Journal of Alzheimer's Disease 65 (2018) 819–842
DOI 10.3233/JAD-170514
IOS Press

Alzheimer's Disease Computer-Aided Diagnosis: Histogram-Based Analysis of Regional MRI Volumes for Feature Selection and Classification

Elena Ruiz[a,*], Javier Ramírez[b], Juan Manuel Górriz[b] and Jorge Casillas[a] for the Alzheimer's Disease Neuroimaging Initiative[1]
[a]*Department of Computer Science and Artificial Intelligence; DaSCI (Data Science and Computational Intelligence Research Institute); University of Granada, Granada, Spain*
[b]*Department of Signal Theory, Networking and Communications; DaSCI (Data Science and Computational Intelligence Research Institute); University of Granada, Granada, Spain*

Accepted 6 May 2018

Abstract. This paper proposes a novel fully automatic computer-aided diagnosis (CAD) system for the early detection of Alzheimer's disease (AD) based on supervised machine learning methods. The novelty of the approach, which is based on histogram analysis, is twofold: 1) a feature extraction process that aims to detect differences in brain regions of interest (ROIs) relevant for the recognition of subjects with AD and 2) an original greedy algorithm that predicts the severity of the effects of AD on these regions. This algorithm takes account of the progressive nature of AD that affects the brain structure with different levels of severity, i.e., the loss of gray matter in AD is found first in memory-related areas of the brain such as the hippocampus. Moreover, the proposed feature extraction process generates a reduced set of attributes which allows the use of general-purpose classification machine learning algorithms. In particular, the proposed feature extraction approach assesses the ROI image separability between classes in order to identify the ones with greater discriminant power. These regions will have the highest influence in the classification decision at the final stage. Several experiments were carried out on segmented magnetic resonance images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) in order to show the benefits of the overall method. The proposed CAD system achieved competitive classification results in a highly efficient and straightforward way.

Keywords: Alzheimer's disease, Alzheimer's Disease Neuroimaging Initiative, classification, computer aided diagnosis, computer-assisted, histogram-based analysis, image processing, MRI, supervised learning

INTRODUCTION

Alzheimer's disease (AD) is the most common cause of dementia among elderly people, affecting 30 million people worldwide. The increasing life expectancy and the aging of the population in developed countries suggests that AD will affect 60 million people worldwide in the next 50 years [1]. This kind of disease has a big impact on today's society, as it involves not only health services, but also a conglomerate of social, psychological, and family care issues. Therefore, neurodegenerative diseases, such as AD,

[1]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

*Correspondence to: Elena Ruiz, Department of Computer Science and Articial Intelligence, University of Granada, E-18071, Granada, Spain. Tel.: +34 958241773; E-mail: eruiz@decsai.ugr.es.

are considered to be one of the biggest challenges of the XXI century. AD is a slow neurodegenerative disease associated with the production of amyloid-β peptide (Aβ) and its extracellular deposition as well as the flame-shaped neurofibrillary tangles of the microtubule binding protein tau [2]. The loss of nerve cells leads to symptoms usually starting with mild memory problems and turning into severe brain damage over the course of time. There is no cure for AD, and nowadays the developed drugs can only offer mild symptomatic benefits but are otherwise palliative [3]. Any approach oriented to anticipate and improve the diagnosis is desirable as early diagnosis facilitates effective treatment.

The emergence and improvement of new imaging technologies, such as magnetic resonance imaging (MRI), electroencefalography (EEG), functional MRI (fMRI), or positron emission tomography (PET) have caused a breakthrough in the procedure of diagnosis. Capturing the anatomical variability caused by AD in the brain structure MRI appears to be an important tool in the AD classification. Beyond manual processing of the images, a trend in the usage of computer-aided systems has been developed to assist in the image pre-processing [4, 5], to perform differential diagnosis [6–8] and even to predict the conversion from prodromal stages such as subjective memory complaints (SMC) [9, 10] or mild cognitive impairment (MCI) to AD [11–13]. In particular, since the AD neurodegeneration progressively affects different brain functions, tomographic functional images, single emission computerized tomography (SPECT) [14–16] or positron emission tomography (PET) [17, 18] have been extensively used in computer aided diagnosis (CAD) systems. MRI has been recently used in many studies for CAD-based automatic diagnosis reaching accuracies up to 90% in AD versus normal controls (NC) [19–22]. Nevertheless, it is worth mentioning that in [19–22] different subjects are assessed within a different set of cross-validation folds and even when they are matched, the cross-validation partitions are seldom the same, thus it is hard to make a fair comparison.

Most of the approaches for brain image processing, such as the one proposed in [23–26], are based on univariate analysis. This kind of analysis is performed using one single feature at a time and frequently consists of a voxel-wise comparison between intensity levels, either over the images themselves (voxel based morphometry, or VBM [27, 28]), regions of interest (ROIs) [29] or models of brain features, such as cortical thickness [30] or intrinsic curvature [31]. These methodologies have been implemented in the most prominent pieces of software used in clinical practice, such as FreeSurfer or SPM [32], and they are used the most in epidemiological studies. Other approaches define weak classifiers on small enough regions [33, 34]. Specifically, [33] use an ensemble of sparse representation classifiers (SRC) defined on equally-sized patches extracted from the gray matter (GM) image. By contrast, Savio and Graña [34] use an ensemble of support vector machines (SVM) to separately classify each area defined by the Automated Anatomical Labelling Atlas (AAL). Despite showing good classification results, both approaches present different drawbacks. In [33] the classification is performed using voxel intensity values instead of computing discriminative features. Therefore, it shares the *curse of dimensionality problem*. There are several studies that show clear advantages of using a reduced number of discriminative features, such as eigenbrains-based methods [16], multivariate Gaussian methods [15, 18], codebook based methods [19] or SVM-based methods [35]. Savio and Graña [34] extract some first order statistics from each brain area to be used as features, and do not consider the spatial relationship among voxels. Both these methods [33, 34] use supervised learning for both, computing the statistical relevance of each brain region and training the classifier, which could be a problem whenever not all the training samples are labelled, or the labels are not reliable enough to use them as ground truth. This is a relatively common problem in the labelling process in medical imaging. In the ADNI dataset, image labels are mainly assigned from the Mini-Mental State Examination (MMSE) score.

There is an increasing interest in multivariate approaches. In contrast to univariate analysis, these can effectively handle information that simultaneously affects the whole brain as well as characterizing the relationship between different ROIs. Brain networks [36, 37], texture features [20, 38–41] and other voxel and region-wise higher-level features [42–46] could therefore reveal other information than only volumetric, complementing and providing new insights into the disease. In [47] a new framework called spherical brain mapping (SBM) was proposed. It performs feature extraction in MR Brain Images by reducing the whole images to bidimensional maps comprising a number of statistical and morphological measures. Each pixel in the latter maps was the result of computing a certain measure on a set of voxels crossed by the mapping vector, centered at the Anterior Commissure (AC) and spanning over

all values of azimuth and elevation angles. The bidimensional maps were related to anatomical changes such as brain atrophy or cortical thickness, yielding high performance in differential diagnosis. Furthermore, they provided a significant feature reduction, as well as a visual representation of the underlying information. In [48] an extension to the framework using texture and volumetric Local Binary Patterns (LBP) used in [39], is proposed. Authors of [49] propose a new path tracing algorithm based on Hidden Markov Models (HMM). This method aims at enhancing the mapping procedure in SBM. It replaces the mapping vector with curvilinear paths that adapt to the structural information present in MRI. This goal is achieved by the computation of the feature maps and the direct use of the intensity distribution along the path, as a characterization of the structural differences in normal or AD-affected subjects. Since the gray level co-occurrence matrix (GLCM), firstly developed by Haralick [50], has been used in brain characterization in [41, 45], an extension of this strategy is proposed to characterize the brain texture along each path and its neighborhood. Finally, in [52] deep learning is employed in order to extract representative features from each brain area defined by the AAL atlas in an unsupervised manner. With this strategy we avoid the need for a ground truth at this stage. Different architectures and voting schemes are implemented and compared to define ensembles of deep belief networks (DBN). An alternative architecture where a SVM is used to fuse the DBN outcomes is also presented. Each unit in the ensemble is responsible not only for classifying the corresponding patch but also for extracting representative features for the different brain regions.

In this paper we propose a novel multivariate CAD system for AD. It uses histograms as a representation of the input MRI images and employs original dissimilarity measurements between those histograms as input attributes for machine learning algorithms. The machine learning algorithms build highly accurate models for the diagnosis of AD, being able to predict if an input MRI image belongs to a subject who has been labelled as patient with AD according to neuropsychological tests. This study has a dual aim: 1) to propose a method that defines a well structured dataset which can be used as input to classification algorithms, as well as 2) proposing a new algorithm to address this particular classification task. The steps proposed reflect the prototypical Data Mining process [53], involving: 1) Problem Specification; 2) Problem Understanding; 3) Data Preprocessing [54] (data preparation and reduction), we perform image preprocessing and feature extraction; 4) Data Mining, we choose classification as the most suitable Data Mining task and apply several machine learning algorithms; 5) Evaluation, we collect results from these machine learning algorithms and assess their performance based on appropriate and interesting measures; and 6) Results Exploitation, we report the discovered knowledge through visualization tools.

The remainder of this paper is organized as follows: First we describe the database of MRI images used, the image preprocessing, and the feature extraction applied. Next, we present the proposed greedy classifier and explain the methodology followed in the experiments. Then we show the results obtained applying the previously explained techniques. Finally, we discuss the primary contributions of this study, analyze the results obtained, provide interesting visual support, and detail future work.

DATASET DESCRIPTION AND PREPROCESSING

In this section we describe the database and explain the image preprocessing and feature extraction processes.

ADNI database

The images used in this paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (http://www.loni.ucla.edu/ADNI and http://www.adni-info.org). The ADNI database contains 1.5T and 3.0T t1w MRI scans for AD, MCI, and cognitively NC at several time points. All structural MR scans used in this paper were acquired from a 1.5T MRI scanner and this database provides data for three groups of patients: healthy patients (NC), Alzheimer disease patients (AD), and patients with mild cognitive symptoms (MCI). All subjects must be willing and able to undergo all test procedures including neuroimaging and agree to longitudinal follow-up. Specific psychoactive medications were excluded.

The dataset used in this study contains 1075 T1-weighted MRI images from 229 NC, 401 MCI (312 stable MCI and 86 progressive MCI), and 188 AD subjects. As only the first exam (Baseline) for each patient has been used in this study, 818 images were used for assessing the proposed approach. Demographic data of patients in the database is summarized in Table 1.

Table 1
Demographic data of patients in the database (ADNI 1075-T1)

Diagnosis	Number	Age	Gender (M/F)	MMSE
NC	229	75.97±5.0	119/110	29.00±1.0
MCI	401	74.85±7.4	258/143	27.01±1.8
AD	188	75.36±7.5	99/89	23.28±2.0

Image preprocessing

Raw magnetic resonance images from the ADNI database were spatially normalized in order to ensure that a given voxel corresponds to the same anatomical position in different subjects. Intensity normalization of images was also performed. After image registration, segmentation is performed using the voxel-based morphometry (VBM) toolbox [55] for Statistical Parametric Mapping (SPM v8) software [56]. For each patient, $121 \times 145 \times 121$ gray matter and white matter membership probability maps were obtained with a $1.5\,\text{mm} \times 1.5\,\text{mm} \times 1.5\,\text{mm}$ voxel size.

Feature extraction

Using all the voxels of the probability maps as direct input to classification algorithms would fall into curse of dimensionality problem in many cases. Thus, we use an atlas to classify voxels according to regions of interest. We have selected the atlas of the AAL toolbox for SPM that includes an anatomical parcellation of 116 regions of the spatially normalized single subject high resolution T1 volume provided by the Montreal Neurological Institute (MNI) atlas [57]. The cerebral region definitions are fully described in [58] and are based on the cerebellum parcellation proposed by Schmahmann et al [59]. The parcellation procedure is described as follows. T1-weighted MRIs are normalized to MNI space and segmented into three different brain tissues: cerebral spinal fluid, gray matter and white matter. Each individual tissue voxel is labeled with the MNI anatomical atlas that was previously corregistered with the template images.

Classification input attributes are going to depend on the distance between the test sample and each class mean for each region of interest, following an idea similar to T-test feature selection. To discard possible outliers, we vectorize each ROI, i.e., form a vector with probability values of all the voxels in the region in all the maps. We compute the four quartiles ($Q1$, $Q2$, $Q3$, and $Q4$) and the interquartile distance (IQR):

$$IQR = Q3 - Q1 \tag{1}$$

for each vectorized region of interest. Therefore, voxel with a probability value greater than $Q3 + 1.5 \times IQR$ or lower than $Q1 - 1.5 \times IQR$ is classified as an *outlier*.

Afterwards, we compute the histogram (50 bins) of the values of the voxels of each ROI for each one of the probability maps. Each histogram is normalized according to the total number of voxels in the ROI with the purpose of avoiding that, in future steps, ROIs relevance may be biased in favour of the biggest ROIs. Then, we classify normalized histograms of training samples according to type of subject (NC, MCI, or AD) and ROI, and calculate the average histogram for each type of subject and ROI. We get 116 average histograms (one per ROI) for each type of subject. All the histograms of a certain ROI are based on the same set of bins (regardless of subject type).

As mentioned before, the idea is to quantify the distance between a certain input image (or probability map) and each type of training subjects' mean. The metric selected in the feature space is the bin-wise mean squared error distance d, defined as:

$$d(A, B) = \frac{\sum_{j=1}^{n} (A_j - B_j)^2}{n} \tag{2}$$

where A_j and B_j are the values of the j^{th} bin in A and B, respectively. We address three binary problems: 1) AD versus NC, 2) MCI versus NC, and 3) AD versus MCI.

