Deformation Based Feature Selection for Computer Aided Diagnosis of Alzheimer's Disease

Alexandre Savio[1], Manuel Graña[1]

[1] *Computational Intelligence Group, UPV/EHU*

Abstract

Deformation-based Morphometry (DBM) allows detection of significant morphological differences of brain anatomy, such as those related to brain atrophy in Alzheimer's Disease (AD). DBM process is as follows: First, performs the non-linear registration of a subject's structural MRI volume to a reference template. Second, computes scalar measures of the registration's deformation field. Third, performs across volume statistical group analysis of these scalar measures to detect effects. In this paper we use the scalar deformation measures for Computer Aided Diagnosis (CAD) systems for AD. Specifically this paper deals with feature extraction methods over five such scalar measures. We evaluate three supervised feature selection methods based on voxel site significance measures given by Pearson correlation, Bhattacharyya distance and Welch's t-test, respectively. The CAD system discriminating between healthy control subjects (HC) and AD patients consists of a Support Vector Machine (SVM) classifier trained on the DBM selected features. The paper reports experimental results on structural MRI data from the cross-sectional OASIS database. Average 10-fold cross-validation classification results are comparable or improve the state-of-the-art results of other approaches performing CAD from structural MRI data. Localization in the brain of the most discriminant deformation voxel sites is in agreement with findings reported in the literature.

1. Introduction

Alzheimer's Disease (AD) is a neurodegenerative disorder which is one of the most common cause of dementia in old people in developed countries [1]. Due to the socioeconomic importance of the disease in occidental countries there is a strong international scientific effort focused in AD. The diagnosis of AD can be done after the exclusion of other forms of dementia but a definitive diagnosis can only be made after a post-mortem study of the brain tissue. Antemortem approaches for diagnosis of AD are under development, but they require neuropathologic confirmation of the characteristic amyloid plaques and neurofibrillary tangles [28]. There has been a strong effort directed to the assessment of medical imaging biomarkers identifying the disease at various stages [14, 10, 18, 19, 27].

Deformation-based Morphometry (DBM) [23, 35] studies aim to detect morphological changes by statistical analysis of deformation scalar measures computed from the Jacobian matrix of the deformation field obtained by non-linear registration with high degrees of freedom of the study's subjects to a template. It allows a comprehensive measurement of structural differences within or across groups, not only in specific structures but throughout the entire brain. In this paper, these scalar measures are the raw data for the construction of the classifier systems.

Computer Aided Diagnosis (CAD) systems apply machine learning techniques to build classifier systems that may help the clinician to come up with a diagnosis on the basis of available data. Most CAD systems applied to imaging data, consist of a spatial normalization (aka registration) of the images to a standard template, the extraction of features from the image and the classification of these features into the target classes. Support Vector Machines (SVM) [36] have become a standard for the construction of such CAD classifiers. CAD systems for AD have been proposed and tested for PET and SPECT data [4, 26, 11], structural Magnetic Resonance Imaging (sMRI) data [9, 12, 32, 21], Difusion Tensor Imaging (DTI) [13]. A critical aspect of these systems is the validation procedure. Circularity analysis effects have been pointed out [22] in neuroscience reports. To avoid this issue, all training procedures must be performed only on the training data. Therefore, feature selection processes must be repeated for each training experiment on the pertinent training data. We have adhered to this procedure in this paper. CAD system construction may have the side effect of identifying voxel locations where discriminant data are extracted. These localizations can be compared with the current findings in the literature.

The contribution of this paper is the proposal of scalar measures of local deformations computed from the Jacobian matrices of the deformation field at each voxel, followed by a supervised feature selection process identifying the most significant voxel sites to extract the feature data for the classifier construction and test. Feature selection is performed using three significance measures, Pearson's correlation, Bhattacharyya distance and the Welch's t-test statistic. Experimental results obtained on the OASIS database of sMRI volumes of AD and Healthy Controls (HC) show that the approach gives state of the art results.

Some preliminar results were reported in the KES 2012 conference [31]. The paper is structured as follows: Section 2 describes the registration procedure and the computation of the deformation scalar measures. Section 3 describes the feature selection processes. Section 4 describes the learning classifier system. Section 5 describes the experimental settings. Section 6 gives the results. Finally, section 7 gives some conclusions.

2. Features from Deformation Based Morphometry

2.1. MRI spatial registration

There is a number of image registration methods reported in the neuroimaging literature [20], the work reported here follows a small deformation model.

A general expression of the non-linear registration coordinate transformation in the small deformation model is as follows:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = M \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} + \begin{bmatrix} d_x(x,y,z) \\ d_y(x,y,z) \\ d_z(x,y,z) \\ 1 \end{bmatrix} \tag{1}$$

where M is an affine transformation matrix, denoting the linear part of the transformation involving scaling, rotation and translation. The term $D = [d_x(x,y,z), d_y(x,y,z), d_z(x,y,z), 1]^T$ denotes the non-linear components of the transformation, the non-linear displacement field, giving the displacement of the registered volume voxel location (x,y,z) in each direction that matches optimally with the reference template volume. Let us denote $g(x,y,z)$ the MRI intensity value, defined as long as x, y and z fall within the original image sampling grid. Therefore, $g(x',y',z')$ denotes the value of g for a set of coordinates after the transformation given by eq. (1). Function $g(x,y,z)$ involves some interpolation procedure allowing to compute values of coordinates that did not coincide with original MRI data voxel sites.

