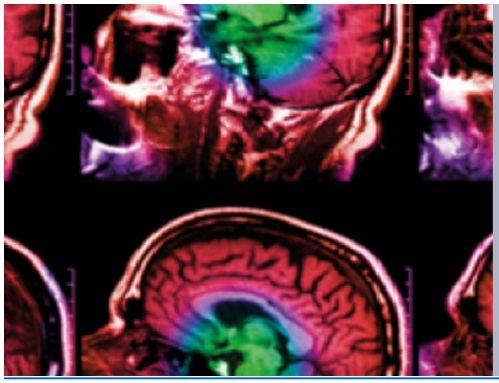
Physics in Medicine & Biology

IPEM
Institute of Physics and Engineering in Medicine

PAPER

Mild cognitive impairment classification using combined structural and diffusion imaging biomarkers

To cite this article: Jorge Perez-Gonzalez *et al* 2021 *Phys. Med. Biol.* **66** 155010



IOP Publishing Phys. Med. Biol. 66 (2021) 155010 https://doi.org/10.1088/1361-6560/ac0e77

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

RECEIVED
5 January 2021

REVISED
27 May 2021

ACCEPTED FOR PUBLICATION
24 June 2021

PUBLISHED
22 July 2021

PAPER

Mild cognitive impairment classification using combined structural and diffusion imaging biomarkers

Jorge Perez-Gonzalez[1], Luis Jiménez-Ángeles[2], Karla Rojas Saavedra[3], Eduardo Barbará Morales[4], Verónica Medina-Bañuelos[5] for the Alzheimer's Disease Neuroimaging Initiative[6]

[1] Unidad Académica del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas en el Estado de Yucatán, UNAM, Yucatán, México
[2] Department of Biomedical Systems Engineering, Engineering Faculty, UNAM, Mexico City, México
[3] Health Sciences Department, Universidad del Valle de México, Mexico City, México
[4] Engineering and Exact Sciences Division. Anahuac Mayab University, Yucatán, México
[5] LINI, Electrical Engineering Department, UAM-Iztapalapa, Mexico City, México
[6] For more information about ADNI see the Acknowledgments section.

E-mail: vera@xanum.uam.mx

Keywords: random forest, MCI, automatic classification

Supplementary material for this article is available online

Abstract

Alzheimer's disease is a multifactorial neurodegenerative disorder preceded by a prodromal stage called mild cognitive impairment (MCI). Early diagnosis of MCI is crucial for delaying the progression and optimizing the treatment. In this study we propose a random forest (RF) classifier to distinguish between MCI and healthy control subjects (HC), identifying the most relevant features computed from structural T1-weighted and diffusion-weighted magnetic resonance images (sMRI and DWI), combined with neuro-psychological scores. To train the RF we used a set of 60 subjects (HC = 30, MCI = 30) drawn from the Alzheimer's disease neuroimaging initiative database, while testing with unseen data was carried out on a 23-subjects Mexican cohort (HC = 12, MCI = 11). Features from hippocampus, thalamus and amygdala, for left and right hemispheres were fed to the RF, with the most relevant being previously selected by applying extra trees classifier and the mean decrease in impurity index. All the analyzed brain structures presented changes in sMRI and DWI features for MCI, but those computed from sMRI contribute the most to distinguish from HC. However, sMRI +DWI improves classification performance in training area under the receiver operating characteristic curve (AUROC = 93.5 ± 8%, accuracy = 88.8 ± 9%) and testing with unseen data (AUROC = 93.79%, accuracy = 91.3%), having a better performance when neuro-psychological scores were included. Compared to other classifiers the proposed RF provide the best performance for HC/MCI discrimination and the application of a feature selection step improves its performance. These findings imply that multimodal analysis gives better results than unimodal analysis and hence may be a useful tool to assist in early MCI diagnosis.

1. Introduction

Alzheimer's disease (AD) is a neurodegenerative, incurable illness that causes language, memory and behavior changes, and can lead to loss of functional independence and finally to dementia. It is usually preceded by a prodromal stage called mild cognitive impairment (MCI), where the first behavioral symptoms can be detected through specifically designed neuro-psychological tests. This pathology evolves from biochemical deficiency on tau proteins and beta-amyloids production, that provokes neurofibrillary tangles and neuritic plaques. These early changes are detectable with biomarkers measured in cerebrospinal fluid (CSF) (with higher specificity) as well as in blood plasma. As a consequence, brain atrophy appears in some specific cerebral regions that lose mass, volume and functionality. Anatomical and functional alterations are observable and measurable at this

stage with several imaging modalities, that provide useful biomarkers to diagnose and monitor AD's evolution. Cognitive and behavioral changes become evident in more advanced stages of the disease and are clinically characterized with the help of specific neuro-psychological tests, whose scores can be considered as surrogate parameters. All these biomarkers are useful for diagnosis at different stages of the disease, although definite corroboration can only be obtained *post mortem* after an autopsy. Even though there is no cure for AD, its progression can be delayed if it can be detected early and therefore, the importance of making a diagnosis during the MCI stage has been highlighted by many researchers.

It is widely assumed that incipient protein pathology in the medial temporal lobe initiate the loss of episodic memory in AD, one of the earliest cognitive deficits in this type of dementia. The cingulate cortex projects back to the entorhinal cortex of parahippocampal gyrus (Papez circuit) who in turn, connect with many other areas of the limbic system that are involved in learning and memory, emotion, social behavior, and emotional experience. The loss of episodic memory in early AD reflects a neurodegeneration in the Papez circuit, which critically involves the limbic thalamus (Aggleton *et al* 2016). Thalamic abnormalities generally occur in the early stages of the disease. It has been observed severe tau immunoreactive cytoskeletal pathology in the thalamus of clinically diagnosed AD patients, with the affection of the extraterritorial nuclei of thalamus (Rüb *et al* 2016). In AD and MCI patients, structural connectivity between hippocampus and thalamus affects the functional connectivity between them. The anterior thalamic nuclei in turn connect to the cingulate cortex. The amygdala, along with neocortical areas, are now known to be centrally involved in emotional experience. Thus, the limbic areas in association with neocortical areas, are affected in the neurological deterioration of patients with AD. In this circuit, hippocampal atrophy is one of the most validated, easily accessible and widely used biomarker of AD. Preliminary studies have indicated that looking at this focal atrophy pattern rather than standard whole hippocampal volumetry by fMRI, improves diagnostic accuracy at the MCI stage (de Flores *et al* 2015). These regions show high levels of amyloid deposition in AD and are both structurally and functionally vulnerable early in the disease. Added to this, amygdala may represent a preferential locus for a pivotal transition from a relatively benign clinical condition to a more aggressive disease wherein multiple protein species are misfolded, Nelson *et al* (2018) proposes the amygdala like an incubator for misfolded proteins that participate in the cognitive and emotional damage of patients with AD.