For each input MRI image (test image) we compute the distance between the obtained GM histograms and the average histograms of each class computed in the training set. This distance is referred as $d_T(I) = d(I, T)$, where I is the histogram of the test individual image for a certain ROI and T is the average training histogram, template, for one class. To quantify the relative proximity of the i^{th} test sample to both classes in one numeric measure for the j^{th} ROI we propose the following index:

$$f_j(I_i) = 2 \cdot \frac{d_1(I_i)}{d_2(I_i) + d_1(I_i)} - 1 \tag{3}$$

where $d_1(I_i)$ and $d_2(I_i)$ are the distances computed between the ith test subject's histogram and the average training histograms of classes c_1 (positive class) and c_2 (negative class), respectively. Thus, for each jth ROI of the test image, there is a $f_j(I_i)$ whose value is in $[-1, +1]$, being closer to -1 as lower $d_1(I_i)$ is

in relation to $d_2(I_i)$ and closer to +1 as lower $d_2(I_i)$ is in relation to $d_1(I_i)$. That is, negative values mean the jth ROI is labeled as class c_1 (analogously, positive values label as class c_2), while the certainty of this labeling becomes higher as the value becomes closer to -1 (or +1).

Fig. 1 illustrates some examples of $d_1(I)$, $d_2(I)$, and $f_j(I)$ for two ROIs and three subjects, each one from a different class and binary problem. This figure shows that, depending on the ROI, a test subject can be closer to its true class or to the wrong one, and how $d_1(I)$, $d_2(I)$, and $f_j(I)$ change accordingly.

METHODOLOGY

Throughout this section we present a proposal of a greedy classifier and we detail how we design the classification experiments and which algorithms are included in them.

Greedy classifier

Not all brain regions suffer AD damages to the same grade. The proposed greedy algorithm[3] tries to incorporate this quality into the learning process by assigning a particular weight to each ROI which determines the influence of the region on the classification. The weight (w) is defined as follows:

$$w_j(p_j^P, p_j^N) = \frac{p_j^P + p_j^N}{2}, \tag{4}$$

where p_j^P represents the ratio of subjects of the positive class whose $f_j(I_i)$ is lower than zero and p_j^N the ratio of subjects of the negative class whose $f_j(I_i)$ is greater than zero for the jth ROI. Thus, those regions with the highest percentage of samples which are more similar to their own class mean than to the opposite class mean are the regions with the highest weights.

Once we have computed $w_j(p_j^P, p_j^N)$ for each ROI the weights are normalized with:

$$\hat{w}_j(p_j^P, p_j^N) = \frac{w_j - \min_k w_k}{\max_k w_k - \min_k w_k} \tag{5}$$

Given histogram I_i and percentages p_j^P and p_j^N of the test image for a certain ROI j, we compute $\overline{f}_{ij}$ as the product of $f_j(I_i)$ and $\widehat{w}_j(p_j^P, p_j^N)$:

$$\overline{f}_{ij} = f_j(I_i) \cdot \widehat{w}_j(p_j^P, p_j^N) \tag{6}$$

We compute eq. 1 for every ROI and obtain a set of m values $(\overline{f}_{i1}, \overline{f}_{i2}, \ldots, \overline{f}_{im})$ where m is the number of ROIs. Finally, we compute F_i as:

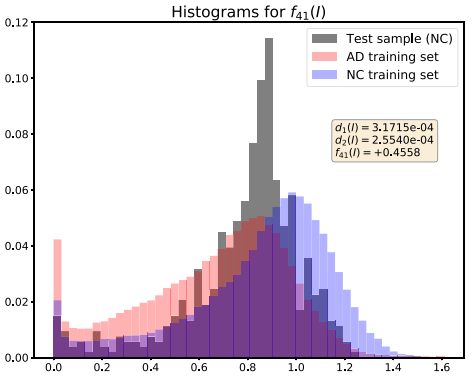
$$F_i = \frac{1}{m} \sum_{j=1}^{m} \overline{f}_{ij} \tag{7}$$

F_i is used to decide if the test subject is classified in the positive or negative class. Fig. 2 provides some visual examples of the relation between the distribution of $f_j(I_i)$ in the training set of subjects, p_j^P, p_j^N, $w_j(p_j^P, p_j^N)$ and $\widehat{w}_j(p_j^P, p_j^N)$. In this figure it is possible to observe the obvious relationship between $f_j(I_i)$ distribution, and p_j^P and p_j^N, as well as, the resulting effect on $w_j(p_j^P, p_j^N)$.

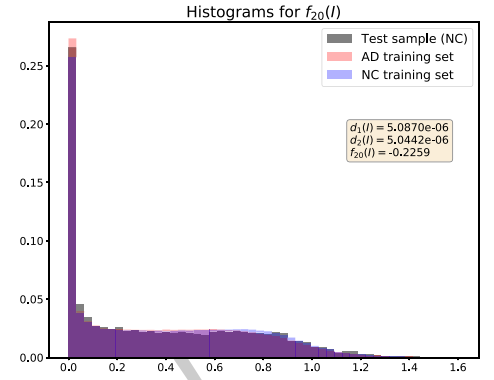
ROIs' weights have a dual aim. In addition to using them as a way to estimate the influence of each ROI on the classification process, we also use them as a feature selection criterion [53]. To reduce the dataset dimensions we remove irrelevant attributes until getting a minimum set of features with which the classification performance is as good as the one obtained using all features or better. We perform a series of experiments where in each experiment the r highest weighted ROIs are included, starting with $r = 1$ and adding one new ROI each time. At the end of each experiment, every sample has a predicted label l. According to the set of predicted labels and to the true class of each sample, we use accuracy and G-mean to assess the performance of the classification task. Figs. 3–5 show accuracy and G-mean for each experiment.

Fig. 3 is the clearest example of the fact that the best results can be obtained taking into account a reduced set of ROIs. Including the full set of input attributes turns out to be unhelpful or can even lead to overfitting. In light of these results, the significance of accomplishing a suitable feature selection to get the optimal results of this greedy classifier is clear. Taking this fact into account, we include a proper feature selection mechanism as part of the algorithm: the algorithm only includes in the classification process those ROIs in which both p_j^P and p_j^N surpass a certain threshold. Such threshold is a configurable input parameter. Only the $f_j(I_i)$ values coming from ROIs which satisfy this condition are included in the computation of F_i for the test samples.

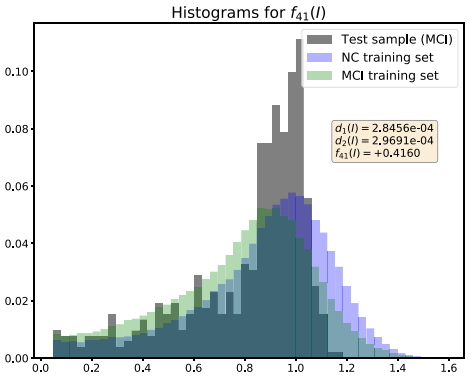
[3]The source code of this classifier and of the previous feature extraction technique is available at the following repository: (https://github.com/eruizsanchez/CAD_Alzheimers_Disease).



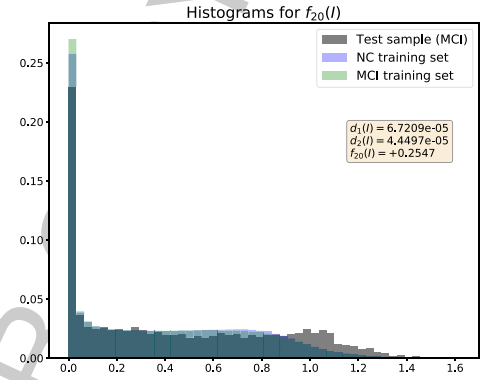
(a) ROI 41 ("Amygdala_L") - AD vs. NC



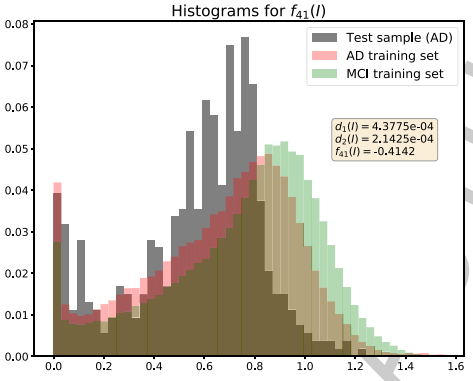
(b) ROI 20 ("Supp_Motor_Area_R") - AD vs. NC



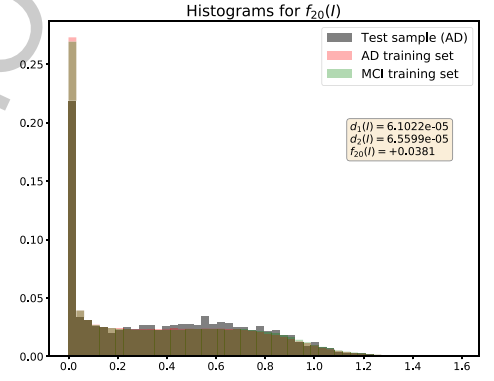
(c) ROI 41 ("Amygdala_L") - MCI vs. NC



(d) ROI 20 ("Supp_Motor_Area_R") - MCI vs. NC



(e) ROI 41 ("Amygdala_L") - AD vs. MCI



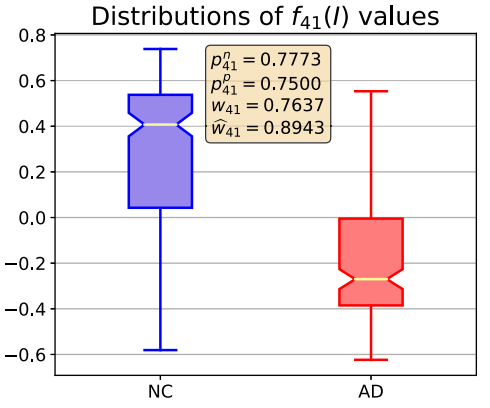
(f) ROI 20 ("Supp_Motor_Area_R") - AD vs. MCI

Fig. 1. Histograms for ROIs 41 ("Amygdala_L") and 20 ("Supp_Motor_Area_R") of a certain test subject (I) along with the average histograms (T) of each binary problem's classes: AD versus NC, MCI versus NC, and AD versus MCI. For each binary problem we choose a different test subject. Each graph includes $d_1(I)$, $d_2(I)$ and $f_j(I)$ computed based on the three correspondent histograms.

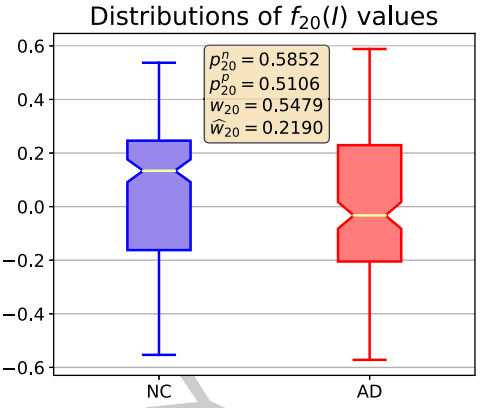
Classification

As mentioned before, we address three binary classification problems using $f(I)$ values for each region of interest as input features. Five machine learning classification algorithms are tested:

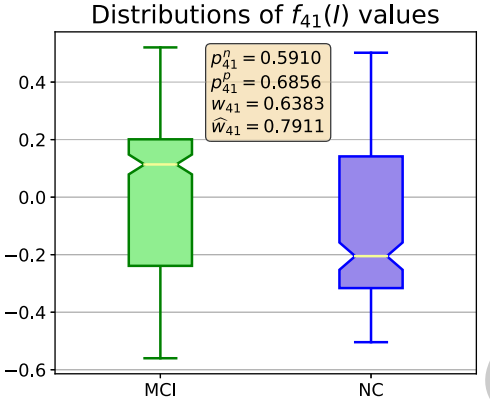
1. *C4.5* [60, 61]: generates a decision tree from a training dataset in the same way as ID3 does, using the concept of *information entropy*
2. *Random-Forest* [62] (RF): combines *bagging* and random attribute selection ideas to build a collection of k decision trees, k being a configurable



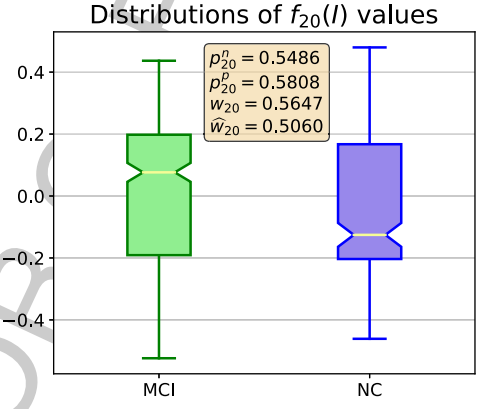
(a) ROI 41 ("Amygdala_L") - AD vs. NC



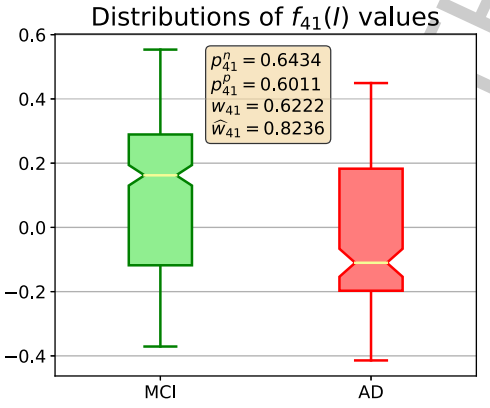
(b) ROI 20 ("Supp_Motor_Area_R") - AD vs. NC