Each component $d_i(x,y,z)$ of the displacement field, where $i \in \{x,y,z\}$ denotes the axis, is built as a linear combination of basis functions, i.e.

$$d_i(x,y,z) = \sum_{l,m,n} c_{l,m,n} B_{l,m,n}(x,y,z), \tag{2}$$

where $B_{l,m,n}(x,y,z)$ is a basis function, i.e. a 3D cubic spline, value at spatial location $[x,y,z]$. It is notationally convenient to gather the basis coefficients of each displacement field into a vector $\mathbf{w}_i$, furhter concatenating all of them into a global parameter vector: $\mathbf{w} = \left[\mathbf{w}^{(x)T}, \mathbf{w}^{(y)T}, \mathbf{w}^{(z)T}\right]^T$. Then, $g_{xyz}(\mathbf{w})$ denotes the value of g at (x',y',z') under the transform determined by $\mathbf{w}$, where M is implicit. This general notation is independent of our choice of basis-function and would be equally valid for e.g. a discrete cosine basis set. The gradient of g at point (x',y',z') is defined as

$$\nabla g_{xyz}(\mathbf{w}) = \left[\left.\frac{\partial g_{xyz}}{\partial x}\right|_{\mathbf{w}} \left.\frac{\partial g_{xyz}}{\partial y}\right|_{\mathbf{w}} \left.\frac{\partial g_{xyz}}{\partial z}\right|_{\mathbf{w}} \right] \tag{3}$$

where the partial derivative with respect to e.g. x, $\left.\partial g_{xyz}/\partial x\right|_{\mathbf{w}}$, denotes the rate of change of g at (x',y',z') as one translates the sampling point in the x-direction. The exact expression of the partial derivatives depends on the interpolating function invoked to compute $g(x',y',z')$, and on the basis functions used to model the displacement fields. Assuming that parameter $w_i^{(x)}$ is the coefficient for the lmn-th 3D spline basis function composing the x-displacement field $d_x$we can write

$$\left.\frac{\partial g_{xyz}}{\partial w_i^{(x)}}\right|_{\mathbf{w}} = \left.\frac{\partial g_{xyz}}{\partial x}\right|_{\mathbf{w}} B_{lmn}(x,y,z), \tag{4}$$

where the $\mathbf{w}$ subscript indicates that the derivative has been calculated at point $\mathbf{w}$ in parameter space, and i identifies the lmn-th 3D spline basis function. To obtain a matrix expression, the vector of signal values obtained after registration with the deformation fields determined by $\mathbf{w}$ is:

$$\mathbf{g}(\mathbf{w}) = \left[g_{x,y,z}(\mathbf{w}); x \in \{1,\ldots,X\}, y \in \{1,\ldots,Y\}, z \in \{1,\ldots,Z\}\right]^T. \quad (5)$$

Analogously,

$$\left.\frac{\partial g_{xyz}}{\partial x}\right|_w = \left[\left.\frac{\partial g_{x,y,z}}{\partial x}\right|_{\mathbf{w}}; x \in \{1,\ldots,X\}, y \in \{1,\ldots,Y\}, z \in \{1,\ldots,Z\}\right]^T \quad (6)$$

and

$$\mathbf{B}_{lmn} = \left[B_{lmn}(x,y,z); x \in \{1,\ldots,X\}, y \in \{1,\ldots,Y\}, z \in \{1,\ldots,Z\}\right]^T. \quad (7)$$

We can write $\mathbf{B}_{lmn}$ as the vector representation of the i-th basis function $\mathbf{B}_i$. Combining equations (4), (6) and (7) we can define the matrix

$$\mathbf{J}_x(\mathbf{w}) = \left[\left.\frac{\partial g}{\partial x}\right|_{\mathbf{w}} \odot \mathbf{B}_1 \;\; \left.\frac{\partial g}{\partial x}\right|_{\mathbf{w}} \odot \mathbf{B}_2 \;\ldots\; \left.\frac{\partial g}{\partial x}\right|_{\mathbf{w}} \odot \mathbf{B}_{LMN}\right], \quad (8)$$

where $\odot$ denotes elementwise Hadamard matrix product. Each element of $\mathbf{J}_x(\mathbf{w})$ is of the form given by equation (4). Finally, the Jacobian matrix $\mathbf{J}$ of the mapping $\mathbf{g}(\mathbf{w}) : \mathbb{R}^{3LMN} \rightarrow \mathbb{R}^{XYZ}$ is composed as:

$$\mathbf{J}(\mathbf{w}) = \left[\mathbf{J}_x(\mathbf{w})\ \mathbf{J}_y(\mathbf{w})\ \mathbf{J}_z(\mathbf{w})\right], \quad (9)$$

where $\mathbf{J}_y(\mathbf{w})$ and $\mathbf{J}_z(\mathbf{w})$ are computed analogously to $\mathbf{J}_x(\mathbf{w})$, $\mathbf{J}(\mathbf{w})$ is of size $XYZ \times 3LMN$, and LMN and XYZ are the number of 3D spline basis functions and the number of samples in the volume, respectively.