Several studies have reported the diagnostic capability of the available imaging modalities and have proposed different biomarkers to identify the disease' stages. Specifically, accuracies of up to 89% have been obtained using parameters extracted from T1-weighted structural magnetic resonance images (sMRI), that reflect the degree of brain atrophy. In the most recent challenge, the proposals made by Dimitriadis and Liparas (2018) and Jiménez-Mesa *et al* (2020) provided accuracies of 61.9% and 67%. Among other imaging modalities, diffusion weighted MRI (DWI) has been employed in several studies for the extraction of useful biomarkers. The most successful approaches have consisted on the combination of multimodal metrics, particularly sMRI, DWI and DTI, as shown in the researches by Zhang *et al* (2013), Lee *et al* (2017), Jung *et al* (2015), Gupta *et al* (2020), Marzban *et al* (2020), Wen *et al* (2021) attaining accuracies of up to 79%. Particularly, Mesrob *et al* (2012) have proposed a multimodal measure that combines anatomical and diffusivity measures taken at multiple anatomical structures, attaining up to 99% accuracy, when relevant regions were selected. Schouten *et al* (2016) combine anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate AD, with the help of an elastic net classifier. They found that the combination of multiple modalities can substantially improve classification performance over unimodal classification. In another study, the same authors (Schouten *et al* 2017) improved feature selection and found that fractional anisotropy (FA) clustered into ICA components was the best performing measure.

On another aspect, analyses of regional changes in specific brain structures has allowed the identification of different stages of AD's evolution. Studies reported in Li *et al* (2013), Nir *et al* (2013), Lee *et al* (2017), Eldeeb *et al* (2018), Gupta *et al* (2019, 2020), Zarei *et al* (2010) focused on hippocampal, thalamic and other cortical and subcortical regions. Also, multiple studies have demonstrated that an efficient feature selection can be helpful in characterizing AD brain changes (Mesrob *et al* 2012, Jung *et al* 2015, Schouten *et al* 2017). Particularly, Eldeeb *et al* (2018) extracted speeded up robust features and scale invariant feature transform parameters, based on the visual diffusion patterns of FA, and mean diffusivity (MD) maps, to build bag-of-words AD-signature, specifically for the hippocampal area, attaining accuracies of 87% and 89% for FA and MD maps respectively. Jiménez-Mesa *et al* (2020) propose a novel multiclass classification approach that addresses the outlier detection problem, uses pairwise t-test feature selection, project the selected features onto a partial-least-squares multiclass subspace, and applies one-versus-one error correction output codes classification, obtaining an accuracy of 67%. Wen *et al* (2021) developed an open-source framework to evaluate AD classification from T1-weighted MRI, PET and diffusion MRI parameters. They found that FS has a positive impact on classification results, that voxel-wise features generally give better performance than regional features and that FA and MD provided comparable results for voxel-wise features. They conclude that, with proper feature rescaling and selection, the performance of diffusion MRI features is comparable to that of structural MRI.

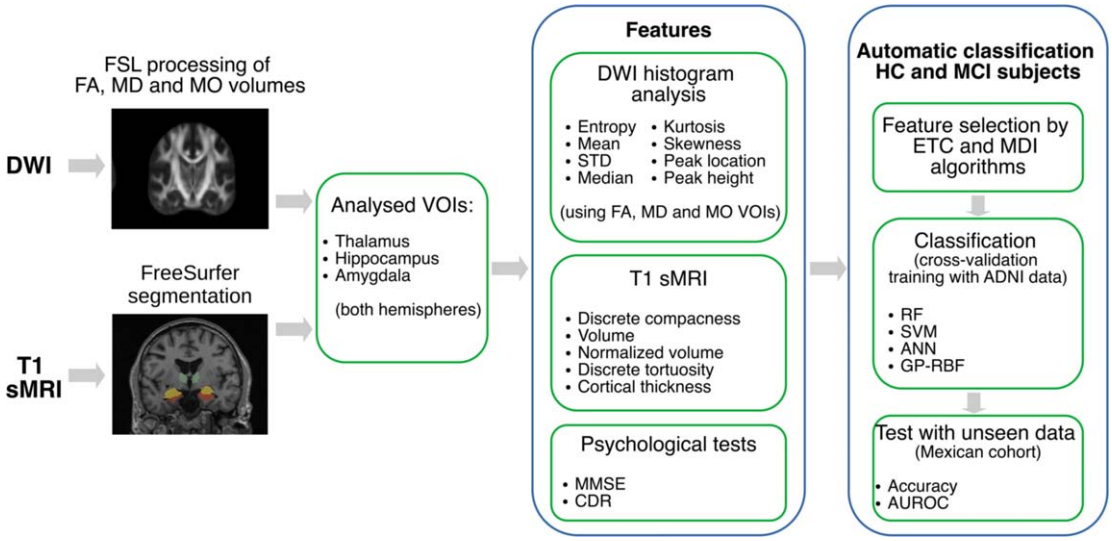


Figure 1. Methodological diagram. DWI—difussion-weighted imaging; sMRI—T1-weighted magnetic resonance imaging; FA—fractional anisotropy; MD—mean diffusivity; MO—mode of anisotropy; MMSE—mini-mental state examination; CDR—clinical dementia rating; HC—healthy controls; MCI—mild cognitive impairment; ETC—extra tree classifier; MDI—mean decrease in impurity; RF—random forest; SVM—support vector machine; ANN—artificial neural network; GP-RBF—Gaussian process with radial basis functions; AUROC—area under the receiver operating characteristic curve.

Table 1. Characteristics of the training and holdout test populations.

	Training populations		Holdout populations	
	(ADNI)		(CI3M)	
	HC ($N = 30$)	MCI ($N = 30$)	HC ($N = 12$)	MCI ($N = 11$)
Age (years) ($\mu \pm \sigma$)	69.0 ± 5.2	68.5 ± 4.3	73.4 ± 6.2	79.9 ± 9.2
Gender (F/M)	15/15	12/18	12/0	8/3
MMSE (rank/$\mu \pm \sigma$)	27–39	24–30	27–30	14–22
	(29.5 ± 0.9)	(27.6 ± 2.5)	(28.6 ± 1.2)	(18.6 ± 2.9)
CDR (rank/$\mu \pm \sigma$)	0	0.5–4	0	0.5–2
		(1.9 ± 1.5)		(1.3 ± 0.6)

In this study, the integration of new morphological metrics, together with information extracted from diffusion imaging is proposed. The latter is analyzed with an original approach, by extracting representative parameters from metabolic histograms measured in specific VOIs. Feature vectors are completed with scores obtained from neuro-psychological tests. The influence of each modality is analyzed individually and in different combinations. The overall pipeline also includes selection of features, that are fed to a random forest (RF) classifier, whose performance is then contrasted with other conventional diagnostic strategies, specifically designed for MCI detection.

2. Methodology

In this section, the steps for volumes' processing, segmentation, feature extraction and classification are described, according to the overall diagram shown in figure 1.

2.1. Studied populations

2.1.1. Alzheimer's disease neuroimaging initiative (ADNI) subjects (training population)

T1-weighted MR images (sMRI) and DWI, as well as mini mental state examination (MMSE) and clinical dementia rating (CDR) scores for 30 control subjects and 30 MCI patients were obtained from the ADNI (adni.loni.usc.edu) database. MMSE and CDR are standard clinical neuro-psychological tests that allow to assess several cognitive areas, such as attention, recall, language and space-time orientation, among others. Subjects were diagnosed by ADNI experts, according to inclusion/exclusion criteria reported in Petersen *et al* (2010), following the ethical statement of the ADNI protocol.