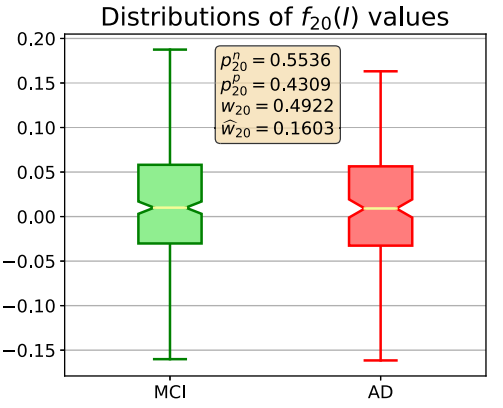
(c) ROI 41 ("Amygdala_L") - MCI vs. NC



(d) ROI 20 ("Supp_Motor_Area_R") - MCI vs. NC

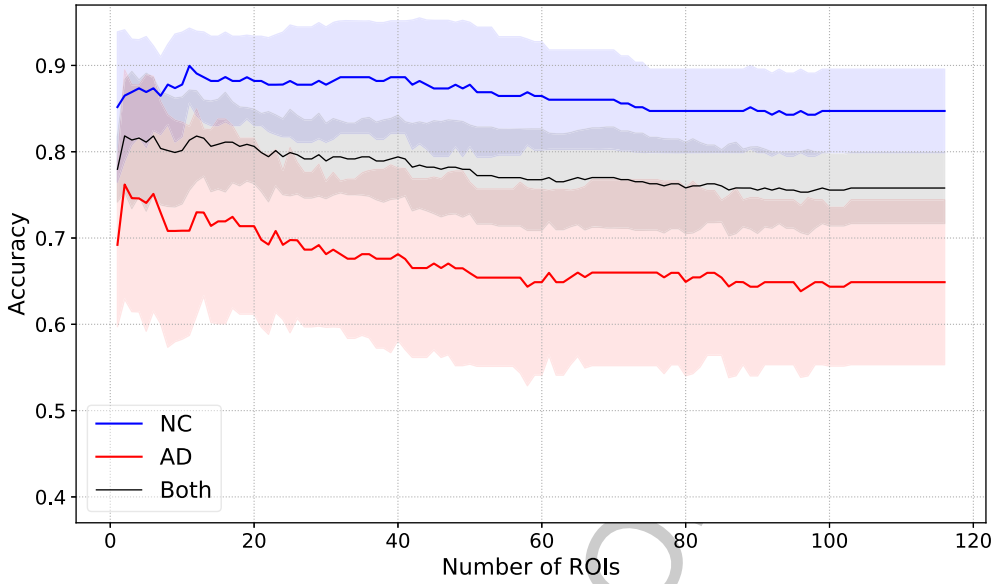


(e) ROI 41 ("Amygdala_L") - AD vs. MCI

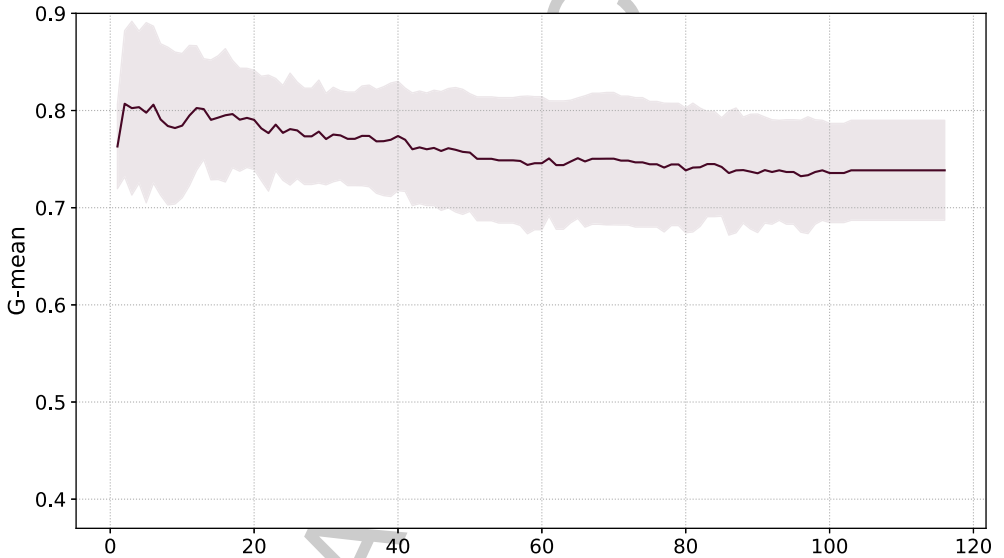


(f) ROI 20 ("Supp_Motor_Area_R") - AD vs. MCI

Fig. 2. Distribution of $f_j(I)$ per class for $j = 20$ ("Supp_Motor_Area_R") and $j = 41$ ("Amygdala_L") in the three binary problems: AD versus NC; MCI versus NC, and AD versus MCI. Plots also include p_j^P, p_j^N, $w_j(p_j^P, p_j^N)$ and $\widehat{w}_j(p_j^P, p_j^N)$.



(a) Accuracy - AD vs. NC

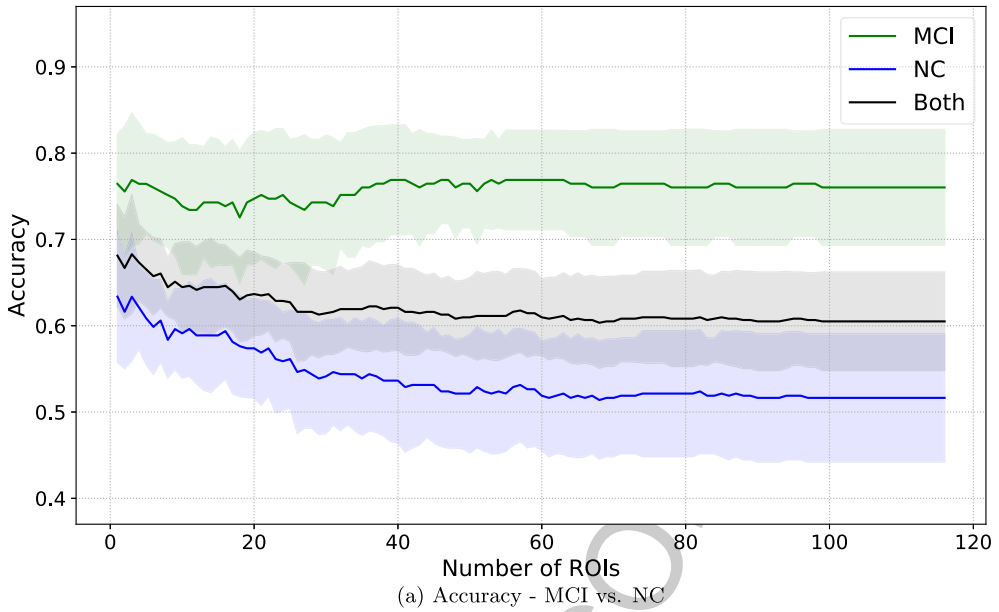


(b) G-mean - AD vs. NC

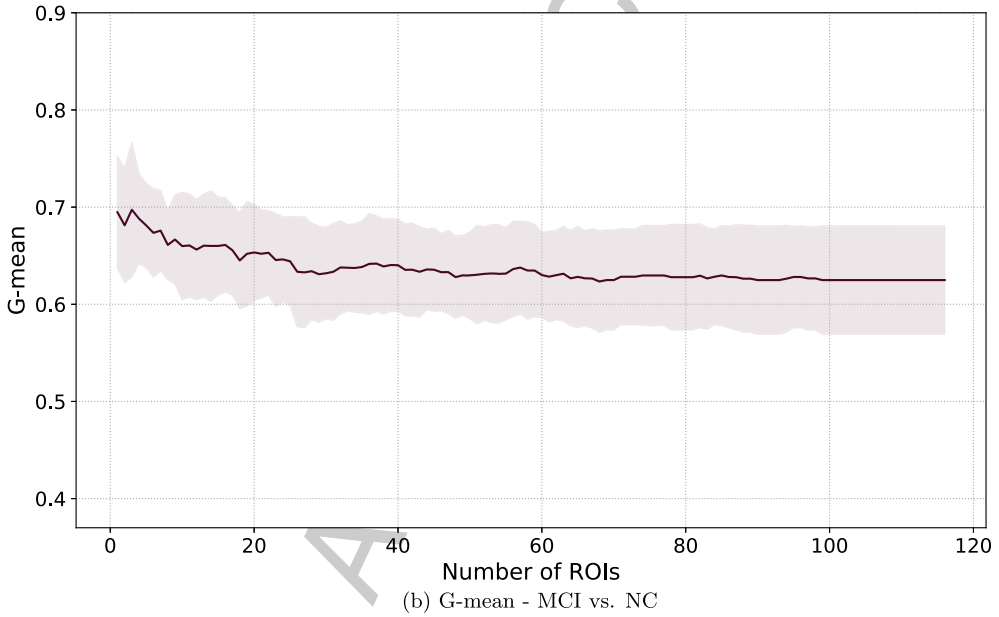
Fig. 3. Comparisons of Accuracy (a) and G-mean (b) registered using *10-fold cross-validation* to afford AD versus NC problem, selecting the n ROIs with higher weights (for n from 1 to 116). Shaded areas cover mean $\pm$ standard deviation.

input parameter ($k = 100$ is used in the experiments of which results are shown in Table 2)

3. *Support Vector Machine* [63] (SVM): generates a representation of the training examples as points in space, mapped so that the examples of the different classes are separated by a gap that is as wide as possible
4. *eXtreme Gradient Boosting* [64] (XGBoost): produces a prediction model in the form of an ensemble of weak prediction models. It builds the model in a stage-wise fashion and generalizes them by allowing optimization of an arbitrary differentiable loss function. XGBoost is applied with the following input parameters: logistic regression for classification as objective function, 1 as step size of each boosting step, 3 as maximum depth of the tree and 500 as maximum number of iterations



(a) Accuracy - MCI vs. NC

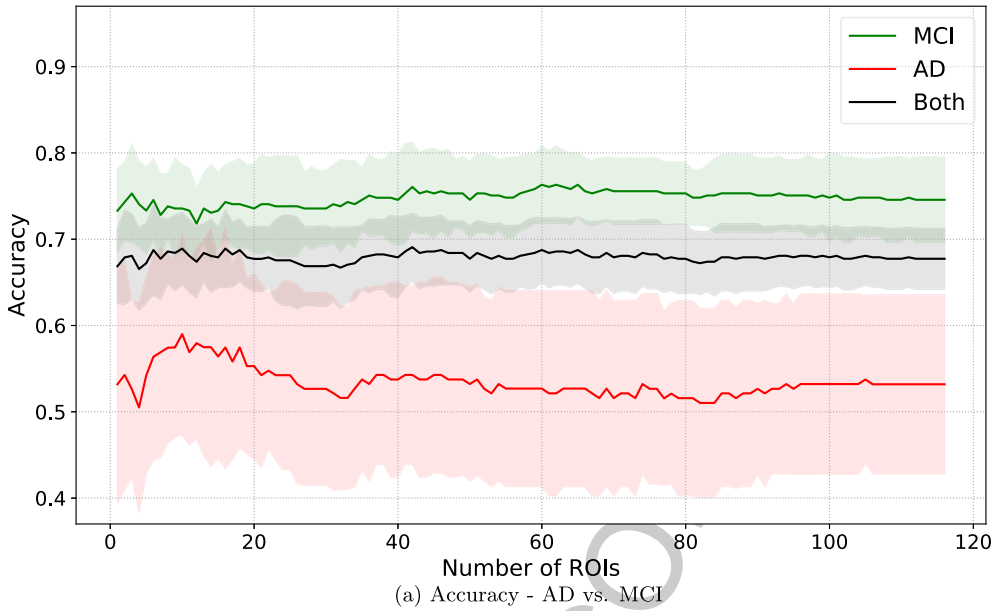


(b) G-mean - MCI vs. NC

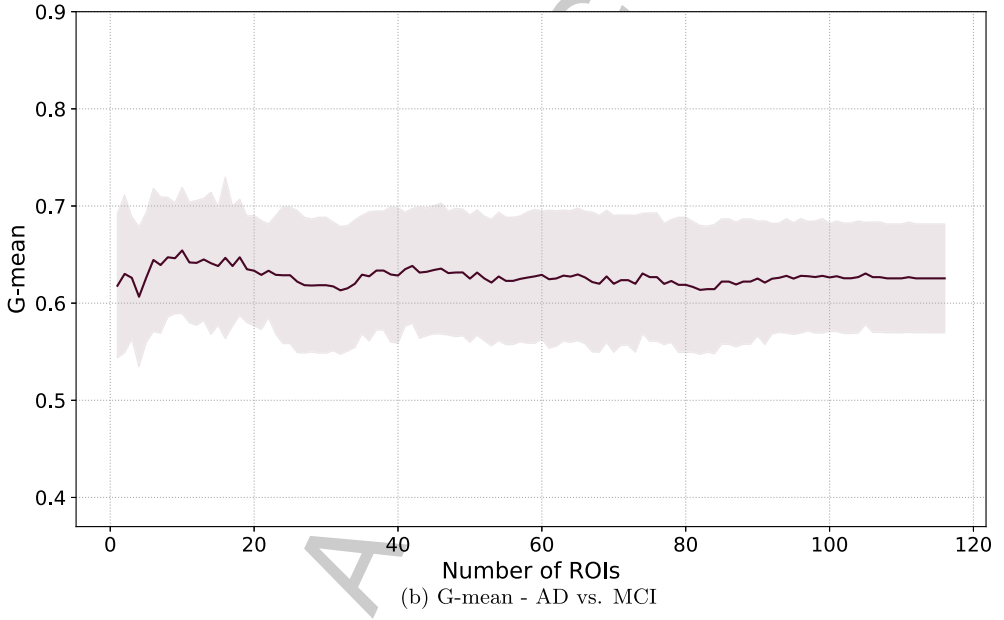
Fig. 4. Comparisons of Accuracy (a) and G-mean (b) registered using *10-fold cross-validation* to afford MCI versus NC problem, selecting the n ROIs with higher weights (for n from 1 to 116). Shaded areas cover mean $\pm$ standard deviation.