The estimation of the parameters $\mathbf{w}$ determining the displacement fields $d_i(x,y,z)$ that transform the image native space into a reference space is performed minimizing an objective function

$$\widehat{\mathbf{w}} = \arg\min_{\mathbf{w}} O(\mathbf{w}) \quad (10)$$

measuring the matching between the warped image $\mathbf{g}(\mathbf{w})$ and the reference template $\mathbf{f} \in \mathbb{R}^{XYZ}$:

$$O(\mathbf{w}) = \frac{1}{XYZ}\left(\mathbf{g}(\mathbf{w}) - \mathbf{f}\right)^T\left(\mathbf{g}(\mathbf{w}) - \mathbf{f}\right). \quad (11)$$

This minimisation is performed by a Levenberg-Marquardt procedure, using the Gauss-Newton approximation of the Hessian matrix of the objective function

$$H_{ij}(\mathbf{w}) = \left.\frac{\partial^2 O}{\partial w_i \partial w_j}\right|_{\mathbf{w}} \quad (12)$$

given by

$$H(w) = \frac{2}{XYZ}\mathbf{J}^T(\mathbf{w})\,\mathbf{J}(\mathbf{w}).$$

2.2. Deformation scalar measures

The determinant of the Jacobian matrix of the deformation field is the most commonly used scalar measure of deformation for Tensor-based Morphometry (TBM) analyses Lepore et al. [23]. The Jacobian matrix of the nolinear deformation field $D = [d_x(x,y,z), d_y(x,y,z), d_z(x,y,z), 1]^T$ is given by

$$J = \begin{bmatrix} \frac{\partial(x-d_x)}{\partial x} & \frac{\partial(x-d_x)}{\partial y} & \frac{\partial(x-d_x)}{\partial z} \\ \frac{\partial(y-d_y)}{\partial x} & \frac{\partial(y-d_y)}{\partial y} & \frac{\partial(y-d_y)}{\partial z} \\ \frac{\partial(z-d_z)}{\partial x} & \frac{\partial(z-d_z)}{\partial y} & \frac{\partial(z-d_z)}{\partial z} \end{bmatrix}. \tag{13}$$

The Jacobian matrix describes the velocity of the deformation in the neighboring area of each voxel. Let us denote $J(\mathbf{x})$ the local Jacobian matrix at voxel site $\mathbf{x}$. Its determinant, the Jacobian, is used to analyze the deformation needed for the spatial registration. A value $\det(J(\mathbf{x})) > 1$ implies that the neighborhood of the voxel $\mathbf{x}$ was stretched to match the template (i.e., local volumetric expansion), while $\det(J(\mathbf{x})) < 1$ is associated with local shrinkage. In TBM the Jacobian map obtained computing the Jacobian at each voxel site is used to compute group statistics at each voxel, in the template reference frame, to identify group differences in anatomical shape.

The following scalar measures are used in the computational experiments

1. The Jacobian map (*jacs*) given by $\det(J(\mathbf{x}))$ at each voxel site $\mathbf{x}$.
2. The modulated GM (*modgm*) segmentation data computed multiplying the GM segmentation probability by the Jacobian at each voxel.
3. The trace of the Jacobian matrix $\text{trace}(J(\mathbf{x}))$ at each voxel site $\mathbf{x}$ (*trace*).
4. The magnitude of the displacement field (*norms*) at each voxel site $m(\mathbf{x}) = \|[d_x(\mathbf{x}), d_y(\mathbf{x}), d_z(\mathbf{x})]\|$
5. The Geodesic Anisotropy Batchelor et al. [2] (*geodan*) defined as the geodesic distance between the Green strain tensor $\mathbf{S}_i = \left(\mathbf{J}_i^T \mathbf{J}_i\right)^{1/2}$ and the isotropic tensor within the affine-invariant metric framework Lepore et al. [23]:

$$\text{GA}(S) = \left(\text{Trace}\left(\log S - \langle \log S \rangle I\right)^2\right)^{1/2}$$

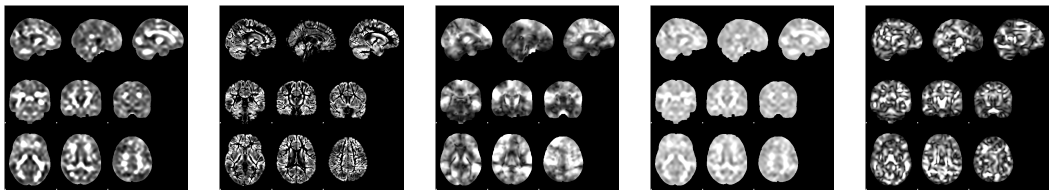
where

$$\langle \log S \rangle = \frac{\text{Trace}(\log S)}{3}.$$

In this paper, we use these maps as the voxel's raw data for the feature selection procedure that will be described later. Figure 1 shows instances of these maps.