For all subjects, the following information was retrieved:

- T1-weighted MR volumes, obtained at 1.5 T, with voxel resolution of $1.2 \times 0.93 \times 0.93$ mm^3.
- DWI volumes.
- Cognitive scores obtained with both MMSE and CDR tests.
- Age-paired and gender-balanced groups.
- All studies obtained in nearby dates.

These rigorous criteria led to a relatively reduced training population size ($N = 60$), for which the subsequent analysis was carried out following the standard methodology for classification (cross-validation during training) as described later in section 2.6. Table 1 shows demographic information and neuro-psychological scores for both groups (second and third columns).

2.1.2. Mexican cohort (holdout dataset)

For testing with unseen data, two populations of Mexican subjects were conformed. They consisted of 12 healthy control (HC) and 11 MCI subjects, previously diagnosed by an expert neurologist. MCI patients received a clinical diagnosis of Alzheimer-related dementia with MMSE $\leqslant 24$ and CDR $\geqslant 0.5$ scores; subjects with other neurodegenerative conditions were excluded. Both groups were paired by age, socio-cultural and academic levels, with the characteristics shown in the last two columns of table 1. All patients gave their written consent, according to the declaration of Helsinki. Images for these populations were acquired at the Centro Nacional de Investigación en Imagenología e Instrumentación Médica (CI3M) of the Universidad Autónoma Metropolitana, unidad Iztapalapa (UAM-I) in Mexico, after approval of project number PND_AC_08_16.

T1-weighted MR volumes were obtained with the following parameters: TR = 31.92 ms, TE = 1.95 ms, voxel resolution $0.42 \times 0.42 \times 5$ mm^3. DTI were acquired with the following characteristics: b-values of 0 and 800 s mm^{-2} and 32 diffusion-weighted directions. Both studies were obtained using a 3T Phillips equipment.

2.2. Segmentation

sMRI were segmented using Freesurfer v 6.0 to extract gray matter (GM), white matter (WM) and CSF. Also, we have analyzed some of the subcortical substructures that have been reported to be mostly affected in MCI and AD: hippocampus, thalamus and amygdala, for left and right hemispheres.

Image processing begins by removing cranium and other extraencephalic structures to proceed to white and GM segmentations. Afterwards, topological correction, intensity normalization, WM tessellation and surface atlas registration are applied, to finally obtain an individual atlas for each subject, consisting of up to 82 brain structures. Final labeling of the regions of interest is carried out based on the Desikan–Killiany Atlas. Details on the whole segmentation procedure can be found in Fischl *et al* (2004, 2002).

2.3. sMRI biomarkers

For all segmented structures, several anatomical biomarkers were computed: volume (V), determined as the sum of voxels belonging to each VOI; normalized volume (NV), obtained as the ratio between each VOI's volume and intracranial volume, determined as the sum of voxels corresponding to the whole brain; and cortical thickness (CT), calculated as the average of the distance from the WM surface to the closest point on the pial surface (Fischl and Dale 2000).

Additionally, Perez-Gonzalez *et al* (2014) demonstrated that discrete compactness (DC), related to the surface area of those voxels that make contact in 3D, can be a valuable biomarker to identify AD's evolution stages and therefore is useful for its early diagnosis. Therefore, in this study, DC measured on the segmented brain structures has been incorporated into the features vector. Another parameter that quantifies the degree of brain atrophy is surface discrete tortuosity (DT), proposed by Barbará-Morales *et al* (2020) as a valuable discriminator between HC, MCI and AD classes. It is computed using the sum of angles method and is also being considered in this study, to complement structural biomarkers. All morphological features were computed using MATLAB 2020b, except for CT which was calculated with Fresurfer v 6.0.

2.4. Diffusion biomarkers

DWI data were processed using a combination of commands in MRtrix3 package (Tournier *et al* 2019) and the FMRIB Software Library (Jenkinson *et al* 2012). Initial preprocessing steps included denoising (Veraart *et al* 2016), eddy-current and EPI distortion correction and motion correction (Andersson and Sotiropoulos 2016), Gibbs ringing removal (Kellner *et al* 2016) and bias-field correction (Tustison *et al* 2010).

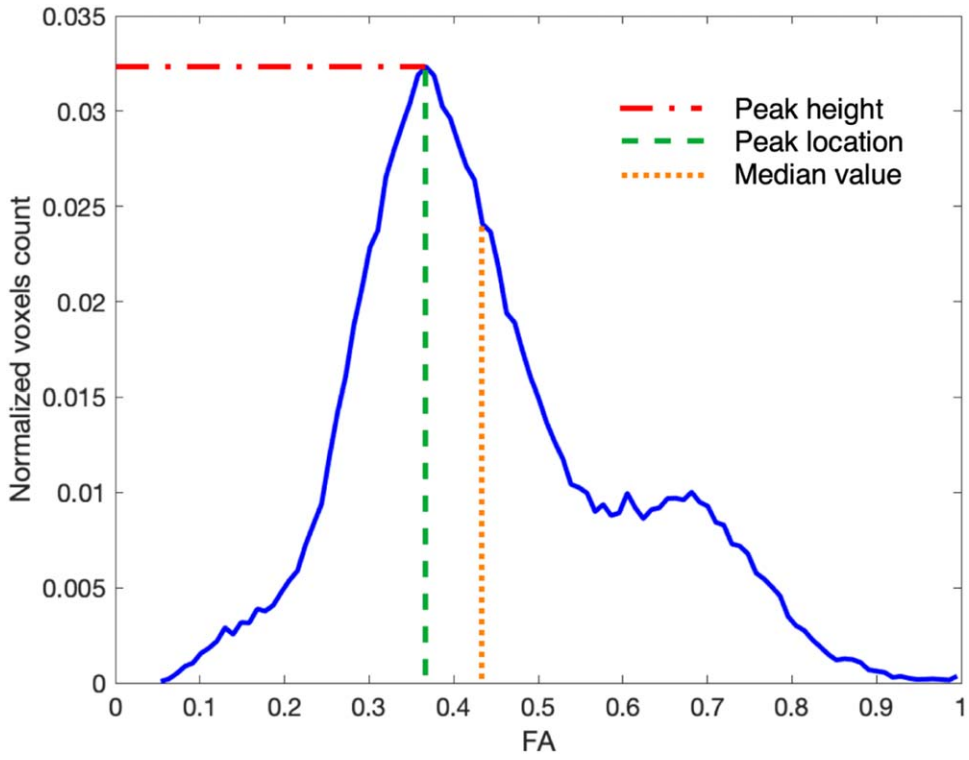


Figure 2. Representative FA histogram of a HC, with the measurement of the peak height in red, peak's location in green and median in yellow color.

Preprocessed DWI data were aligned to b0 image, followed by removing the non-brain tissues using the brain extraction tool (Smith 2002). WM segmentation was based on the subsequent TBSS procedures (Smith *et al* 2007). Eigenvalues of the diffusion tensor model were computed at each voxel to generate rotationally invariant indices like the FA, MD and mode of anisotropy (MO). FA is computed as the normalized standard deviation of the eigenvalues and measures the degree of water diffusion anisotropy, ranging from 0 (isotropic diffusion) to 1 (completely anisotropic diffusion). MD is proportional to the trace of the diffusion tensor and quantifies water molecules diffusivity independently of direction. MO is a measure of anisotropy type, ranging from −1 to +1: negative MO values describe planar anisotropy whereas positive MO values indicate linear anisotropy (Ennis and Kindlmann 2006, Kindlmann *et al* 2007).