5. The proposed greedy algorithm is applied with the following feature selection thresholds: 0.65 in the case of AD versus NC problem, 0.55 in MCI versus NC, and 0.6 in AD versus MCI. We decide to choose lower feature selection thresholds for AD versus MCI and MCI versus NC to avoid an excessive reduction of the number of ROIs since p_p and p_n values are, in general, lower in these two problems than in AD versus NC.

The results of applying these algorithms are compared with the voxel-as-features (VAF) [65] baseline method. A variant of this method, referred as VAF-FS, is also included in the experimentation. VAF-FS employs the same feature selection criterion as the greedy classifier. Only those voxels of the selected regions are used as input features of VAF. Since J48, RF, SVM, and XGB have their own feature selection mechanisms, we do not employ any other additional



(a) Accuracy - AD vs. MCI



(b) G-mean - AD vs. MCI

Fig. 5. Comparisons of Accuracy (a) and G-mean (b) registered using *10-fold cross-validation* to afford AD versus MCI problem, selecting the n ROIs with higher weights (for n from 1 to 116). Shaded areas cover mean $\pm$ standard deviation.

feature selection. Fig. 6 shows the relevance of the 30 main features according to RF in AD versus NC.

Given the limited number of images in the dataset, the best option is to perform a series of experiments varying the test and training sets in a way that each sample will be in the test set once and in the training set for the other experiments. This allows to get a wider set of test samples to check the quality of the tested classifiers. With this approach, known as cross-validation, the dataset is partitioned into k subsets, k experiments are performed using a different test subset each time and the union of the remaining $k-1$ subsets for training. In this particular case, 10-fold cross-validation is used. In the proposed greedy algorithm, $f(I)$ and $w(p^P, p^N)$ values are computed from the training set.

The evaluation measures used to assess the quality of the results are 1) Accuracy, 2) Sensitivity,

Table 2
Performance results obtained with the six algorithms in the three binary problems (AD versus NC, MCI versus NC and AD versus MCI) using 50-bins histograms. Performance is assessed in terms of accuracy, sensitivity, specificity, AUC, and G-mean

Subjects	Classifier	Accuracy	Sensitivity	Specificity	AUC	G-mean
AD vs. NC	VAF	0.7772±0.064	0.7658±0.122	0.7862±0.059	0.7760±0.068	0.7728±0.070
	VAF-FS	0.8586±0.054	0.8404±0.079	0.8735±0.086	0.8569±0.053	0.8545±0.052
	C4.5	0.7298±0.069	0.6769±0.125	0.7733±0.094	0.7166±0.080	0.7183±0.072
	RF	0.8279±0.065	0.7880±0.122	0.8605±0.067	0.8889±0.052	0.8249±0.078
	SVM	0.8131±0.034	0.7661±0.089	0.8516±0.062	0.8853±0.031	0.8071±0.041
	XGBoost	0.7966±0.056	0.7509±0.148	0.8342±0.050	0.8667±0.048	0.7856±0.070
	Greedy	0.8039±0.072	0.7243±0.142	0.8692±0.058	0.8942±0.048	0.7886±0.087
MCI vs. NC	VAF	0.6157±0.062	0.7104±0.099	0.4498±0.095	0.5801±0.057	0.5676±0.082
	VAF-FS	0.6777±0.061	0.8032±0.067	0.5107±0.096	0.6388±0.064	0.6362±0.055
	C4.5	0.6398±0.066	0.7332±0.091	0.4769±0.109	0.5993±0.068	0.5860±0.072
	RF	0.7192±0.052	0.8729±0.042	0.4502±0.091	0.7611±0.067	0.6240±0.073
	SVM	0.7130±0.082	0.8479±0.066	0.4769±0.137	0.7562±0.076	0.6301±0.112
	XGBoosting	0.6922±0.050	0.7709±0.073	0.555±0.077	0.7532±0.052	0.6510±0.051
	Greedy	0.6431±0.049	0.5912±0.051	0.7342±0.068	0.7373±0.068	0.6582±0.051
AD vs. MCI	VAF	0.6332±0.031	0.3301±0.063	0.7755±0.058	0.5528±0.027	0.5020±0.042
	VAF-FS	0.6435±0.031	0.3936±0.083	0.7607±0.050	0.5771±0.037	0.5424±0.057
	C4.5	0.6247±0.048	0.3927±0.109	0.7334±0.056	0.5655±0.054	0.5302±0.078
	RF	0.6739±0.031	0.1330±0.073	0.9276±0.040	0.7008±0.059	0.3584±0.085
	SVM	0.6841±0.030	0.1810±0.089	0.9201±0.047	0.6919±0.057	0.3958±0.063
	XGBoost	0.6892±0.035	0.3673±0.077	0.8404±0.054	0.7008±0.052	0.5511±0.059
	Greedy	0.6469±0.022	0.5164±0.122	0.7084±0.053	0.6459±0.059	0.5981±0.052

3) Specificity, 4) G-mean, and 5) Area Under ROC Curve (AUC) [66, 67] described as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$Sensitivity = TPR = \frac{TP}{TP + FN} \tag{9}$$

$$Specificity = TNR = \frac{TN}{TN + FP} \tag{10}$$

$$G\text{-}mean = \sqrt{Sensitivity \cdot Specificity} \tag{11}$$

where TP is the number of true positives: number of samples of the positive class correctly classified; TN is the number of true negatives: number of samples of the negative class correctly classified; FP is the number of false positives: number of negative samples classified as positive; and FN is the number of false negatives: number of positive samples classified as negative. Sensitivity is also known as True Positive Rate (TPR) and specificity is also known as True Negative Rate (TNR). AUC is computed based on the probability assigned to each test example of belonging to the positive class. Let p_i be such probability for a certain test example as follows:

$$p_i = 0.5 + \begin{cases} \frac{|F_i|}{2}, & \text{if } F_i \leq 0 \\ 1 - \frac{|F_i|}{2}, & \text{if } F_i > 0 \end{cases} \tag{12}$$

Since the three problems are to a greater or lesser degree imbalanced, accuracy may be not the most fitting measure. High accuracy can be maintained despite of high error rate in the minority class. Therefore, we also use sensitivity, specificity, AUC and G-mean to assess if the error is shared out between classes and to obtain more information about how the error is distributed. G-mean emphasizes not only the importance of achieving good results in both classes, but also a balance between the two results. This measure turns out to be specially suitable in this type of medical problem where there can be a high imbalance ratio. Due to this imbalance ratio, classifiers may tend to high percentage of misclassified samples from the minority class, even though this class might be the most relevant.

In each binary problem we consider the positive class to be the one which represents a closer state to illness, which implies that: 1) AD versus NC: AD is the positive class and NC is the negative one; 2) MCI versus NC: MCI is the positive class and NC is the negative one; and 3) AD versus MCI: AD is the positive class while MCI is the negative one. In regards to the number of samples of positive and negative classes, we can make one distinction between problems. On one hand, in AD versus NC and AD versus MCI the positive class is the minority one (AD). On the other hand, in MCI versus NC the positive class (MCI) is the majority one. This fact along with the

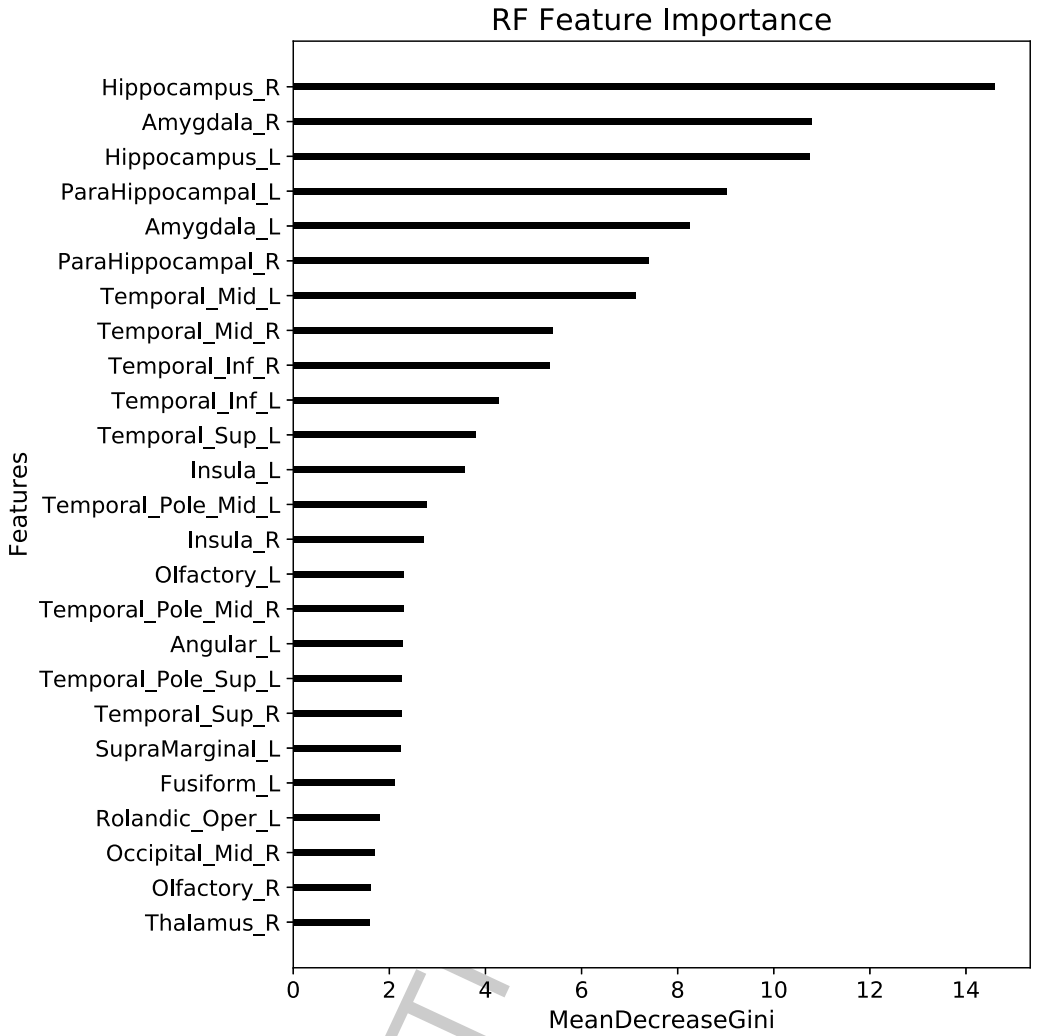


Fig. 6. Importance of the 30 most relevant features for AD versus NC problem according to RF.

imbalance ratio must be taken into account during the results analysis.

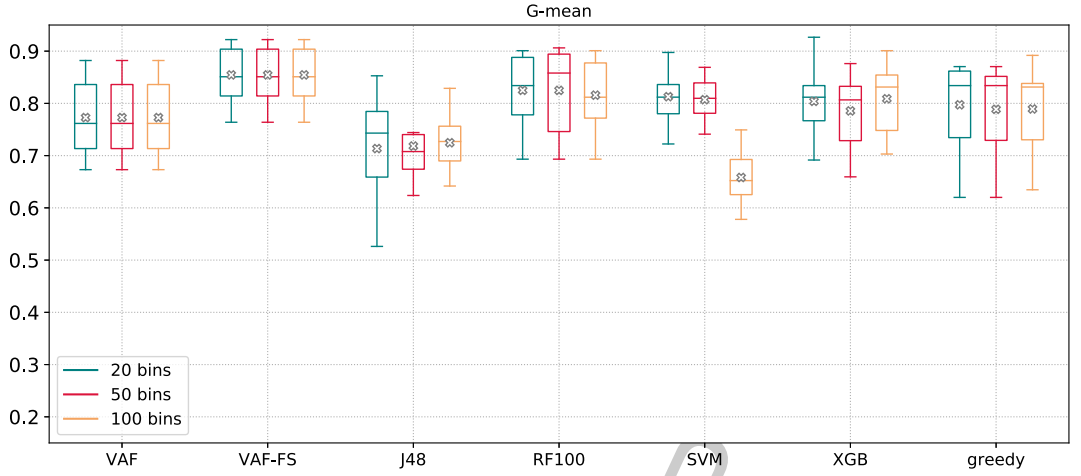
RESULTS

In this section we present and analyze the results obtained from the different tested techniques. In this sense, Table 2 shows the quality of the classification results obtained applying each of the classifiers specified in the previous section for each binary problem (AD versus NC, MCI versus NC, and AD versus MCI).