3. Feature selection

On the deformation data we calculate the significance measure of each voxel site. This significance measure has several three forms in this paper:



(a) (b) (c) (d) (e)

Figure 1: Example deformation measure maps from the OASIS database (a) Jacobian map , (b) modulated GM, (c) displacement norm, (d) Jacobian matrix trace and (e) geodesic anisotropy.

- The absolute value of the Pearson's correlation (PC) between voxel data and the subject class labels. Big absolute values imply high voxel significance.
- The Bhattacharyya distance (BD) between classes. This is a class separability measure, so that high values mean good class separation, and thus high voxel significance.
- Welch's t-test (WT) between both groups testing if the data belong to the same distribution. Its value indicates the difference between the mean of two normally distributed populations when the variances of the two populations are not assumed to be equal, based on two independen samples.

The feature selection procedure performed for all source data and significance measures is as follows:

1. Compute the significance measure at each voxel site gathering the data from such voxel site across volumes. This results in a volume mapping of voxel site significances.
2. Compute the empirical distribution of the significance map.
3. Set a robust Huber [15] threshold on the significance map selecting a percentile of empirical distribution. In the experiments reported below we have used the following percentiles: 0.80, 0.90, 0.95, 0.99, 0.995 and 0.999.
4. Select as feature extraction sites the voxel sites with significance above the threshold.

3.1. Pearson's correlation

The Pearson's correlation coefficient Cohen [7] (PC) typically denoted by r, is a measure of the correlation between two random variables $\mathbf{x}$ and $\mathbf{y}$, taking values in the interval $r \in [-1, 1]$. A value of $r = 1$ means that the two variables are in complete agreement. A value $r = -1$ means that the two variables take opposite values. Pearson's correlation coefficient between two variables is defined as the covariance of the two variables normalized by the product of their standard deviations:

$$r\left(\mathbf{x},\mathbf{y}\right)=\frac{\sum_{i=1}^{n}(x_i-\mu_x)(y_i-\mu_y)}{\sqrt{\sum_{i=1}^{n}(x_i-\mu_x)^2}\sqrt{\sum_{i=1}^{n}(y_i-\mu_y)^2}}, \tag{14}$$

where μ_x and μ_y are the means of variables $\mathbf{x}$ and $\mathbf{y}$, respectively. In our case, considering each voxel site independently, we build a vector $\mathbf{x}_j$ composed of the voxel intensities at the j-th voxel site across all the subjects in the training set (fully independent cross-validation method) or in the whole dataset (partially independent cross-validation method). Vector $\mathbf{y}$ contains the subjects class labels, defined as:

$$y_i=\begin{cases}1 & \text{if } i-\text{th subject is patient}\\ 0 & \text{otherwise}\end{cases}$$

Then for each voxel site we calculate $\mathrm{abs}\left(r\right)$ as the significance measure for feature extraction.

3.2. Bhattacharyya distance

The Bhattacharyya distance Bhattacharyya [3] is a measure of dissimilarity between two discrete or continuous probability distributions. Assuming that the data from the same voxel site values in patients ($\mathbf{p}$) and controls ($\mathbf{q}$) follow two univariate Gaussian distributions, the Battacharyya distance has the following form Nielsen and Boltz [29]Kailath [17]

$$b(\mathbf{p},\mathbf{q})=\frac{1}{4}\frac{(\mu_p-\mu_q)}{\sigma_p^2+\sigma_q^2}+\frac{1}{2}\ln\frac{\sigma_p^2+\sigma_q^2}{2\sigma_p\sigma_q}, \tag{15}$$

where (μ_p, σ_p^2) and (μ_q, σ_q^2) are the sample mean and sample variance of the intensity values of patients and controls, respectively.

3.3. Welch's t-test

Welch's t-test Welch [37] is an adaptation of Student's t-test to compare two samples that may have unequal variances. The t-statistics is computed as follows

$$t(\mathbf{p},\mathbf{q})=\frac{(\mu_p-\mu_q)}{\sqrt{\frac{\sigma_p^2}{n_p}+\frac{\sigma_q^2}{n_q}}}, \tag{16}$$

where (μ_p, σ_p^2, n_p) and (μ_q, σ_q^2, n_q) are the sample mean, sample variance and sample size of the of intensity values of patients and controls, respectively. The conventional way of working with these will be to compute the degrees of freedom of the t-distribution, performing a two-tailed test on the statistic to reject or accept the null hypothesis. Instead, we will use the t-statistic's absolute value as the voxel site significance measure.