All individual FA, MD and MO images were linearly and nonlinearly aligned to FMRIB58_FA template (Douaud *et al* 2011) using the mean FA skeleton that served as the study-specific template and represents the center of the common WM tracts. The FA skeleton was labeled to identify the 50 WM regions with reference to the JHU ICBM-DTI-81 WM atlas (Mori *et al* 2005) and to generate the binary masking images of thalamus, hippocampus, and amygdala.

For each segmented VOI on both hemispheres, normalized histograms of FA, MD and MO were computed and analyzed, following the methodology proposed in Mascalchi *et al* (2018). Normalization, obtained after dividing each histogram by the number of voxels of each structure, allows to correct differences of the subject's brain size. Eight parameters were extracted from normalized histograms: mean (μ, MEAN), standard deviation (σ, STD), as well as skewness (SKEW), kurtosis (KUR) and entropy (ENT) defined by:

$$\mathrm{SKEW} = \frac{\frac{\sum_{i=1}^{n}(x_i-\mu)^3}{n}}{\sigma^3} \tag{1}$$

$$\mathrm{KUR} = \frac{\frac{\sum_{i=1}^{n}(x_i-\mu)^4}{n}}{\sigma^4} \tag{2}$$

$$\mathrm{ENT} = -\sum_{i=0}^{m} p(x_i)\log(p(x_i)), \tag{3}$$

where n is the number of voxels of each structure; x_i is the voxel's intensity for FA, MD and MO; and $p(x_i)$ is a probability estimator of m bins. Additionally, histograms' median value (MED), peak location (LOC) and peak

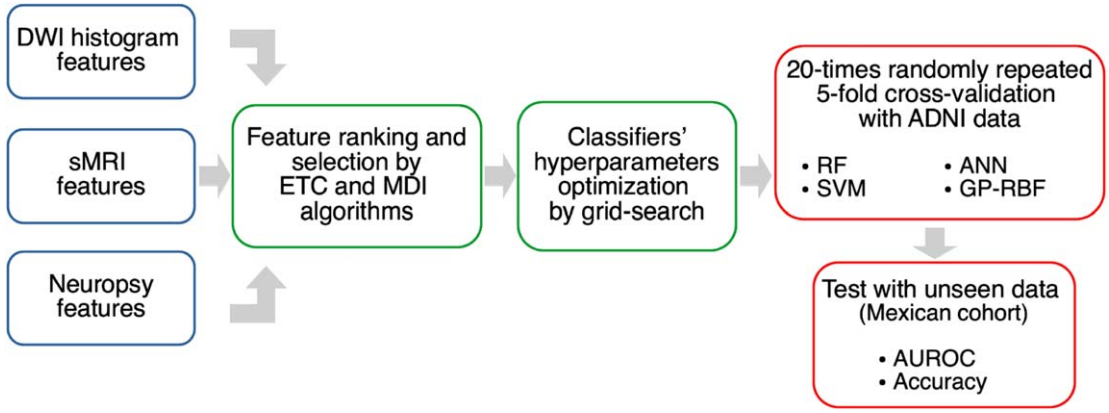


Figure 3. Classification algorithms pipeline between HC and MCI subjects. Blue boxes represent feature extraction stage described in sections 2.3 and 2.4. The steps of feature selection and classifiers' parameters adjustment are shown in green (described in this section). Finally, in red, the validation stages described in section 2.6 are shown.

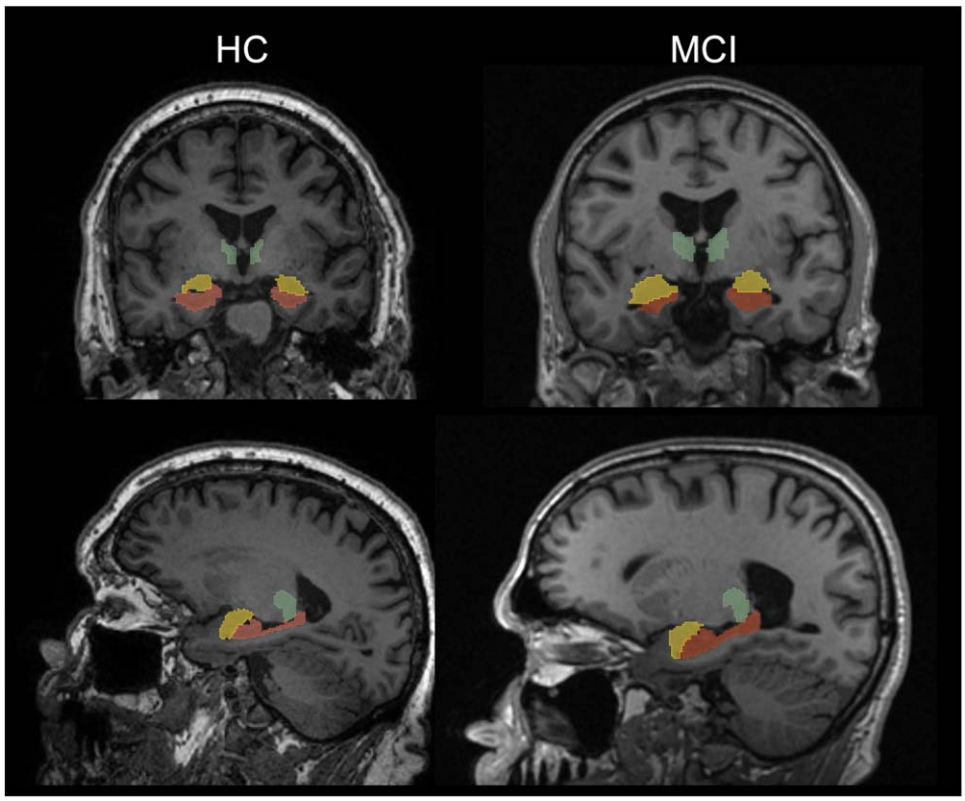


Figure 4. Examples of the segmented structures for HC and MCI subjects: amygdala in yellow, hippocampus in red and thalamus in green.

height (HEIGHT) were determined as shown in figure 2. The algorithm to measure these features was implemented in MATLAB 2020b.

2.5. Classification and feature selection

Automatic classification between HC and MCI subjects was carried out with the help of a RF. It is an algorithm widely used for different medical applications due to its high adaptability to biological data. RF is based on majority voting of multiple decision trees with optimal thresholds to separate classes contained on the training dataset (Criminisi and Shotton 2013).