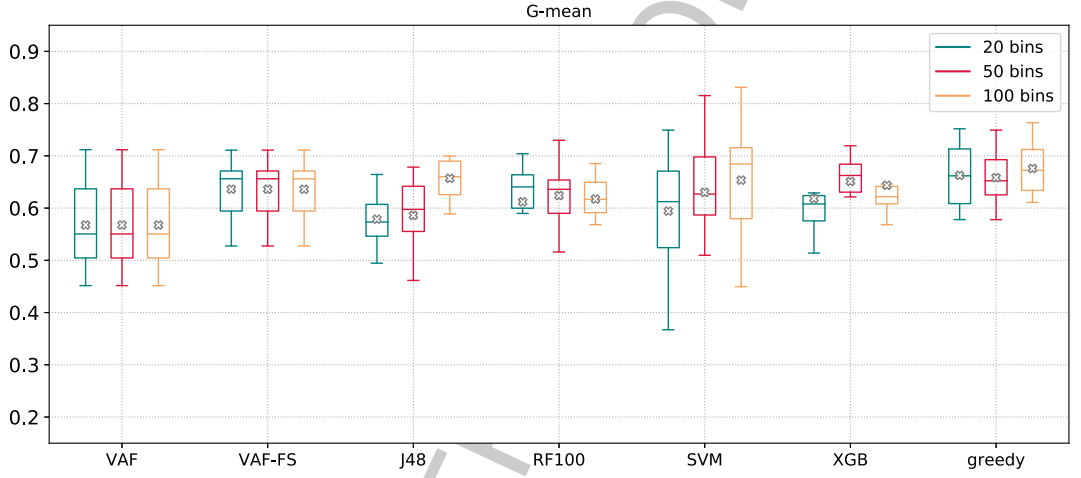
In Table 2 it is possible to appreciate how specificity tends to be higher than sensitivity when the positive class is the minority and how this trend is reversed when the positive class is the majority. This means that it is usually easier for most of the classifiers to correctly classify samples from the majority class. We can also notice that these differences between sensitivity and specificity increase as the imbalance ratio increases.

Fig. 7 shows a comparative analysis of the different algorithms and experiments executed. Regarding G-mean we find differences between problems. G-mean values for AD versus NC are the highest, followed by MCI versus NC and AD versus MCI in last place. This is possibly due to the higher imbalance ratio in AD versus MCI.

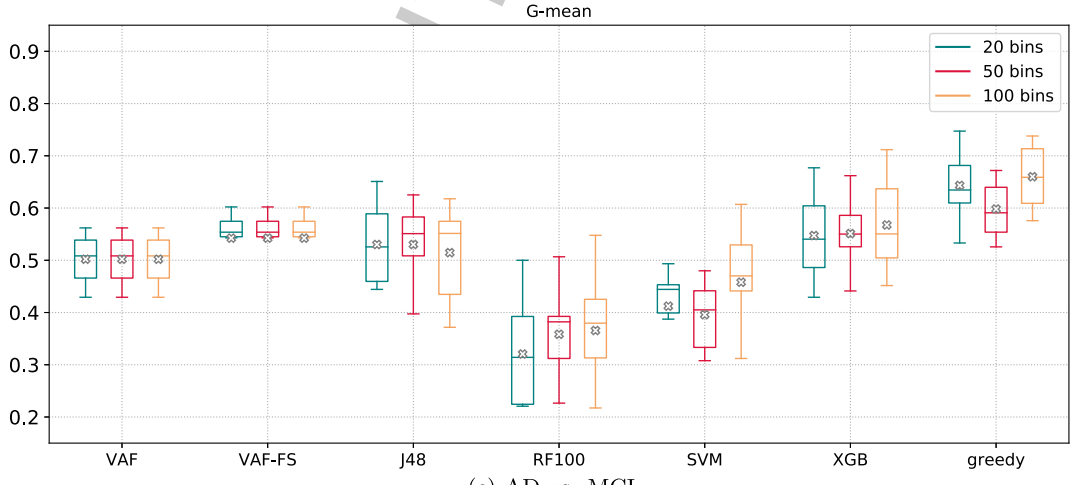
We can also analyze Table 2 and Fig. 7 from other perspectives, i.e., studying if the number of bins that form the histograms has a systematic and significant impact on the quality of the classification and



(a) AD vs. NC



(b) MCI vs. NC



(c) AD vs. MCI

Fig. 7. G-mean values for tests performed with histograms of three different sizes (20, 50 or 100 bins), seven classification algorithms and 10-fold cross-validation for each binary problem addressed: (a) AD versus NC, (b) MCI versus NC and (c) AD versus MCI.

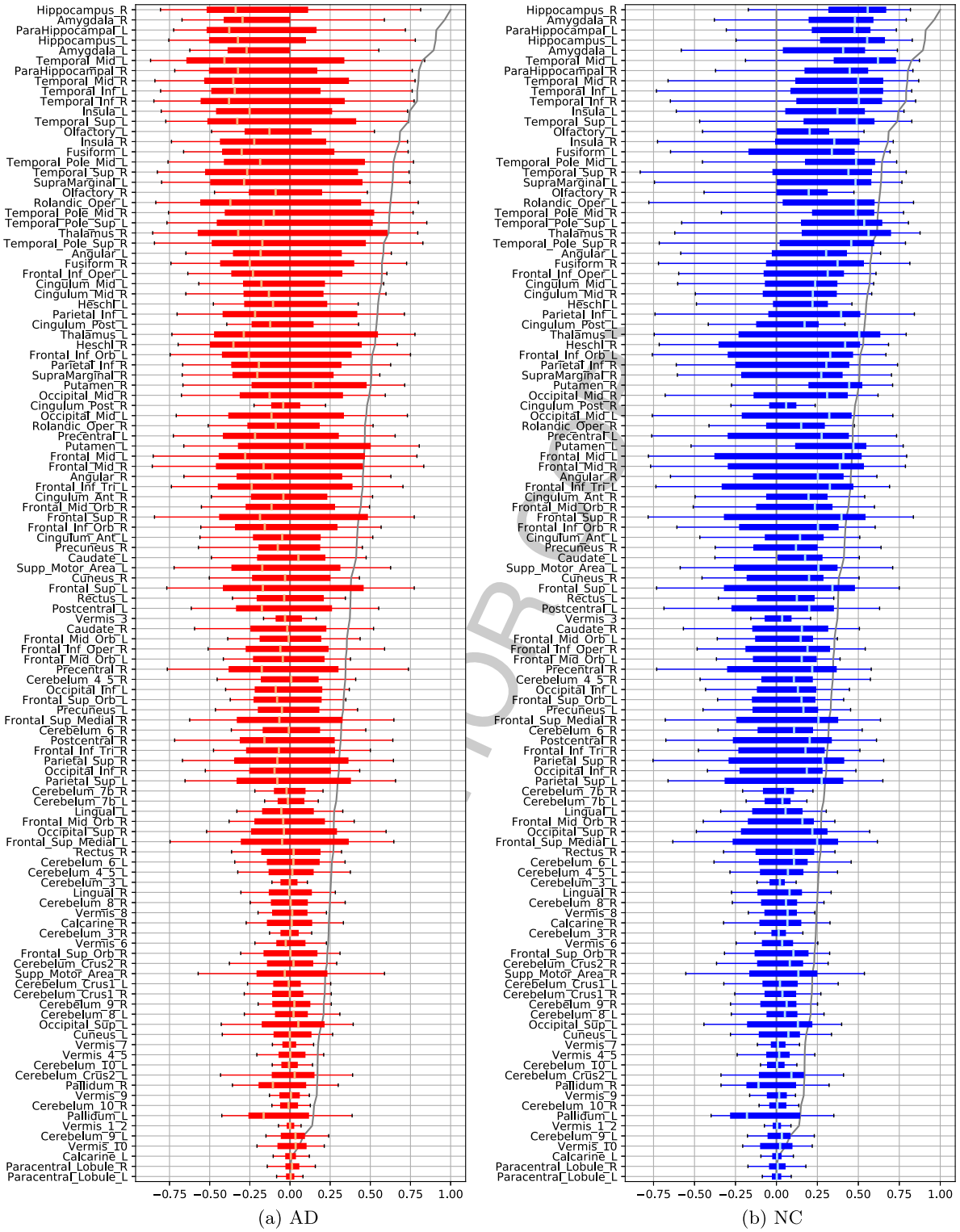


Fig. 8. Distribution of $f(I)$ (shown in boxplots) along with weight of each ROI (gray line) for AD subjects (left) and NC subjects (right) inside the AD versus NC problem. ROIs are sorted from higher to lower weight. We compute both measures using *10-fold cross validation*.

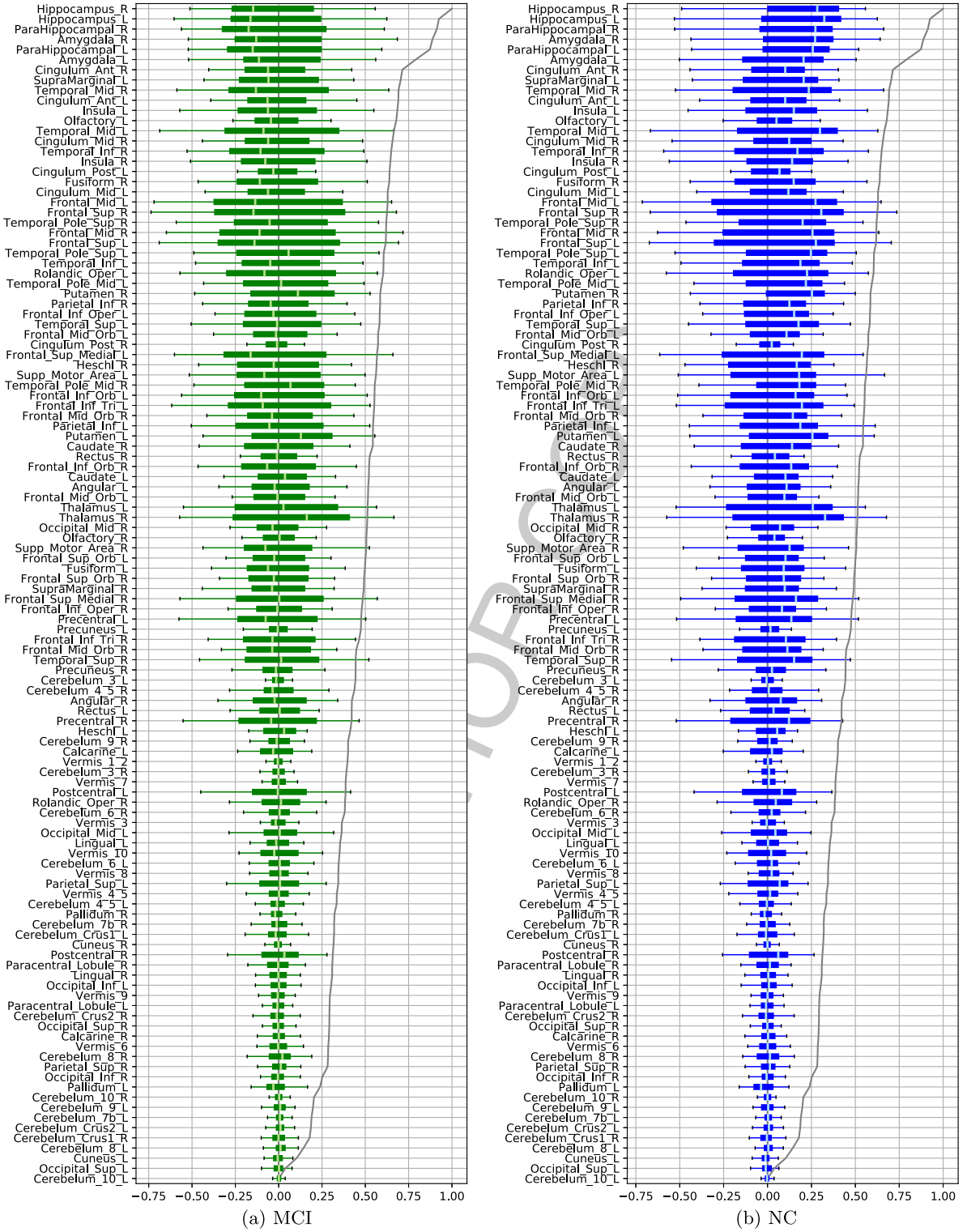


Fig. 9. Distribution of $f(I)$ (shown in boxplots) along with weight of each ROI (gray line) sorting ROIs by weigth for MCI subjects (left) and NC subjects (right) inside the MCI versus NC problem. ROIs are sorted from higher to lower weight. We compute both measures using *10-fold cross validation*.

