24(4):775-83.
77. Ye BS, Choi SH, Han SH, et al. Clinical and Neuropsychological Comparisons of Early-Onset Versus Late-Onset Frontotemporal Dementia: A CREDOS-FTD Study. Journal of Alzheimer's disease : JAD. 2015;45(2):599-608.
78. Yew B, Alladi S, Shailaja M, et al. Lost and forgotten? Orientation versus memory in Alzheimer's disease and frontotemporal dementia. Journal of Alzheimer's disease : JAD. 2013;33(2):473-81.
79. Noh Y, Jeon S, Lee JM, et al. Anatomical heterogeneity of Alzheimer disease: based on cortical thickness on MRIs. Neurology. 2014;83(21):1936-44.
80. Ossenkoppele R, Pijnenburg YA, Perry DC, et al. The behavioural/dysexecutive variant of Alzheimer's disease: clinical, neuroimaging and pathological features. Brain : a journal of neurology. 2015;138(Pt 9):2732-49.
81. Amieva H, Lafont S, Rouch-Leroyer I, et al. Evidencing inhibitory deficits in Alzheimer's disease through interference effects and shifting disabilities in the Stroop test. Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists. 2004;19(6):791-803.
82. Sgaramella TM, Borgo F, Mondini S, et al. Executive deficits appearing in the initial stage of Alzheimer's disease. Brain and cognition. 2001;46(1-2):264-8.

83. Jory JI, Bruna AA, Munoz-Neira C, Chonchol AS. Chilean version of the INECO Frontal Screening (IFS-Ch): psychometric properties and diagnostic accuracy. Dementia & neuropsychologia. 2013;7(1):40-7.
84. Canu E, Agosta F, Mandic-Stojmenovic G, et al. Multiparametric MRI to distinguish early onset Alzheimer's disease and behavioural variant of frontotemporal dementia. NeuroImage Clinical. 2017;15:428-38.
85. Sorensen L, Igel C, Liv Hansen N, et al. Early detection of Alzheimer's disease using MRI hippocampal texture. Hum Brain Mapp. 2016;37(3):1148-61.
86. Biswal BB, Mennes M, Zuo XN, et al. Toward discovery science of human brain function. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(10):4734-9.
87. Kloppel S, Abdulkadir A, Jack CR, Jr., et al. Diagnostic neuroimaging across diseases. NeuroImage. 2012;61(2):457-63.
88. Kloppel S, Stonnington CM, Barnes J, et al. Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. Brain : a journal of neurology. 2008;131(Pt 11):2969-74.
89. Koikkalainen J, Rhodius-Meester H, Tolonen A, et al. Differential diagnosis of neurodegenerative diseases using structural MRI data. NeuroImage Clinical. 2016;11:435-49.
90. Forman MS, Farmer J, Johnson JK, et al. Frontotemporal dementia: clinicopathological correlations. Annals of neurology. 2006;59(6):952-62.
91. McMillan CT, Avants BB, Cook P, et al. The power of neuroimaging biomarkers for screening frontotemporal dementia. Hum Brain Mapp. 2014;35(9):4827-40.
92. Barnett JH, Blackwell AD, Sahakian BJ, Robbins TW. The Paired Associates Learning (PAL) Test: 30 Years of CANTAB Translational Neuroscience from Laboratory to Bedside in Dementia Research. Current topics in behavioral neurosciences. 2016;28:449-74.
93. Giedraitiene N, Kaubrys G. Distinctive Pattern of Cognitive Disorders During Multiple Sclerosis Relapse and Recovery Based on Computerized CANTAB Tests. Frontiers in neurology. 2019;10:572.
94. Janssen G, van Aken L, De Mey H, et al. Decline of executive function in a clinical population: age, psychopathology, and test performance on the Cambridge Neuropsychological Test Automated Battery (CANTAB). Applied neuropsychology Adult. 2014;21(3):210-9.
95. Smith PJ, Need AC, Cirulli ET, et al. A comparison of the Cambridge Automated Neuropsychological Test Battery (CANTAB) with "traditional" neuropsychological testing instruments. Journal of clinical and experimental neuropsychology. 2013;35(3):319-28.
96. Moller C, Pijnenburg YA, van der Flier WM, et al. Alzheimer Disease and Behavioral Variant Frontotemporal Dementia: Automatic Classification Based on Cortical Atrophy for Single-Subject Diagnosis. Radiology. 2016;279(3):838-48.
97. Whitwell JL, Jack CR, Jr., Przybelski SA, et al. Temporoparietal atrophy: a marker of AD pathology independent of clinical diagnosis. Neurobiology of aging. 2011;32(9):1531-41.
98. Pini L, Pievani M, Bocchetta M, et al. Brain atrophy in Alzheimer's Disease and aging. Ageing research reviews. 2016;30:25-48.
99. Pievani M, de Haan W, Wu T, et al. Functional network disruption in the degenerative dementias. The Lancet Neurology. 2011;10(9):829-43.
100. Pievani M, Filippini N, van den Heuvel MP, et al. Brain connectivity in neurodegenerative diseases--from phenotype to proteinopathy. Nature reviews Neurology. 2014;10(11):620-33.