With the purpose of evaluating the discrimination capability of each data modality, features were fed to the classifier either individually or combined, as follows: morphological biomarkers (sMRI); diffusion biomarkers (DWI); sMRI+DWI; and sMRI+DWI+Neuro-psychological scores (Neuropsy). Additionally, for each feature

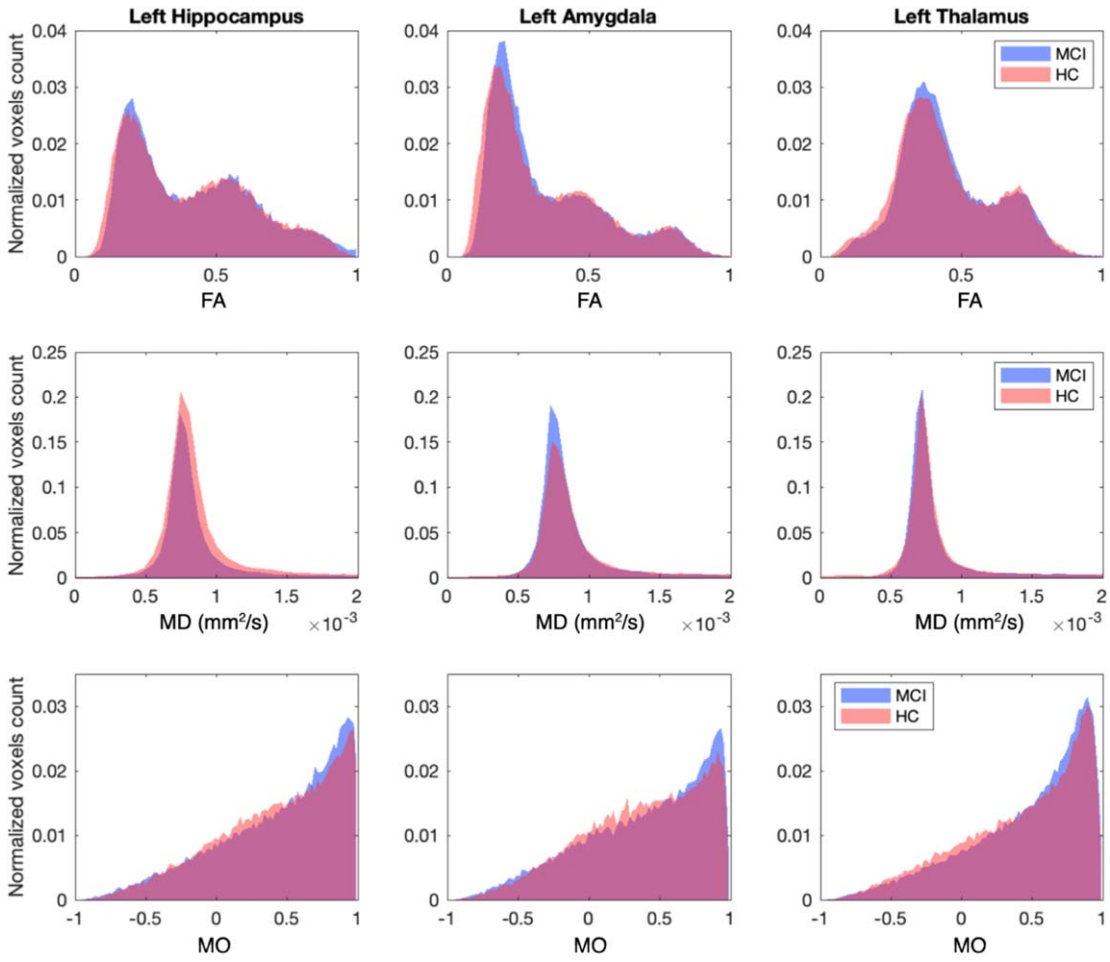


Figure 5. Histogram of DWI measurements on the left hemisphere.

set, the most relevant parameters were selected, and the dimensionality reduced by applying Extra Trees Classifier (ETC), together with the mean decrease in impurity (MDI) index (Tuv *et al* 2009, Breiman 2001). ETC is an ensemble algorithm based on a set of decision trees (500 in this case) with random separation thresholds and bootstrap feature sampling.

For each feature, the MDIs are determined through ETC and sorted in decreasing order. Afterwards, the area under the receiver operating characteristic curve (AUROC) is calculated by gradually aggregating features and the combination that maximizes the AUROC is finally selected.

RF's performance was compared to other conventional classifiers: a support vector machine (SVM), an artificial neural network (ANN) and a radial basis function Gaussian process classifier (GP-RBF). All classifiers' hyper-parameters were optimized during the training stage through grid searching and were used afterwards during the test stage. The list and final values of the tuned parameters are included as supplementary material (available online at stacks.iop.org/PMB/66/155010/mmedia).

Extracted biomarkers were fed to the classification step in different combinations: only sMRI, only DWI, sMRI + DWI and sMRI + DWI + Neuropsy, with and without feature selection, totalling eight tests for each classifier. The most relevant features for each combination are also presented as supplementary material.

In all cases the procedure shown in figure 3 was followed.

2.6. Classifiers' performance

To evaluate classification performance, a 20-times randomly repeated 5-fold cross validation with the training datasets (see section 2.1.1) and a final test with the holdout datasets (see section 2.1.2) were carried out. As described in the previous section, RF results were compared to SVM, ANN and GP-RBF, under the same conditions (same feature vectors, with and without dimensionality reduction). All performances, either in training and in final test stages, were determined with two parameters: AUROC and accuracy (Acc) with a 50% cutoff threshold. Feature selection, classification, optimization and validation algorithms were programmed in Python 3.7 using Scikit-learn 0.024.0 library, and the statistical analysis was performed in R 4.0.5.

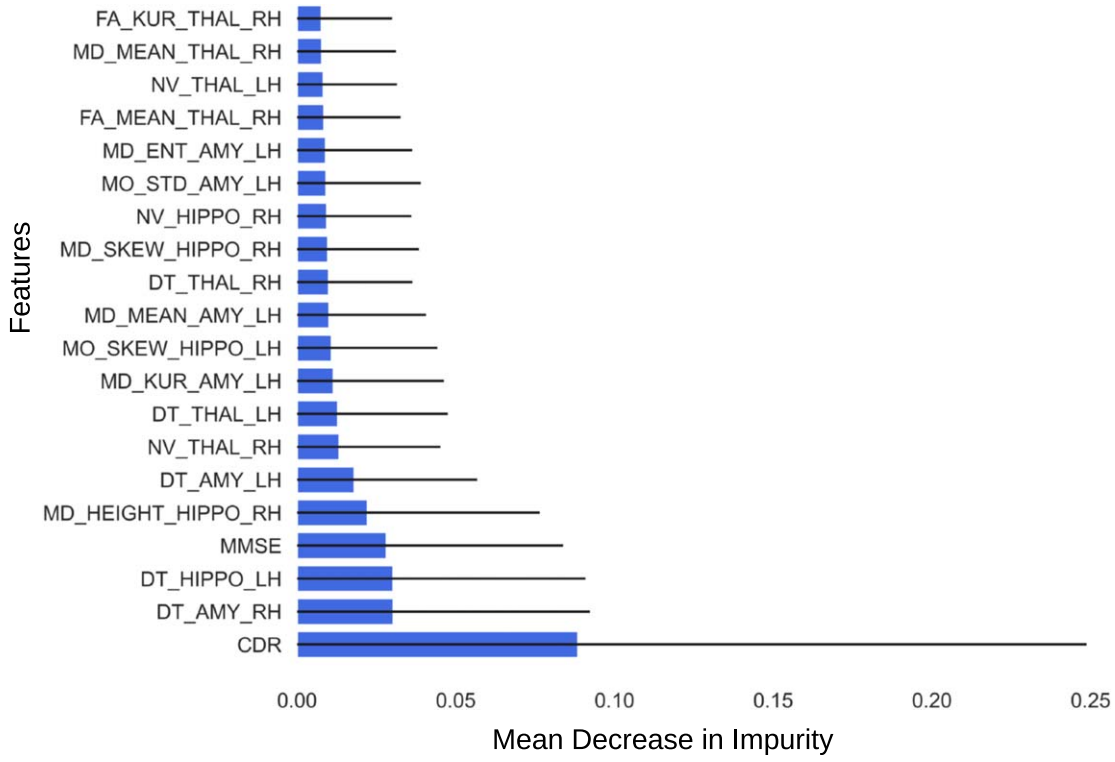


Figure 6. Box plots of the 20 (sMRI + DWI + Neuropsy) most relevant features, selected for classification. AMY = Amygdala, HIPPO = Hippocampus and THAL = Thalamus.