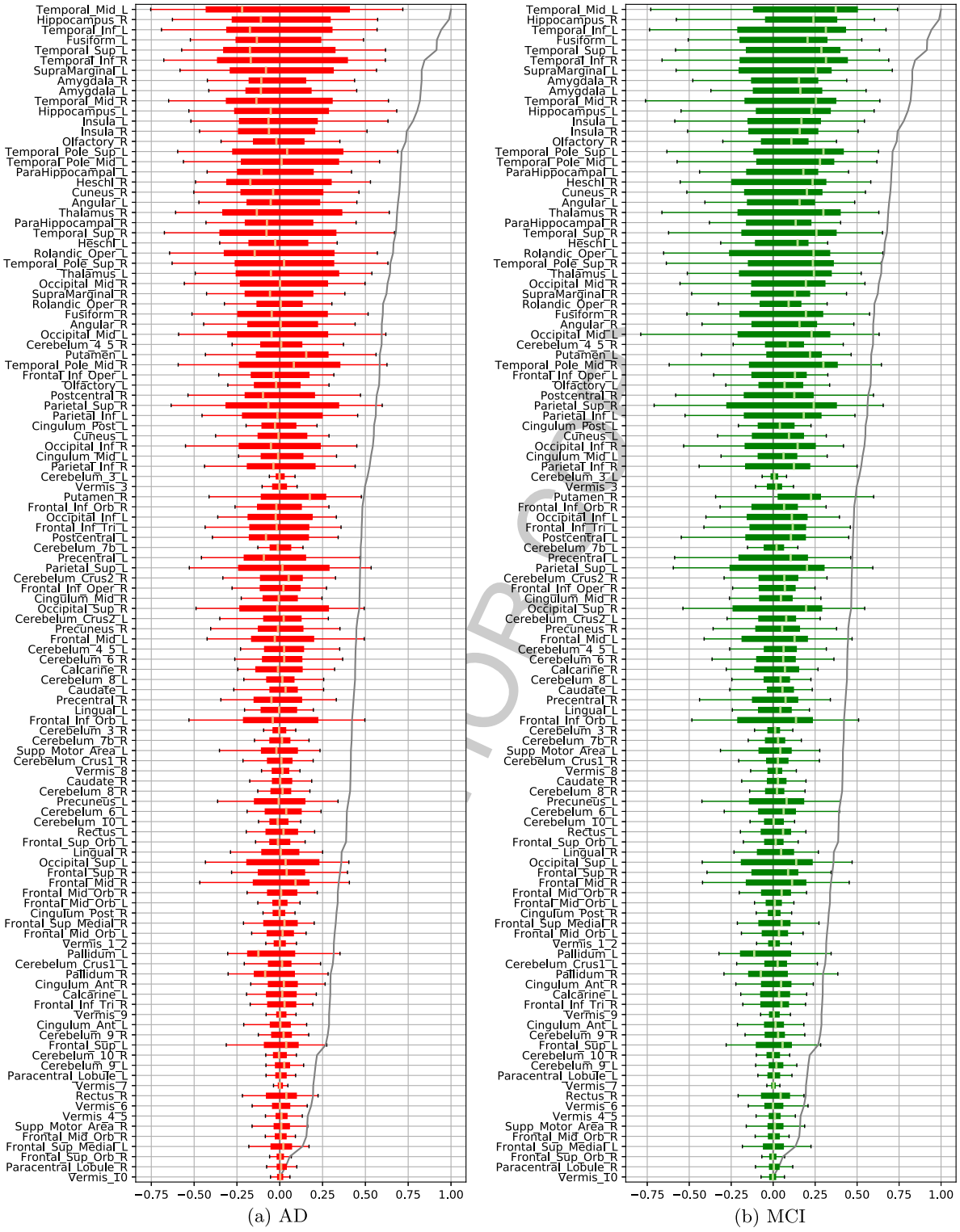
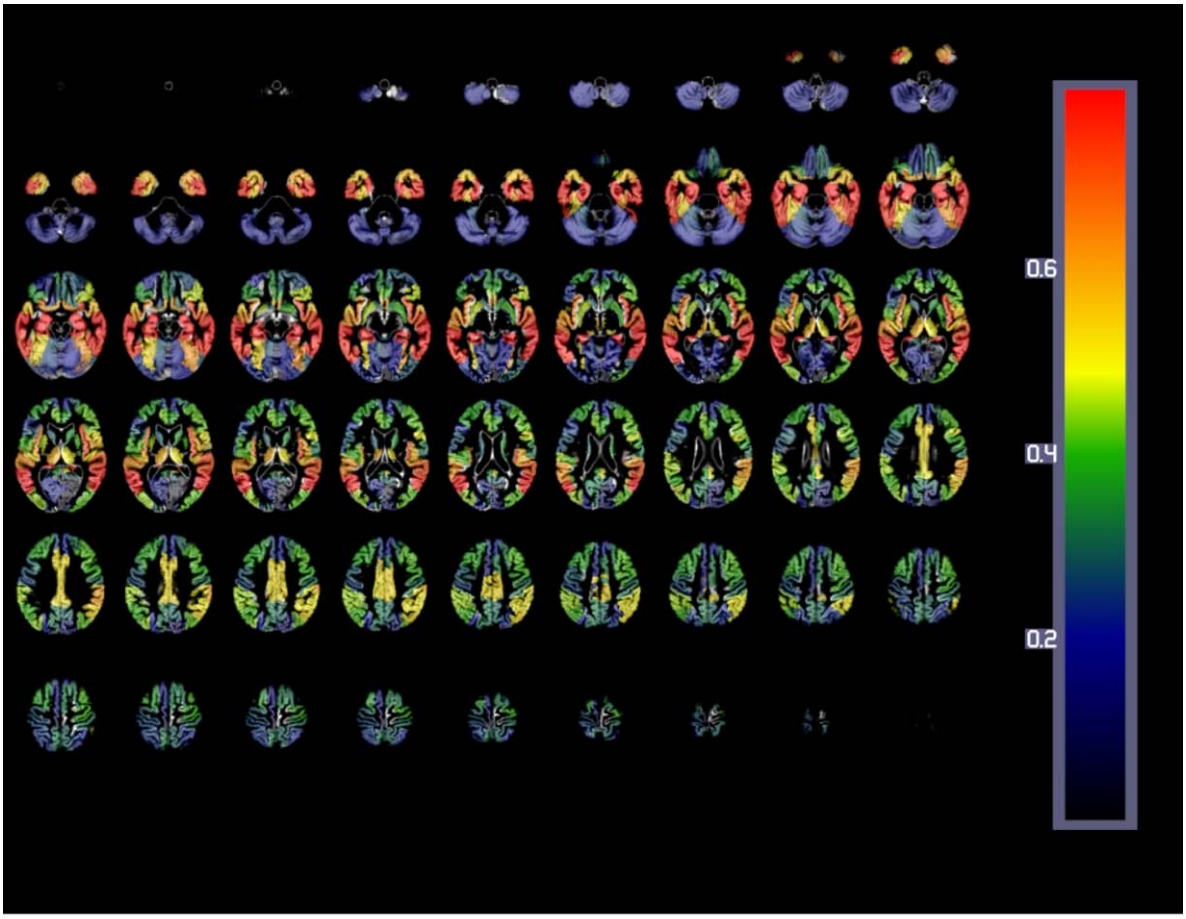
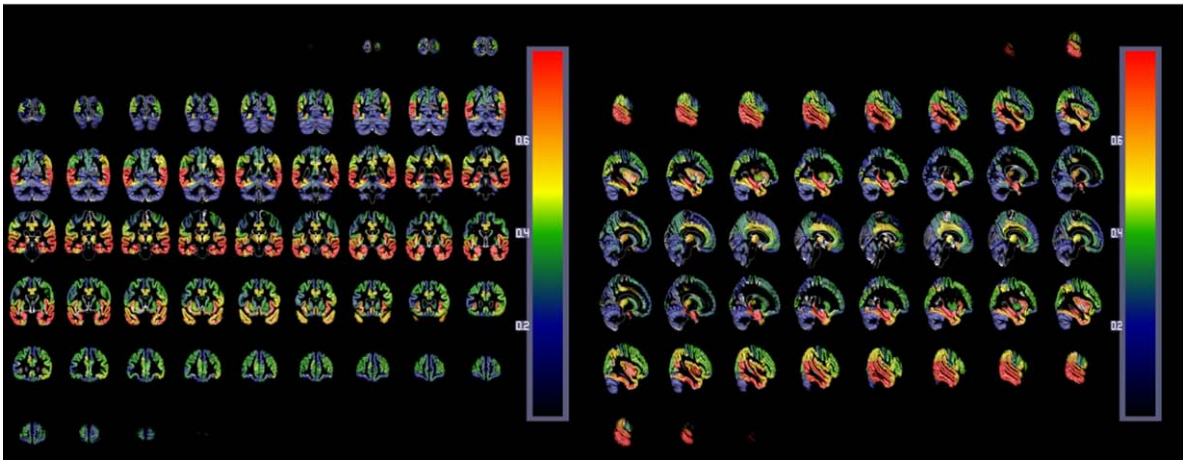


Fig. 10. Distribution of $f(I)$ (shown in boxplots) along with weight of each ROI (gray line) sorting ROIs by weigth for AD subjects (left) and MCI subjects (right) inside the AD versus MCI problem. ROIs are sorted from higher to lower weight. We compute both measures using *10-fold cross validation*.



(a) Axial



(b) Coronal (c) Sagittal

Fig. 11. Human brain representation in (a) axial, (b) coronal and (c) sagittal views colouring each of the 116 ROIs according to their weight in AD versus NC problem. Blue tones for the lowest weights and red tones for the highest.

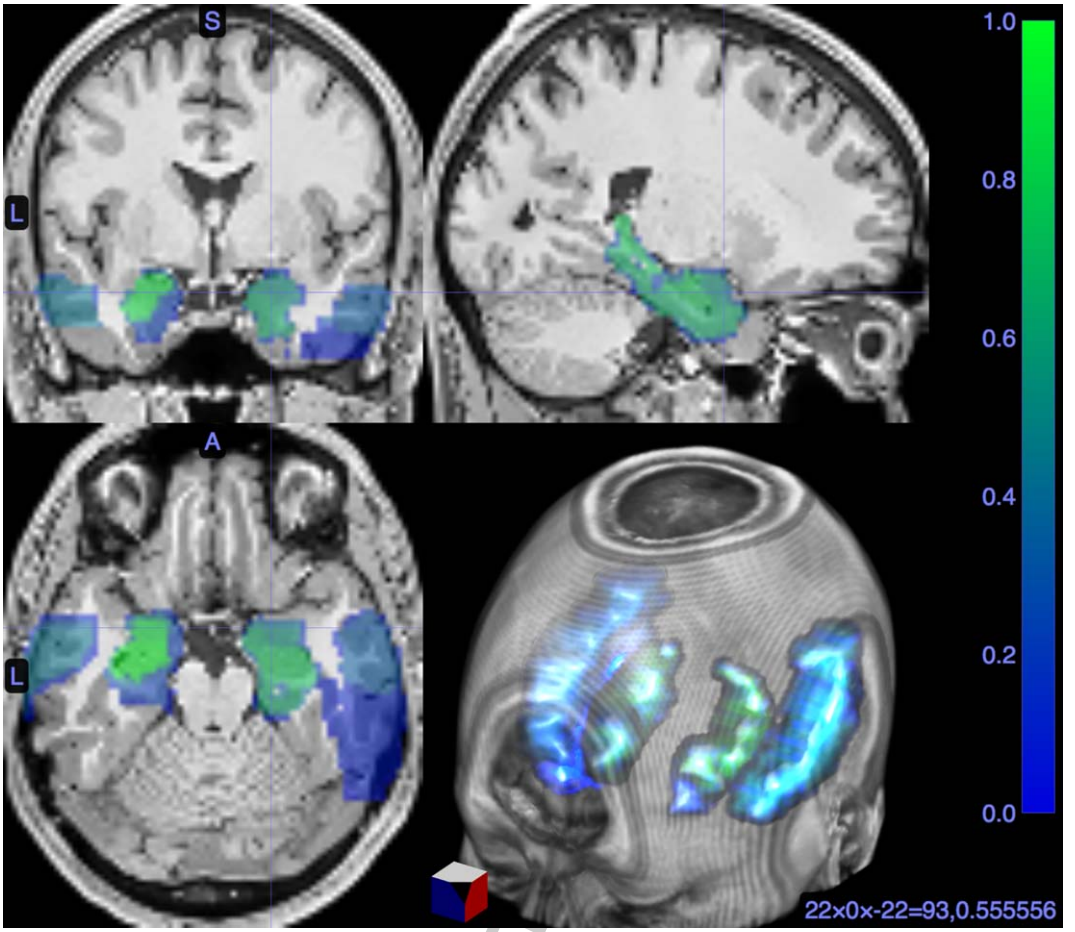


Fig. 12. Human brain representation in colouring the ten most discriminant ROIs according to their weight in AD versus NC problem. Green tones for the highest weights and blue tones for the lowest.

comparing tested classifiers to determine which of them would be the most desirable option. In Fig. 7 we can appreciate that VAF and VAF-FS results do not depend on the number of bins since both options imply that the classifier is applied directly to voxels. The results of the other classifiers undergo certain variations when the number of bins is modified. Nevertheless, these changes are not significant in most cases. It may seem that 100 bins is the option which maintains a good position in a greater number of cases but still not in all of them. No significant patterns are appreciated regardless of whether we perform the analysis by problem, measure, or classifier.

VAF-FS clearly improves VAF results in all cases, as shown in both Table 2 and Fig. 7. Furthermore, the fact of applying feature selection before VAF implies these better results obtained by VAF-FS have been achieved using a lower number of input features. It is also important to have in mind that to make this selection of features possible it is necessary to carry out almost the full process of the greedy proposal. Indeed, the process must be completed until the computation of $\widehat{w}_j(p_j^P, p_j^N)$ for every ROI to be able to select only those that surpass the threshold imposed.

G-mean is quite high for all the tested classifiers in AD versus NC and even in MCI versus NC, as shown in Table 2 and Fig. 7. In both problems, results achieved by all the classifiers are quite similar. The specially quite high G-mean values in AD versus NC are possibly because the low error rate is well balanced between both classes. However, it can be noted that VAF and C4.5 achieve the worst results.