4. Support Vector Machine (SVM)

The Support Vector Machines (SVMs) Vapnik [36] separate a given set of binary labelled training data with a hyperplane that is maximally distant from the two classes, the maximal margin hyperplane. The objective is to build a discriminating function that will correctly classify new examples $(\mathbf{x}, y)$. The kernel trick allows to obtain non-linear discriminating functions when no linear separation of the training data is possible, so that the SVM's hyperplane corresponds to a nonlinear decision boundary in the input space mapped to a higher-dimensional space where the discriminating solution is linear. The decision function is expressed in terms of the support vectors as follows:

$$f(\mathbf{x}) = sign\left(\sum \alpha_i y_i K(\mathbf{s}_i, \mathbf{x}) + w_0\right) \tag{17}$$

where $K(.,.)$ is a kernel function, α_i is a weight of the support vector, and the $\mathbf{s}_i$ are the support vectors.

The SVM training algorithm used for this study is a generalization of the conventional SVM included in the SVM-Perf [1] software package Joachims [16]. This brand of SVM learning optimizes non-linear performance measures treating the learning problem as a multivariate prediction problem. Defining the hypotheses functions $\bar{h}$ that map a tuple $\bar{\mathbf{x}} \in \bar{\mathcal{X}}$ of n feature vectors $\bar{\mathbf{x}} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ to a tuple $\bar{y} \in \bar{\mathcal{Y}}$ of n labels $\bar{y} = (y_1, \ldots, y_n)$

$$\bar{h} : \bar{\mathcal{X}} \rightarrow \bar{\mathcal{Y}},$$

where $\bar{\mathcal{X}} = \mathcal{X} \times \ldots \times X$ and $\bar{\mathcal{Y}} \subseteq \{-1, +1\}^n$ is the set of all admissible label vectors. A linear discriminant function is used for this multivariate mapping:

$$\bar{h}_{\mathbf{w}}(\bar{x}) = \arg\max_{\bar{y}' \in \bar{\mathcal{Y}}} \left\{\mathbf{w}^T \Psi(\bar{\mathbf{x}}, \bar{y}')\right\},$$

where $\mathbf{w}$ is a parameter vector and Ψ is a function that returns a feature vector describing the match between $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $(y'_1, \ldots, y'_n)$. Ψ is restricted to be of the following simple form:

$$\Psi(\bar{\mathbf{x}}, \bar{y}') = \sum_{i=1}^{n} y'_i \mathbf{x}_i.$$

The prediction rule $\bar{h}_{\mathbf{w}}(\bar{\mathbf{x}})$ returns the tuple of labels $\bar{y}$' which scores highest according to a linear function.

Having this, the multivariate SVM optimization problem is formulated as:

$$\begin{array}{ll} & \min_{\mathbf{w}, \xi \geq 0} \frac{1}{2} \|\mathbf{w}\|^2 + C\xi \\ s.t. & \forall \bar{y}' \in \bar{\mathcal{Y}} \setminus \bar{y} : \ \mathbf{w}^T \left[\Psi(\bar{\mathbf{x}}, \bar{y}) - \Psi(\bar{\mathbf{x}}, \bar{y}')\right] \geq \Delta(\bar{y}', \bar{y}) - \xi \end{array}$$

[1] `http://svmlight.joachims.org/svm_perf.html`

where Δ is the training loss function and its upper bound ξ, also called slack variable. This formulation allows to apply the training procedure to different loss functions based on the contingency tables. Efficient algorithms to perform this training in polynomial time are given in Joachims [16].

5. Experimental design and materials

Registration. The nonlinear spatial registration of each subject's MRI volume to the MNI152 template has been performed applying FNIRT [33] from the FMRIB Software Library (FSL). A four step incremental transformation resolution process has been conducted in the experiments reported below.

Modulated gray matter. A modulated grey matter (GM) volume has been calculated for each subject. We used FSL FAST [33] to obtain a GM partial volume estimation (PVE) map and then modulated it multiplying by the Jacobian determinant voxelwise. Features extracted from the modulated grey matter are denoted *modgm* in the results section.

Non-circular analysis. To avoid double dipping analysis [22], the feature selection process is performed only on the training data at each cross-validation fold, accomplishing fully independence between test data and the system learn from the training data. Therefore, feature selection is performed as many times as cross-validation tests are performed. The brain localizations reported below as significative are those that have been found significative at least in 90% of the feature selection processes.