Table 2. AUROC and accuracy (%, $\mu \pm \sigma$) of HC and MCI classification, obtained during the training step, for different biomarkers' combinations, with and without feature selection (FS).

Biomarkers' combination	sMRI		DWI		sMRI + DWI		sMRI + DWI + Neuropsy	
Classifier	Without FS	With FS	Without FS	With FS	Without FS	With FS	Without FS	With FS
				AUROC (%, $\mu \pm \sigma$)				
RF	89.6 ± 8	**92.1 ± 9**	67.3 ± 16	**82.9 ± 12**	84.5 ± 13	**93.5 ± 8**	92.5 ± 8	**99.9 ± 1**
SVM	78.9 ± 11	85 ± 10	72.8 ± 13	73.1 ± 12	81.5 ± 11	91.9 ± 4	86 ± 11	99.1 ± 2
ANN	80.1 ± 11	84 ± 12	60.3 ± 14	77.6 ± 13	77.8 ± 15	88.6 ± 9	85.6 ± 10	98.7 ± 3
GP-RBF	83.3 ± 10	85.8 ± 9	71.7 ± 14	77.8 ± 12	79.9 ± 11	89.6 ± 9	85.8 ± 11	98.4 ± 3
				Accuracy (%, $\mu \pm \sigma$)				
RF	84.1 ± 10	**85.4 ± 9**	65.5 ± 13	**75.7 ± 11**	76.3 ± 13	**88.8 ± 9**	84.6 ± 11	**98.1 ± 5**
SVM	71.1 ± 11	74 ± 12	65 ± 13	68.7 ± 11	77.7 ± 10	83.6 ± 8	80.3 ± 10	94.4 ± 7
ANN	72.9 ± 12	73.3 ± 13	60.3 ± 14	69.9 ± 12	72.6 ± 12	79.3 ± 11	75.9 ± 12	93.6 ± 7
GP-RBF	73.4 ± 11	75.6 ± 10	66.1 ± 10	71.9 ± 11	72.9 ± 10	82.4 ± 10	76.1 ± 10	93.5 ± 6

3. Results

Figure 4 shows examples of the segmented VOIs for control and MCI subjects, in coronal and saggital projections: amygdala in yellow, hippocampus in red and thalamus in green.

In a first exploratory inter-group analysis, box plots and statistical significance levels were determined for each individual biomarker (included as supplementary material). Although all features were considered for the subsequent classification task, this analysis allowed us to depict biomarkers relevant for HC versus MCI discrimination.

As indicated in section 2.4, the histograms of diffusion biomarkers for all structures and both hemispheres, were computed to extract representative features from this modality. Figure 5 shows the distributions obtained for FA, MO and MD of the studied brain structures, corresponding to the left hemisphere as an example. Histograms for the right hemisphere show a similar distribution. Differences can be observed between HC and

Table 3. AUROC and accuracy (%) of HC and MCI classification, obtained during testing with the holdout dataset, for different biomarkers' combinations, with and without feature selection (FS).

Biomarkers' combination	sMRI		DWI		sMRI + DWI		sMRI + DWI + Neuropsy	
Classifier	Without FS	With FS	Without FS	With FS	Without FS	With FS	Without FS	With FS
				AUROC (%)				
RF	77.94	**87.22**	82.58	**87.12**	76.78	**93.79**	82.08	**98.29**
SVM	72.79	85.95	74.24	79.24	74.46	91.54	80.45	98.13
ANN	64.54	75.65	81.81	82.57	65.83	87.97	80.76	95.32
GP-RBF	66.57	78.34	71.21	84.09	73.15	88.76	81.1	97.41
				Accuracy (%)				
RF	69.57	**82.61**	82.61	**86.96**	86.96	**91.30**	82.61	**100**
SVM	69.57	73.91	60.87	78.26	73.91	86.96	78.26	**100**
ANN	60.87	65.22	73.91	82.61	69.57	82.61	73.91	91.30
GP-RBF	65.22	69.57	65.21	82.61	65.22	82.61	78.26	95.65

MCI, which suggests that parameters extracted from these histograms can be useful in the subsequent classification step, as discussed later.

Results of the repetitive cross-validation, carried out during training, are shown in table 2. Final tests were conducted with unseen data, as described in section 2.6. Table 3 shows the corresponding results. For both tables, the eight tested combinations appear in columns, while the compared classifiers correspond to each row. The highest performances for each biomarkers' combination are highlighted in bold.

In order to identify those biomarkers that contribute the most in enhancing the AUROC, they were ordered accordingly to their MDI, as separate modalities (sMRI and DWI) and in combination (sMRI+DWI, sMRI +DWI+Neuropsy). These rankings are included as supplementary material. As a particular case, figure 6 shows the 20 most relevant parameters of sMRI+DWI combination, where it can be observed that both modalities influence classification and that the proposed biomarkers appear as relevant features.

4. Discussion

The complete statistical analysis (see supplementary material), carried out to depict between-group changes in the proposed biomarkers' set, showed that thalamus suffers important morphological differences reflected by the NVs of both hemispheres and DT of the left side. Also, the three diffusion biomarkers measured in this structure were significantly different between control and MCI subjects. Regarding the amygdala, surface tortuosity and MD show differences on both hemispheres, as does the hippocampus, that also presents changes on tortuosity (only on the left side) and overall MD. These results concur with those obtained by Li *et al* (2013), Lee *et al* (2017) and Coupé *et al* (2019) that have reported microstructural and diffusivity changes in these structures, along the progression of the disease. Even though not all biomarkers indicate changes in the analyzed structures, they were all included as a global feature vector, trying to maximally exploit the complementarity of the morphological and diffusion information that they contain. Among the advantages of the proposed morphological biomarkers, the translation, rotation and scaling invariance of DC, as well as the angle's normalization applied to DT, contribute to the adequate characterization of morphological changes of the analyzed structures, regardless of subject's brain size and position, such as reported by Perez-Gonzalez *et al* (2014) and Barbará-Morales *et al* (2020).

A qualitative analysis carried out on FA, MO and MD histograms (figure 5) allows to appreciate differences between the two populations' distributions and suggests that all parameters that can synthesize these behaviors are potential biomarkers to facilitate groups' classification. In the research reported by Mascalchi *et al* (2018) the authors had already demonstrated that histogram-derived indicators reflect microstructural changes in brain degeneration for another clinical application. In our study, we have proposed an expanded set of biomarkers that constitute an enhanced feature vector derived from DWI. Unlike voxel-based morphometry, the analysis based on histogram-extracted parameters allows the comparison between structures, by only aligning the reference atlas to each individual, without the need of inter-subject volumes' registration.