In the case of AD versus MCI, differences between classifiers are much more noticeable. G-mean values for algorithms like RF or SVM are specially low since most misclassified instances belong to the same class (positive and minority one). It is worth mentioning that greedy approach is capable of correctly

Table 3
ROI's index, name and weight of the 116 ROIs in the AD versus NC problem. ROIs are sorted from higher to lower weight. ROIs written in italics have p_p or p_n lower than 0.5 and, consequently, are dismissed

Index	Area	Weight	Index	Area	Weight	Index	Area	Weight
38	Hippocampus_R	1.0000	32	Cingulum_Ant_R	0.6930	48	Lingual_R	0.5788
42	Amygdala_R	0.9728	13	Frontal_Inf_Tri_L	0.6888	54	Occipital_Inf_R	0.5783
37	Hippocampus_L	0.9544	51	Occipital_Mid_L	0.6866	23	Frontal_Sup_Medial_L	0.5762
39	ParaHippocampal_L	0.9512	7	Frontal_Mid_L	0.6844	102	Cerebelum_7b_R	0.5757
41	Amygdala_L	0.9380	18	Rolandic_Oper_R	0.6826	50	Occipital_Sup_R	0.5721
86	Temporal_Mid_R	0.8899	*73*	*Putamen_L*	*0.6815*	10	Frontal_Mid_Orb_R	0.5716
85	Temporal_Mid_L	0.8873	66	Angular_R	0.6790	*99*	*Cerebelum_6_L*	*0.5645*
40	ParaHippocampal_R	0.8777	1	Precentral_L	0.6784	47	Lingual_L	0.5638
89	Temporal_Inf_L	0.8763	36	Cingulum_Post_R	0.6775	*114*	*Vermis_8*	*0.5573*
90	Temporal_Inf_R	0.8719	8	Frontal_Mid_R	0.6695	95	Cerebelum_3_L	0.5567
29	Insula_L	0.8547	31	Cingulum_Ant_L	0.6637	91	Cerebelum_Crus1_L	0.5562
81	Temporal_Sup_L	0.8454	4	Frontal_Sup_R	0.6596	92	Cerebelum_Crus1_R	0.5559
30	Insula_R	0.8106	16	Frontal_Inf_Orb_R	0.6559	*97*	*Cerebelum_4_5_L*	*0.5488*
21	Olfactory_L	0.8038	*71*	*Caudate_L*	*0.6504*	*104*	*Cerebelum_8_R*	*0.5483*
82	Temporal_Sup_R	0.7998	26	Frontal_Mid_Orb_R	0.6474	*103*	*Cerebelum_8_L*	*0.5466*
55	Fusiform_L	0.7987	19	Supp_Motor_Area_L	0.6370	*94*	*Cerebelum_Crus2_R*	*0.5465*
87	Temporal_Pole_Mid_L	0.7886	46	Cuneus_R	0.6326	101	Cerebelum_7b_L	0.5463
63	SupraMarginal_L	0.7881	68	Precuneus_R	0.6324	112	Vermis_6	0.5417
78	Thalamus_R	0.7746	57	Postcentral_L	0.6283	*44*	*Calcarine_R*	*0.5401*
22	Olfactory_R	0.7741	3	Frontal_Sup_L	0.6262	*6*	*Frontal_Sup_Orb_R*	*0.5369*
83	Temporal_Pole_Sup_L	0.7688	27	Rectus_L	0.6262	20	Supp_Motor_Area_R	0.5347
88	Temporal_Pole_Mid_R	0.7684	67	Precuneus_L	0.6172	*106*	*Cerebelum_9_R*	*0.5343*
17	Rolandic_Oper_L	0.7683	110	Vermis_3	0.6130	*93*	*Cerebelum_Crus2_L*	*0.5189*
84	Temporal_Pole_Sup_R	0.7526	12	Frontal_Inf_Oper_R	0.6122	*49*	*Occipital_Sup_L*	*0.5180*
65	Angular_L	0.7462	25	Frontal_Mid_Orb_L	0.6119	*115*	*Vermis_9*	*0.5175*
33	Cingulum_Mid_L	0.7457	2	Precentral_R	0.6112	*45*	*Cuneus_L*	*0.5157*
56	Fusiform_R	0.7448	*72*	*Caudate_R*	*0.6110*	*76*	*Pallidum_R*	*0.5063*
34	Cingulum_Mid_R	0.7393	*98*	*Cerebelum_4_5_R*	*0.6110*	109	Vermis_1_2	0.5048
11	Frontal_Inf_Oper_L	0.7317	5	Frontal_Sup_Orb_L	0.6077	*75*	*Pallidum_L*	*0.5008*
35	Cingulum_Post_L	0.7307	53	Occipital_Inf_L	0.6071	*105*	*Cerebelum_9_L*	*0.4966*
77	Thalamus_L	0.7299	9	Frontal_Mid_Orb_L	0.5983	*111*	*Vermis_4_5*	*0.4945*
61	Parietal_Inf_L	0.7291	58	Postcentral_R	0.5972	108	Cerebelum_10_R	0.4944
79	Heschl_L	0.7264	24	Frontal_Sup_Medial_R	0.5942	107	Cerebelum_10_L	0.4903
80	Heschl_R	0.7254	60	Parietal_Sup_R	0.5910	*113*	*Vermis_7*	*0.4882*
15	Frontal_Inf_Orb_L	0.7109	100	Cerebelum_6_R	0.5907	*70*	*Paracentral_Lobule_R*	*0.4837*
74	*Putamen_R*	*0.7096*	*28*	*Rectus_R*	*0.5848*	*43*	*Calcarine_L*	*0.4698*
62	Parietal_Inf_R	0.7033	14	Frontal_Inf_Tri_R	0.5838	*116*	*Vermis_10*	*0.4611*
64	SupraMarginal_R	0.7028	96	Cerebelum_3_R	0.5834	*69*	*Paracentral_Lobule_L*	*0.0000*
52	Occipital_Mid_R	0.6997	59	Parietal_Sup_L	0.5829			

classifying a high percentage of instances of the positive class, despite the high imbalance ratio and the simplicity of the proposal. Thus, the greedy classifier achieves very good G-mean and maintains quite balanced sensitivity and specificity values. It obtains the highest sensitivity in AD versus MCI while other algorithms (e.g., RF or SVM) fail, as shown in Table 2. This type of scenario where the positive class which suffers from a lack of samples turns out to be the most relevant is quite common in disease diagnosis problems.

Furthermore, from the greedy approach we can extract some additional information which has proven to be really valuable to help us gain more knowledge about the problem, i.e., we can analyze the values of $f(I)$ of each region and subject as well as examine weight distribution among the different regions. With this aim, in Figs. 8–10 distribution of $f(I)$ and weight of each ROI are represented. Fig. 10 shows human brain representations in three different views coloring each ROI according to its weight in AD versus NC problem. Finally, Fig. 12 also shows human brain representations but focuses only on the ten most discriminant regions according to the greedy classifier.

DISCUSSION

The aims of this study were, on the one hand, to process the brain image dataset until compressing

the information into a well structured dataset which is valid and efficient for supervised-learning approaches and to apply certain classiffication algorithms to said dataset; and on the other hand, to develop an ad hoc greedy algorithm which is easily comprehensible but obtains competitive results.

VAF is used as a baseline classifier with which we compare the proposed methods. This technique uses all voxels in each image as input features for a classifier. In this case, the only classifier tested in the VAF approximation is linear SVM, due to the high number of input features. Because of its simplicity and the quality of its performance, VAF is considered as a reference: in [65], a pathological confirmed database was used to estimate a performance superior to visual assessments by experts.

The method proposed to reduce the number of input features passed to the machine learning algorithms (having one feature per ROI instead of per voxel) works really well. The number of features is significantly reduced improving efficiency and widening the range of viable classification algorithms. Furthermore, this is achieved without losing accuracy since the performance results stay at the same level or even improve depending on the classification algorithm used, as shown in Table 2. All these results correspond to experiments performed considering *10-fold cross-validation*.

The original greedy algorithm gets very competitive results in the three binary problems, being better than other approaches which are much more complex. One of the principles behind this algorithm is to take into account that not all the brain regions suffer AD to the same effect. As we have explained before, this greedy algorithm includes a mechanism to assign a specific weight to each ROI. Those ROIs with the greatest distances between subjects of one class and of the other have the most influence on the classification result.

Furthermore, a feature selection threshold is included in this greedy proposal with the purpose of avoiding favoring ROIs whose weight are high because they are strong in one class at the expense of being weak in the other. ROIs' weights are determined by the mean of p_p and p_n (percentages of subjects whose $f(I)$ values for the ROI are in the right range for the positive or negative class, respectively), the weight of a certain ROI may be high or medium due to high enough value of one of these two measures (p_p or p_n), even though the other one is low. Subjects of the positive class are expected to have values of $f(I)$ in $[-1, 0)$ while subjects of the negative class are expected to have values higher than zero between $(0, +1]$. However, in Figs. 8–10 we can observe how for certain ROIs there is a high percentage of subjects (reaching 50% or higher in some cases) in the wrong interval. This means that for these ROIs, at least, a significant percentage of the examples of a certain class have histograms that are closer to the average histogram of the wrong class than to the one of their own class.

From this weight-base proposal (greedy algorithm) we can infer some additional knowledge that helps us to better understand the problem. Table 3 along with Figs. 11 and 12 are some examples of such extra knowledge. Table 2 shows the 116 brain regions of interest sorted from higher to lower weight for the AD versus NC classification; while Fig. 11 shows a representation of the human brain in which each of these 116 ROIs is colored according to its weight. Similar knowledge can be also extracted from some of the others classifiers (not from all of them) as evidenced in Fig. 6. This figure shows the most relevant features according to RF in AD versus NC. Most of the relevant features for this classifier match with the most relevant features in accordance to the greedy approach. Concretely, the ten most relevant features for RF are also the most relevant for greedy. This type of interpretable results are much easier to attain if the input features are at ROI level instead of at voxel level. The contents of this table and the noted figures corroborate progressive cerebral atrophy as a characteristic feature of neurodegeneration in patients progressing from a cognitive normal healthy state to MCI and AD [68]. Traditional studies of regional MRI volumes have shown that AD is characterized by a progression of atrophy in the medial temporal lobe [69] the entorhinal cortex typically being the earliest region of atrophy, closely followed by the hippocampus, amygdala, and parahippocampus [70–72], which fits perfectly with the information shown in Table 3, and Figs. 11 and 12. Even more, other ROIs within the limbic lobe (ie., posterior cingulate) are also affected during the early stage of the disease.

CONCLUSION

In this study we propose a new AD CAD system based on the estimation of distance per ROI in MRI between AD, MCI, and NC samples and the use of machine learning algorithms for classification (e.g.,

classification trees, ensembles, SVM, boosting or greedy techniques). Among these algorithms, we include a new proposal based on discovering which are the brain regions (ROIs) that exhibit the greatest loss of GM caused by AD, and also on the particular distances per ROI between a certain sample and the mean computed for each possible type of subject (AD, MCI, or NC). The source code of this technique can be found in the following repository: https://github.com/eruizsanchez/CAD_Alzheimers_Disease. Furthermore, a meaningful visual support is offered together with the proposed classifier's results, including several graphics related with relevance of ROIs in the classification process and how these discriminative regions are actually distributed in brain. This enriches the CAD system and allows to better understand the underlying logic in the CAD output.

The performance results have shown that most of the algorithms tested on the datasets built with the originally proposed histogram-based preprocessing method obtain better results than the baseline (VAF), reaching an improvement of 12% of AUC and 5% of G-mean in AD versus NC; 14% of AUC and 9% of G-mean in AD versus MCI, and 18% of AUC and 9% of G-mean in MCI versus NC. In addition, the greedy algorithm originally proposed in this study surpasses the baseline method and some of the other complex algorithms tested. This algorithm maintains a desirable balance between sensitivity and specificity. It should be noted that this greedy proposal is efficient and easy to program, notwithstanding it obtains competitive results facing some more advanced and complex machine learning algorithms. This suggests that the proposed technique of histogram analysis works in the current problem setting. However, the proposal still has one important limitation since classifiers are applied to three different binary problems (AD versus NC, MCI versus NC and AD versus MCI) instead of directly to the raw three-class problem. Furthermore, some work still has to be done to minimize the impact of the imbalanced classification problem on the results of the classifiers.

To conclude, we want to extend our study by including other databases that allow the algorithms to learn new information, improve the classification models making them even more accurate and reliable. It is also our intention to work to achieve accurate classification models facing the problem with multiclass approaches. Furthermore, we will apply association rules mining on the information generated by this study to delve deeper and better understand the relationship between different brain ROIs.

ACKNOWLEDGMENTS

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (http://www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

This work was supported by the MINECO/FEDER under the Spanish National Research Projects TEC 2015-64718-R, TIN2014-57251-P and TIN2017-89517-P; the Consejería de Economía, Innovación, Ciencia, y Empleo of the Junta de Andalucía under the P11-TIC-7103 Excellence Project and the Project BigDaP-TOOLS – Ayudas Fundación BBVA a Equipos de Investigación Científica 2016. E. Ruiz holds a grant from the Spanish Ministry of Economy and Competitiveness (project TIN2014-57251-P).

Authors' disclosures available online (https://www.j-alz.com/manuscript-disclosures/17-0514r3).

REFERENCES

[1] World Heath Organization (2012) *Dementia: A public health priority*. World Health Organization.

[2] Murphy MP, LeVine H (2010) Alzheimer's disease and the β-amyloid peptide. *J Alzheimers Dis* **19**, 311-318.

[3] Alzheimer's Disease Society (2014) *Factsheet: Drug Treatments for Alzheimer's Disease.* http://alzheimers.org.uk.

[4] Pérez G, Conci A, Moreno AB, Hernandez-Tamames JA (2014) Rician noise attenuation in the wavelet packet transformed domain for brain MRI. *Integr Comput Aided Eng* **21**, 163-175.

[5] Goncalves N, Nikkilä J, Vigario R (2014) Self-supervised MRI tissue segmentation by discriminative clustering. *Int J Neural Syst* **24**, 1450004.