Performance measures. The quantitative learning performance measures used in the results reported below are: *accuracy* computed as $((TP+TN)/N)$, *precision* computed as $(TP/(TP+FP))$, sensitivity aka *recall* computed as $(TP/(TP+FN))$, where TP denote true positives, TN denote true negatives, FP denote false positives, and FN denote false negatives. Results include the area under the ROC curve which can not be computed from the classification contigency table, but can be estimated from the ranking of the SVM responses and its relation to the class labels of the samples. Suppose that we have ordered the training data samples$\{\mathbf{x}_i, y_i\}_{i=1}^{n}$ according to the linear response of the SVM: $\mathbf{w}^T\mathbf{x}_1 \geq \mathbf{w}^T\mathbf{x}_2 \geq \ldots \geq \mathbf{w}^T\mathbf{x}_n$, then a pair of samples is swaped if the order of the labels does not match with the order of the samples, the number of swaped pairs is given by

$$SP = \left|\left\{(i,j)\,\middle|(y_i > y_j) \text{ and } \left(\mathbf{w}^T\mathbf{x}_i < \mathbf{w}^T\mathbf{x}_j\right)\right\}\right|,$$

these pairs of samples would correspond to an error for some setting of the decision threshold in the computation of the ROC curve. The area under the ROC can be computed as:

$$AUC = \frac{SP}{\#pos \cdot \#neg},$$

where $\#pos \cdot \#neg$ gives the number of correct pairs (i,j) of a positive label sample $(\mathbf{x}_i, 1)$ and a negative label sample $(\mathbf{x}_j, -1)$.

		Without Dementia			*With Dementia*			
Age Group	*Total n*	*n*	*Mean age*	*M/F**	*n*	*Mean age*	*M/F**	*CDR 0.5/1/2*
<20	19	19	18.53	10/9	0		0/0	0/0/0
20s	119	119	22.82	51/68	0		0/0	0/0/0
30s	16	16	33.38	11/5	0		0/0	0/0/0
40s	31	31	45.58	10/21	0		0/0	0/0/0
50s	33	33	54.36	11/22	0		0/0	0/0/0
60s	40	25	64.88	7/18	15	66.13	6/9	12/3/0
70s	83	35	73.37	10/25	48	74.42	20/28	32/15/1
80s	62	30	84.07	8/22	32	82.88	13/19	22/9/1
≥90	13	8	91.00	1/7	5	92.00	2/3	4/1/0
Total	416	316		119/197	100		41/59	70/28/2

Table 1: Age and Diagnosis Characteristics of the Data Set. *Male/Female

5.1. Subjects

In this study we used all the subjects of the first Open Access Series of Imaging Studies (OASIS) [25]. This release of OASIS consists of a cross-sectional collection of 416 male (119 controls and 41 patients) and female (197 controls and 59 patients) subjects aged 18 to 96 years (218 aged 18 to 59 years and 198 subjects aged 60 to 96 years). One hundred subjects older than 60 have been clinically diagnosed with very mild to moderate Alzheimer's disease (70 with CDR=0.5, 28 with CDR=1 and 2 with CDR=2). Table 1 summarizes the demographics of the dataset. Further demographic and image acquisition details can be found in [25].

The database includes at least 3 raw anatomical MP-RAGE images *per* subject as well as post-processed images in diverse stages normalization. To carry out our experiments, our point of departure are the MRI volumes which have been corrected for interscan head movement, rigidly aligned to the Talairach and Tournoux space [34], spatially registered to the MNI152 template with a 12-parameter affine transformation, resampled into a 1-mm isotropic resolution, skull-stripped, and corrected for intensity inhomogeneity.

5.2. Experimental design

The computational experiments consists of the cross-validation of the linear SVM classifier over different feature vectors extracted from the OASIS data following diverse feature extraction processes. The procedure we followed in this work was: (1) nonlinearly register each subject to the MNI152 template, (2) calculate 2 measures using the displacement fields and GM volume estimation,

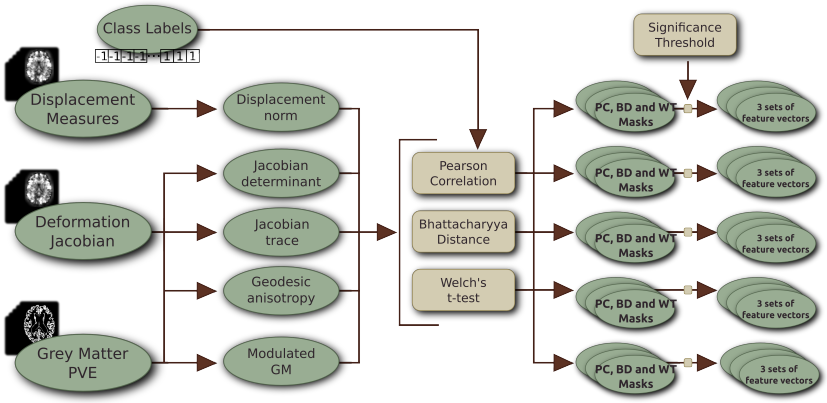


Figure 2: A summary of the pipelines of the feature selection methods. Each path from left to right specifies a feature selection process tested in the computational experiments.

	jacs	modgm
None	812498.4 (6343.65)	775086.8 (2318.37)
80%	141705.7 (3682.77)	128598.3 (1567.37)
90%	82261.3 (1971.43)	77540.0 (768.25)
95%	60427.6 (1020.87)	57981.0 (413.33)
99%	45338.3 (238.56)	44718.9 (178.46)
99.5%	43564.7 (174.22)	43207.5 (160.54)
99.9%	42136.3 (123.47)	42032.0 (130.96)

Table 2: Pearson's correlation significance measure feature set sizes mean(std) for varying percentile threshold.