As previously indicated, the contribution of the proposed biomarkers, either in individual or combined modalities, can be seen in table 2 for the training step. A first observation indicates that, compared to other classifiers, RFs provide the best performance for HC/MCI discrimination, given the biomarker's vector proposed in this research. RF and SVM classifiers had already been demonstrated to be suitable for this task, as

reported in several studies, using other feature sets (Tanveer *et al* 2020). A second finding indicates that the application of a feature selection step prior to classification, considerably improves its performance, as concluded by Sivapriya *et al* (2015) and by Wen *et al* (2021) who also reported that with proper dimensionality reduction, the performance of diffusion MRI features is comparable to that of sMRI. The analysis of individual modalities seems to indicate that morphological biomarkers (1st and 2nd columns in table 2) contribute the most to identify both groups, in comparison to DWI-extracted parameters (3rd and 4th columns). However, the combination of the two modalities (5th and 6th columns) improves classification performance, up to $93.5 \pm 8\%$ AUROC and $88.8 \pm 9\%$ accuracy in our study, using RF. This is consistent with findings reported by Gupta *et al* (2020) (AUROC $= 0.96$) and Zhang *et al* (2013) (AUROC $= 94.9\%$). Finally, in the two last columns of table 2 it can be observed that, when incorporating neuro-psychological scores into the feature vector, classification performances substantially improve. This is expected, due to the contribution of limbic structures in patients' cognitive decline. Also, MMSE and CDR scores are considered as determinant criteria to diagnose cases in the ADNI database. Again, it can be appreciated that a previous selection of features provides the best RF performance.

A similar behavior was observed during classifiers' final testing with unseen data (table 3), obtaining 93.79% AUROC and 91.3% accuracies, when structural and diffusion biomarkers were combined (6th column). These results outperform those obtained by Marzban *et al* (2020), who use deep learning for the same task (AUROC $= 0.84$, accuracy of 79.6%); by Sheelakumari *et al* (2018) (AUROC $= 0.89$) that also consider spectroscopy indicators; and by Jung *et al* (2015) that attain an accuracy of 84.4% in HC/MCI classification. This demonstrates that a previous feature selection (ETC) and an adequate classification strategy (RF) are necessary for this application.

Features that resulted more relevant for classification (figure 6) include parameters extracted from both modalities, such as right-amygdala and left-hippocampus DT peak's height of the MD for the right-hippocampus. Moreover, it can be seen that all the analyzed structures (amygdala, hippocampus and thalamus) resulted relevant in structural parameters (DT and volumes) as well as in diffusion biomarkers (MD, MO and FA). This corroborates the previous statistical analysis and concours with findings of Zarei *et al* (2010), Lee *et al* (2017) and Li *et al* (2013) that report that diffusivity is strongly correlated with the cognitive performance.

5. Conclusions

In this study, a robust method for the automatic classification of HC and MCI subjects, using anatomical (sMRI) and diffusion (DWI) magnetic resonance volumes was presented. An enhanced feature set was proposed, that includes shape-invariant morphological biomarkers, such as DC and tortuosity. Also, a strategy to extract biomarkers from DWI histograms is incorporated, that proved to be useful to recover microstructural brain changes between studied populations. The analysis focused on specific limbic structures, that are involved in MCI, such as amygdala, hippocampus and thalamus.

The combination of an adequate classifier, together with a previous feature selection step provided the best performance. In this study, given the proposed feature set, RF turned out to be the most adequate for populations' separation. The discrimination between MCI and HC subjects reached a high accuracy when the combination of sMRI and DWI biomarkers for specific brain regions (amygdala, thalamus and hippocampus) of both brain hemispheres, and further when neuro-psychological scores were included. These results imply that multimodal analysis gives better results than unimodal analysis and hence may be a useful tool to assist in early MCI diagnosis, attaining an AUROC of 93.79% and an accuracy of 91.3% during final testing with the holdout set.

As a future work, a more detailed analysis of MCI's involved substructures, the proposal of other more specific biomarkers, as well as the application of deep learning methods to improve diagnosis, would be necessary. Also, the incorporation of other diagnostic modalities, such as biochemical or functional biomarkers can help to better identify MCI condition in an early stage.

Acknowledgments

This work was supported by UNAM-PAPIIT IA102920. Authors also wish to acknowledge technical assistance of Adrian Duran Chavesti for server management used for data processing.

Data collection and sharing for this project was funded by the 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: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir,

Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. 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 (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

ORCID iDs

Jorge Perez-Gonzalez https://orcid.org/0000-0002-4069-4268
Luis Jiménez-Ángeles https://orcid.org/0000-0002-7390-0814
Verónica Medina-Bañuelos https://orcid.org/0000-0001-8072-3927

References

Aggleton J P, Pralus A, Nelson A J D and Hornberger M 2016 Thalamic pathology and memory loss in early alzheimer's disease: moving the focus from the medial temporal lobe to papez circuit *Brain: J. Neurol.* **139** 1877–90

Andersson J L and Sotiropoulos S N 2016 An integrated approach to correction for off-resonance effects and subject movement in diffusion mr imaging *NeuroImage* **125** 1063–78

Barbará-Morales E, Pérez-González J, Rojas-Saavedra K C and Medina-Bañuelos V 2020 Evaluation of brain tortuosity measurement for the automatic multimodal classification of subjects with alzheimer's disease *Comput. Intell. Neurosci.* **2020** 1–11

Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32

Coupé P, Manjón J V, Lanuza E and Catheline G 2019 Lifespan changes of the human brain in alzheimer's disease *Sci. Rep.* **9** 1–12

Criminisi A and Shotton J 2013 *Decision Forests for Computer Vision and Medical Image Analysis* (Berlin, London: Springer)

de Flores R, La Joie R and Chételat G 2015 Structural imaging of hippocampal subfields in healthy aging and alzheimer's disease *Neuroscience* **309** 29–50

Dimitriadis S I and Liparas D 2018 How random is the random forest? random forest algorithm on the service of structural imaging biomarkers for alzheimer's disease: from alzheimer's disease neuroimaging initiative (adni) database *Neural Regener. Res.* **13** 962–70

Douaud G *et al* 2011 Dti measures in crossing-fibre areas: Increased diffusion anisotropy reveals early white matter alteration in mci and mild alzheimer's disease *NeuroImage* **55** 880–90

Eldeeb G W, Zayed N and Yassine I A 2018 Alzheimer's disease classification using bag-of-words based on visual pattern of diffusion anisotropy for dti imaging *2018 40th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC) (Honolulu, HI, 18–21 July 2018)* (Picastaway, NJ: IEEE) pp 57–60

Ennis D B and Kindlmann G 2006 Orthogonal tensor invariants and the analysis of diffusion tensor magnetic resonance images *Magn. Reson. Med.* **55** 136–46

Fischl B and Dale A M 2000 Measuring the thickness of the human cerebral cortex from magnetic resonance images *Proc. Natl Acad. Sci.* **97** 11050–5

Fischl B *et al* 2002 Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain *Neuron* **33** 341–55

Fischl B *et al* 2004 Automatically parcellating the human, cerebral cortex *Cerebral Cortex* **14** 11–22

Gupta Y, Kim J-I, Kim B C and Kwon G-R 2020 Classification and graphical analysis of alzheimer's disease and its prodromal stage using multimodal features from structural, diffusion, and functional neuroimaging data and the apoe genotype *Frontiers Aging Neurosci.* **12** 1–33