[6] Morabito FC, Campolo M, Labate D, Morabito G, Bonanno L, Bramanti A, Salvo S de, Marra A, Bramanti P (2015) A longitudinal EEG study of Alzheimer's disease progression based on a complex network approach. *Int J Neural Syst* **25**, 1-18.

[7] Ahmadlou M, Adeli H, Adeli A (2010) New diagnostic EEG markers of the Alzheimer's disease using visibility graph. *J Neural Transm* **117**, 1099-1109.

[8] Sankari Z, Adeli H (2011) Probabilistic neural networks for EEG-based diagnosis of Alzheimer's disease using conventional and wavelet coherence. *J Neurosci Methods* **197**, 165-170.

[9] Montejo P, Montenegro M, Fernández MA, Maestú F (2011) Subjective memory complaints in the elderly: Prevalence and influence of temporal orientation, depression and quality of life in a population- based study in the city of Madrid. *Aging Mental Health* **15**, 85-96.

[10] Adamczuk K, Schaeverbeke J, Vanderstichele HM, Lilja J, Nelissen N, Van Laere K, Dupont P, Hilven K, Poesen K, Vandenberghe R (2015) Diagnostic value of cerebrospinal fluid Aβ ratios in preclinical Alzheimer's disease. *Alzheimers Res Ther* **7**, 75.

[11] Schroeter ML, Stein T, Maslowski N, Neumann J (2009) Neural correlates of Alzheimer's disease and mild cognitive impairment: A systematic and quantitative meta-analysis involving 1351 patients. *Neuroimage* **47**, 1196-1206.

[12] Aksu Y, Miller DJ, Kesidis G, Bigler DC, Yang QX (2011) An MRI-derived definition of MCI-to-AD conversion for long-term, automatic prognosis of MCI patients. *PLoS One* **6**, e25074.

[13] Adaszewski S, Dukart J, Kherif F, Frackowiak R, Draganski B, Alzheimer's Disease Neuroimaging Initiative (2013) How early can we predict Alzheimer's disease using computational anatomy? *Neurobiol Aging* **34**, 2815-2826.

[14] Ramírez J, Chaves R, Gorriz JM, Lopez M, Alvarez IA, Salas-Gonzalez D, Segovia F, Padilla P (2009) Computer aided diagnosis of the Alzheimer's disease combining SPECT-based feature selection and random forest classifiers. In *Proc IEEE Nuclear Science Symp Conf Record (NSS/MIC)*, pp. 2738-2742.

[15] Górriz J, Segovia F, Ramírez J, Lassl A, Salas-González D (2011) GMM based SPECT image classification for the diagnosis of Alzheimer's disease. *Appl Soft Comput* **11**, 2313-2325.

[16] López M, Ramírez J, Górriz J, Álvarez I, Salas-González D, Segovia F, Chaves R, Padilla P, Gómez-Río M (2011) Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease. *Neurocomputing* **74**, 1260-1271.

[17] Álvarez I, Gorriz J, Ramírez J, Salas-Gonzalez D, Lopez M, Segovia F, Chaves R, Gomez-Rio M, Garcia-Puntonet C (2011) 18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis. *Inf Sci* **184**, 903-196.

[18] Segovia F, Górriz J, Ramírez J, Salas-González D, Álvarez I, López M, Chaves R, Alzheimer's Disease Neuroimaging Initiative (2012) A comparative study of the feature extraction methods for the diagnosis of Alzheimer's disease using the ADNI database. *Neurocomputing* **75**, 64-71.

[19] Ortiz A, Górriz J, Ramírez J, Martínez-Murcia F (2014) Automatic ROI selection in structural brain MRI using SOM 3D projection. *PLoS One* **9**, 1-12.

[20] Ortiz A, Górriz JM, Ramírez J, Martínez-Murcia FJ, Alzheimers Disease Neuroimaging Initiative (2013) LVQ-SVM based CAD tool applied to structural MRI for the diagnosis of the Alzheimer's disease. *Pattern Recognit Lett* **34**, 1725-1733.

[21] Chyzhyk D, Graña M, Savio A, Maiora J (2012) Hybrid dendritic computing with kernel-LICA applied to Alzheimer's disease detection in MRI. *Neurocomputing* **75**, 72-77.

[22] Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehéricy S, Habert M, Chupin M, Benali H, Colliot O, Alzheimer's Disease Neuroimaging Initiative (2010) Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *Neuroimage* **56**, 766-781.

[23] Piaggi P, Menicucci D, Gentili C, Handjaras G, Gemignani A, Landi A (2014) Singular spectrum analysis and adaptive filtering enhance the functional connectivity analysis of resting state FMRI data. *Int J Neural Syst* **24**, 1450010.

[24] Michalopoulos K, Bourbakis N (2015) Combining EEG microstates with fMRI structural features for modeling brain activity. *Int J Neural Syst* **25**, 1550041.

[25] Ferdowsi S, Sanei S, Abolghasemi V (2015) A predictive modeling approach to analyze data in EEG–fMRI experiments. *Int J Neural Syst* **25**, 1440008.

[26] Chyzhyk D, Graña M, Öngür D, Shinn AK (2015) Discrimination of schizophrenia auditory hallucinators by machine learning of resting-state functional MRI. *Int J Neural Syst* **25**, 1550007.

[27] Shiino A, Watanabe T, Maeda K, Kotani E, Akiguchi I, Matsuda M (2006) Four subgroups of Alzheimer's disease based on patterns of atrophy using VBM and a unique pattern for early onset disease. *Neuroimage* **33**, 17-26.

[28] Ashburner J (2007) A fast diffeomorphic image registration algorithm. *Neuroimage* **38**, 95-113.

[29] Ayache N (1996) Analyzing 3D images of the brain. *Neuroimage* **4**, S34-35.

[30] Han X, Jovicich J, Salat D, Kouwe, A van der, Quinn B, Czanner S, Busa E, Pacheco J, Albert M, Killiany R, Maguire P, Rosas D, Makris N, Dale A, Dickerson B, Bruce F (2006) Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. *Neuroimage* **32**, 180-194.

[31] Ronan L, Pienaar R, Williams G, Bullmore E, Crow TJ, Roberts N, Jones PB, Suckling J, Fletcher PC (2011) Intrinsic curvature: A marker of millimeter-scale tangential cortico-cortical connectivity? *Int J Neural Syst* **21**, 351-366.

[32] Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE (2011) *Statistical parametric mapping: The analysis of functional brain images*, Academic Press.
[33] Liu M, Zhang D, Shen D, Alzheimer's Disease Neuroimaging Initiative (2012) Ensemble sparse classification of Alzheimer's disease. *Neuroimage* **60**, 1106-1116.
[34] Savio A, Graña M (2013) An ensemble of classifiers guided by the AAL Brain Atlas for Alzheimer's disease detection. In *Advances in Computational Intelligence* 7903, Lecture Notes in Computer Science, Springer Berlin Heidelberg, pp. 107-114.
[35] Dukart J, Mueller K, Barthel H, Villringer A, Sabri O, Schroeter ML, Alzheimer's Disease Neuroimaging Initiative (2013) Meta-analysis based SVM classification enables accurate detection of Alzheimer's disease across different clinical centers using FDG-PET and MRI. *Psychiatry Res Neuroimaging* **212**, 230-236.
[36] Ahmadlou M, Adeli H, Adeli A (2011) Fractality and a wavelet-chaos-methodology for EEG-based diagnosis of Alzheimer disease. *Alzheimer Dis Assoc Disord* **25**, 85-92.
[37] Ahmadlou M, Adeli A, Bajo R, Adeli H (2014) Complexity of functional connectivity networks in mild cognitive impairment patients during a working memory task. *Clin Neurophysiol* **125**, 694-702.
[38] Kovalev VA, Kruggel F, Gertz H-J, Cramon, DY von (2001) Three-dimensional texture analysis of MRI brain datasets. *IEEE Trans Med Imaging* **20**, 424-433.
[39] Unay D, Ekin A, Cetin M, Jasinschi R, Ercil A (2007) Robustness of local binary patterns in brain MR image analysis. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 2098-2101.
[40] Acharya UR, Yanti R, Zheng JW, Krishnan, MMR, TAN JH, Martis RJ, Lim CM (2013) Automated diagnosis of epilepsy using CWT, HOS and texture parameters. *Int J Neural Syst* **23**, 1350009.
[41] Martinez-Murcia F, Górriz J, Ramírez J, Moreno-Caballero M, Gómez-Río M (2014) Parametrization of textural patterns in 123I-ioflupane imaging for the automatic detection of Parkinsonism. *Med Phys* **41**, 012502.
[42] Yoon U, Lee J-M, Im K, Shin Y-W, Cho BH, Kim IY, Kwon JS, Kim SI (2007) Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia. *Neuroimage* **34**, 1405-1415.
[43] Ahmadlou M, Adeli H, Adeli A (2012) Fractality analysis of frontal brain in major depressive disorder. *Int J Psychophysiol* **85**, 206-211.
[44] Illan IA, Górriz JM, Ramírez J, Meyer-Base A (2014) Spatial component analysis of MRI data for Alzheimer's disease diagnosis: A Bayesian network approach. *Front Comput Neurosci* **8**, 156.
[45] Chaves R, Ramírez J, Górriz JM, Illán IA, Gómez-Río M, Carnero C (2012) Effective diagnosis of Alzheimer's disease by means of large margin-based methodology. *BMC Med Inform Decis Mak* **12**, 1-17.
[46] Chaves R, Ramírez J, Górriz JM, Illán IA (2012) Functional brain image classification using association rules defined over discriminant regions. *Pattern Recognit Lett* **33**, 1666-1672.
[47] Ramírez J, Illán IA, Salas-González D (2015) Projecting MRI brain images for the detection of Alzheimer's disease. *Innovation in Medicine and Healthcare 2014* 207, pp. 225.
[48] Martínez-Murcia FJ, Ortiz A, Górriz JM, Ramírez J, Illán I (2015) A volumetric radial LBP projection of MRI brain images for the diagnosis of Alzheimer's disease. In *International Work-Conference on the Interplay Between Natural and Artificial Computation*, pp. 19-28.
[49] Martínez-Murcia FJ, Górriz JM, Ramírez J, Ortiz A (2016) A structural parametrization of the brain using hidden Markov models-based paths in Alzheimer's disease. *Int J Neural Syst* **26**, 1650024.
[50] Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. *IEEE Trans Syst Man Cybern* **3**, 610-621.
[51] Sikiö M, Holli-Helenius KK, Harrison LC, Ryymin P, Ruottinen H, Saunamaki T, Eskola HJ, Elovaara I, Dastidar P (2015) MR image texture in Parkinson's disease: A longitudinal study. *Acta Radiol* **56**, 97-104.
[52] Ortiz A, Munilla J, Górriz JM, Ramírez J (2016) Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease. *Int J Neural Syst* **26**, 1650025.
[53] García S, Luengo J, Herrera, F (2015) *Data preprocessing in data mining*, Springer.
[54] Pyle D (1999) *Data preparation for data mining*, Morgan Kaufmann.
[55] Structural Brain Mapping Group, DoP, *VBM manual*, University of Jena, Germany 2010.
[56] (2007). In *Statistical Parametric Mapping*, Academic Press.
[57] Collins DL, Zijdenbos AP, Kollokian V, Sled JG, Kabani NJ, Holmes CJ, Evans AC (1998) Design and construction of a realistic digital brain phantom. *IEEE Trans Med Imaging* **17**, 463-468.
[58] Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**, 273-289.
[59] Schmahmann JD, Doyon J, McDonald D, Holmes C, Lavoie K, Hurwitz AS, Kabani N, Toga A, Evans A, Petrides M (1999) Three-dimensional MRI atlas of the human cerebellum in proportional stereotaxic space. *Neuroimage* **10**, 233-260.
[60] Quinlan JR (1993) *C4.5: Programs for machine learning*.
[61] Quinlan JR (1996) Improved use of continuous attributes in C4. 5. *J Artif Intell Res* **4**, 77-90.
[62] Breiman L (2001) Random forests. *Mach Learn* **45**, 5-32.
[63] Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* **20**, 273-297.
[64] Friedman JH (2001) Greedy function approximation: A gradient boosting machine. *Ann Stat* 1189-1232.
[65] Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, Fox NC, Jack CR, Ashburner J, Frackowiak RSJ (2008) Automatic classification of MR scans in Alzheimer's disease. *Brain* **131**, 681-689.
[66] Huang J, Ling CX (2005) Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans Knowl Data Eng* **17**, 299-310.
[67] Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit* **30**, 1145-1159.
[68] Johnson KA, Fox NC, Sperling RA, Klunk WE (2012) Brain imaging in Alzheimer disease. *Cold Spring Harb Perspect Med* **2**, a006213.
[69] Scahill RI, Schott JM, Stevens JM, Rossor MN, Fox NC (2002) Mapping the evolution of regional atrophy in Alzheimer's disease: Unbiased analysis of fluid-registered serial MRI. *Proc Natl Acad Sci U S A* **99**, 4703-4707.

[70] Lehericy S, Baulac M, Chiras J, Pierot L, Martin N, Pillon B, Deweer B, Dubois B, Marsault C (1994) Amygdalohippocampal MR volume measurements in the early stages of Alzheimer disease. *AJNR Am J Neuroradiol* **15**, 929-937.

[71] Chan D, Fox NC, Scahill RI, Crum WR, Whitwell JL, Leschziner G, Rossor AM, Stevens JM, Cipolotti L, Rossor MN (2001) Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Ann Neurol* **49**, 433-442.

[72] Dickerson BC, Goncharova I, Sullivan M, Forchetti C, Wilson R, Bennett D, Beckett LA, DeToledo- Morrell L (2001) MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimers disease. *Neurobiol Aging* **22**, 747-754.