(3) for each measure calculate 3 distance metrics to separate patients and controls and finally (4) for each distance metric, threshold it and create feature vectors. Figure 2 shows the a graph summarizing the diverse pipelines that lead to the actual feature vectors employed in the cross-validation experiments. From left to right, raw MRI data provide the deformation jacobian and the grey matter segmentation that may be used to compute either the Jacobian maps or the modulated GM map. From these measures, computing Pearson's correlation, Battacharyya separability measure or Welch's t-test we obtain the significance maps. Setting significance thresholds according to various setting of the percentile on the significance map empirical distribution we obtain the feature voxel sites where data is extracted to build the feature vectors. Tables 2, 3 and 4 contain the number of features extracted for each setting of the percentile threshold. As feature selection is performed in each cross-validation folder, the standard deviation of the number of features is provided. Only two of the five scalar deformation measures (modulated GM and Jacobian maps) are shown in these tables.

	jacs	modgm
None	2102154.4 (0.49)	2098251.5 (77.98)
80%	87644.8 (1775.03)	62757.5 (516.61)
90%	62642.4 (772.39)	50604.6 (204.75)
95%	51749.6 (403.02)	45889.1 (83.38)
99%	43757.5 (164.41)	42657.5 (32.21)
99.5%	42792.2 (151.90)	42281.5 (32.17)
99.9%	42039.3 (156.62)	41991.8 (25.25)

Table 3: Bhattacharyya significance measure feature set sizes mean(std) for varying percentile threshold.

	jacs	modgm
None	1289656.4 (6343.79)	1323163.8 (2313.34)
80%	84545.1 (1108.19)	71914.6 (652.86)
90%	59979.4 (579.14)	54066.6 (247.73)
95%	50228.2 (220.14)	47383.8 (138.54)
99%	43432.1 (78.49)	42904.0 (82.05)
99.5%	42637.7 (71.82)	42397.6 (88.42)
99.9%	42004.6 (80.34)	41987.7 (82.18)

Table 4: T-test significance measure feature set sizes mean(std) for varying percentile threshold.

5.3. Grid search for SVM tuning

Parameter setting for SVM is critical to obtain good performance. Thus, each learning experiment is wrapped with a grid search procedure [24]. Thus, every time an algorithm is trained on a data set T, an internal stratified two-fold cross-validation (repeated 3 times) is performed for each combination of parameter values, looking for the best parameter setting for T. The grid search mechanism selects the combination of parameters maximizing the F1-score for the SVM minimizing the error rate. The values examined for the regularization parameter are C $C \in \left\{10^{-2}, 3{\cdot}10^{-2}, 10^{-1}, 3{\cdot}10^{-1}, 1\right\}$, the values for the RBF kernel parameter σ ranged from 10^{-3} to 1, varying the exponents in steps of 1.

6. Results

6.1. Classification results

We have tested all combinations of feature selection processes illustrated in figure 2 with all kind of deformation scalar measures. The results are shown in figures 3, 4 and 5 for the Pearson's correlation, Battacharyya separability and Welch's t-test, respectively. Each figure shows the plots of the average accuracy, precision, recall and AUC for varying number of features, depending on the percentile threshold, for each deformation scalar measure. The plots contain also the visualization of the uncertainty as bars giving the span of one standard deviation. The qualitative inspection of these plots shows that in each