Gupta Y, Lama R K, Kwon G-R and the Alzheimer's Disease Neuroimaging Initiative 2019 Prediction and classification of alzheimer's disease based on combined features from apolipoprotein-e genotype, cerebrospinal fluid, mr, and fdg-pet imaging biomarkers *Front. Comput. Neurosci.* **13** 72

Jenkinson M, Beckmann C F, Behrens T E, Woolrich M W and Smith S M 2012 Fsl *NeuroImage* **62** 782–90

Jiménez-Mesa C, Illán I A, Martín-Martín A, Castillo-Barnes D, Martinez-Murcia F J, Ramírez J and Górriz J M 2020 Optimized one vs one approach in multiclass classification for early alzheimer's disease and mild cognitive impairment diagnosis *IEEE Access* **8** 96981–93

Jung W B, Lee Y M, Kim Y H and Mun C W 2015 Automated classification to predict the progression of alzheimer's disease using whole-brain volumetry and dti *Psychiatry Invest.* **12** 92–102

Kellner E, Dhital B, Kiselev V G and Reisert M 2016 Gibbs-ringing artifact removal based on local subvoxel-shifts *Magn. Reson. Med.* **76** 1574–81

Kindlmann G, Ennis D B, Whitaker R T and Westin C 2007 Diffusion tensor analysis with invariant gradients and rotation tangents *IEEE Trans. Med. Imaging* **26** 1483–99

Lee P, Ryoo H, Park J and Jeong Y 2017 Morphological and microstructural changes of the hippocampus in early mci: a study utilizing the alzheimer's disease neuroimaging initiative database *J. Clin. Neurol.* **13** 144–54

Li Y-D, Dong H-B, Xie G-M and Zhang L-J 2013 Discriminative analysis of mild alzheimer's disease and normal aging using volume of hippocampal subfields and hippocampal mean diffusivity: an in vivo magnetic resonance imaging study *Am. J. Alzheimer's Dis. Other Dementias* **28** 627–33

Marzban E N, Eldeib A M, Yassine I A, Kadah Y M and for the Alzheimer's Disease Neurodegenerative Initiative 2020 Alzheimer's disease diagnosis from diffusion tensor images using convolutional neural networks *PLoS One* **15** e0230409

Mascalchi M *et al* 2018 Histogram analysis of DTI-derived indices reveals pontocerebellar degeneration and its progression in SCA2 *PLoS One* **13** e0200258

Mesrob L, Sarazin M, Hahn-Barma V, de Souza L C, Dubois B, Gallinari P and Kinkingnhun S 2012 Dti and structural mri classification in alzheimers disease *Adv. Mol. Imaging* **02** 12–20

Mori S, Wakana S, van Zijl P C M and Nagae-Poetscher L M 2005 *MRI Atlas of Human White Matter* 1st edn (Amsterdam, Netherlands: Elsevier Science) https://www.elsevier.com/books/mri-atlas-of-human-white-matter/mori/978-0-444-51741-8

Nelson P T *et al* 2018 The amygdala as a locus of pathologic misfolding in neurodegenerative diseases *J. Neuropathol. Exp. Neurol.* **77** 2–20

Nir T M, Jahanshad N, Villalon-Reina J E, Toga A W, Jack C R, Weiner M W and Thompson P M 2013 Effectiveness of regional dti measures in distinguishing alzheimer's disease, mci, and normal aging *NeuroImage: Clin.* **3** 180–95

Perez-Gonzalez J L, Yanez-Suarez O, Bribiesca E, Cosío F A, Jiménez J R and Medina-Bañuelos V 2014 Description and classification of normal and pathological aging processes based on brain magnetic resonance imaging morphology measures *J. Med. Imaging* **1** 1–7

Petersen R C *et al* 2010 Alzheimer's disease neuroimaging initiative (adni): Clinical characterization *Neurology* **74** 201–9

Rüb U, Stratmann K, Heinsen H, Del Turco D, Ghebremedhin E, Seidel K, den Dunnen W and Korf H-W 2016 Hierarchical distribution of the tau cytoskeletal pathology in the thalamus of alzheimer's disease patients *J. Alzheimers Dis.* **49** 905–15

Schouten T M, Koini M, de Vos F, Seiler S, de Rooij M, Lechner A, Schmidt R, van den Heuvel M, van der Grond J and Rombouts S A 2017 Individual classification of alzheimer's disease with diffusion magnetic resonance imaging *NeuroImage* **152** 476–81

Schouten T M *et al* 2016 Combining anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate alzheimer's disease *NeuroImage: Clin.* **11** 46–51

Sheelakumari R, Sarma S P, Kesavadas C, Thomas B, Sasi D, Sarath L V, Justus S, Mathew M and Menon R N 2018 Multimodality neuroimaging in mild cognitive impairment: a cross-sectional comparison study *Ann. Indian Acad. Neurol.* **21** 133–9

Sivapriya T R, Nadira Banu Kamal A R and Jeba Thangaiah P R 2015 Ensemble merit merge feature selection for enhanced multinomial classification in Alzheimer's dementia *Comput Math Methods Med.* **2015** 676129

Smith S M 2002 Fast robust automated brain extraction *Hum. Brain Mapp.* **17** 143–55

Smith S M *et al* 2007 Acquisition and voxelwise analysis of multi-subject diffusion data with tract-based spatial statistics *Nat. Protocols* **2** 499–503

Tanveer M, Richhariya B, Khan R U, Rashid A H, Khanna P, Prasad M and Lin C T 2020 Machine learning techniques for the diagnosis of alzheimer's disease: a review *ACM Trans. Multimedia Comput. Commun. Appl.* **16** 1–35

Tournier J-D, Smith R, Raffelt D, Tabbara R, Dhollander T, Pietsch M, Christiaens D, Jeurissen B, Yeh C-H and Connelly A 2019 Mrtrix3: a fast, flexible and open software framework for medical image processing and visualisation (https://doi.org/10.1016/j.neuroimage.2019.116137)

Tustison N J, Avants B B, Cook P A, Zheng Y, Egan A, Yushkevich P A and Gee J C 2010 N4itk: Improved n3 bias correction *IEEE Trans. Med. Imaging* **29** 1310–20

Tuv E, Borisov A, Runger G and Torkkola K 2009 Feature selection with ensembles, artificial variables, and redundancy elimination *J. Mach. Learn. Res.* **10** 1341–66

Veraart J, Fieremans E and Novikov D S 2016 Diffusion mri noise mapping using random matrix theory *Magn. Reson. Med.* **76** 1582–93

Wen J *et al* 2021 Reproducible evaluation of diffusion mri features for automatic classification of patients with Alzheimer's disease *Neuroinform* **19** 57–78

Zarei M, Patenaude B, Damoiseaux J, Morgese C, Smith S, Matthews P M, Barkhof F, Rombouts S, Sanz-Arigita E and Jenkinson M 2010 Combining shape and connectivity analysis: an mri study of thalamic degeneration in alzheimer's disease *NeuroImage* **49** 1–8

Zhang Y *et al* 2013 Mri markers for mild cognitive impairment: Comparisons between white matter integrity and gray matter volume measurements *PLoS One* **8** 1–10