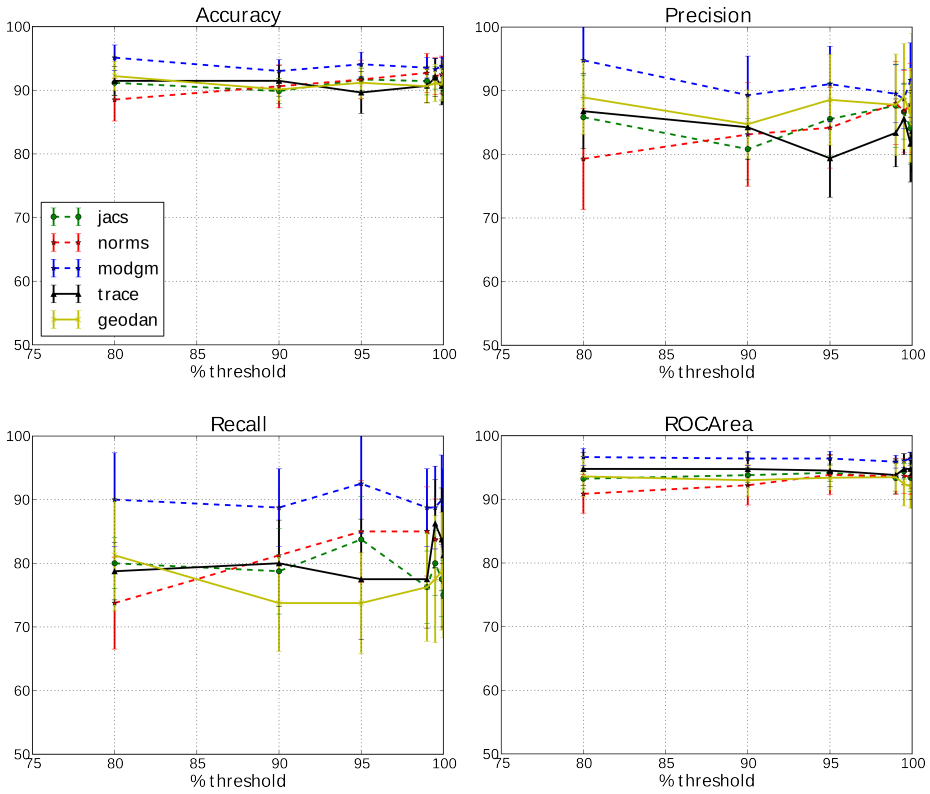


Figure 3: Average and standard deviation of Accuracy, Precision, Recall and ROC area over the 10-fold cross-tests results of a linear SVM trained to minimize the error rate, features selected using **Pearson correlation**. Note that the cross-validation population subsets were all the same for all the experiments.

of them there seems to be no significant differences as all the average values fall inside the rough confidence interval given by the standard deviation bars of the other approaches. Performing a two tail t-test on the data of each plot separately reaches the same conclusion, there are not significative differences between approaches ($p<0.05$). Overall, the acuracy and AUC results have little variability within each plot, while the precision and recall have more fluctuations feature sets. For the Pearson's correlation, the best results are obtained with the Modulated GM data, while for the other significance measures, best results are given by the trace and geodesic anisotropy data. For us it is unexpected that the Jacobian maps produce classification results below other approaches, because they have been used as the reference value for the TBM studies reported in the literature.

Table 5 shows the average and standar deviation of cross-validation results obtained for each voxel site significance measure applying the 95% percentile for feature selection. Notice that Modulated GM provides the best results for

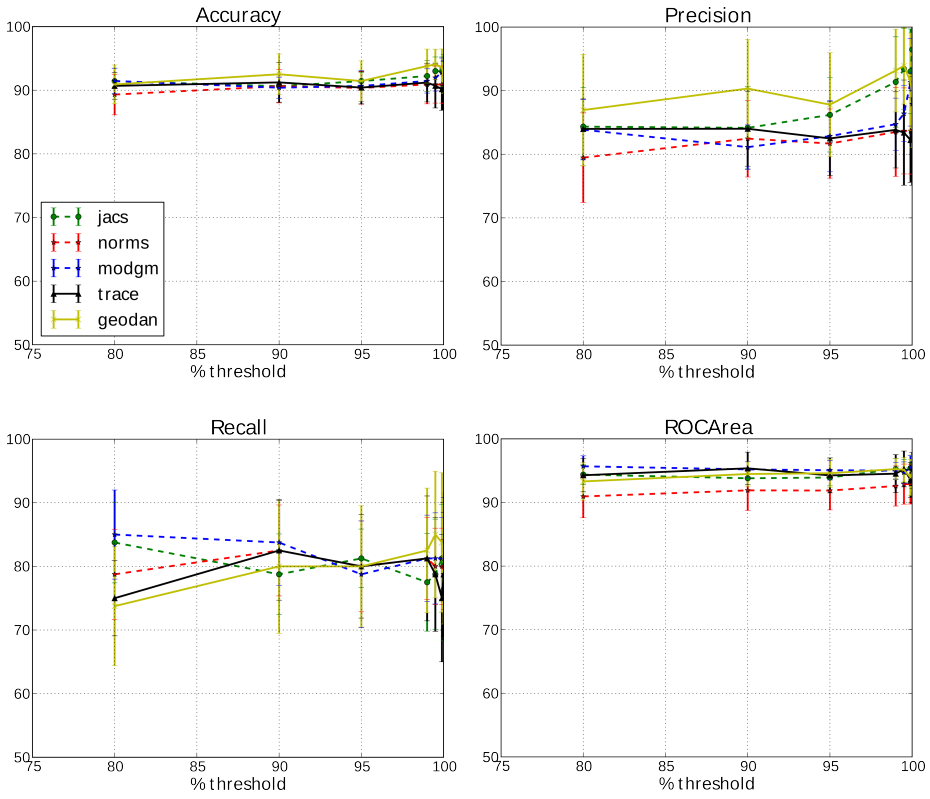


Figure 4: Average and standard deviation of Accuracy, Precision, Recall and ROC area over the 10-fold cross-tests results of linear SVM trained to minimize error rate, features selected using **Bhattacharyya distance**. Note that the cross-validation population subsets were all the same for all the experiments.

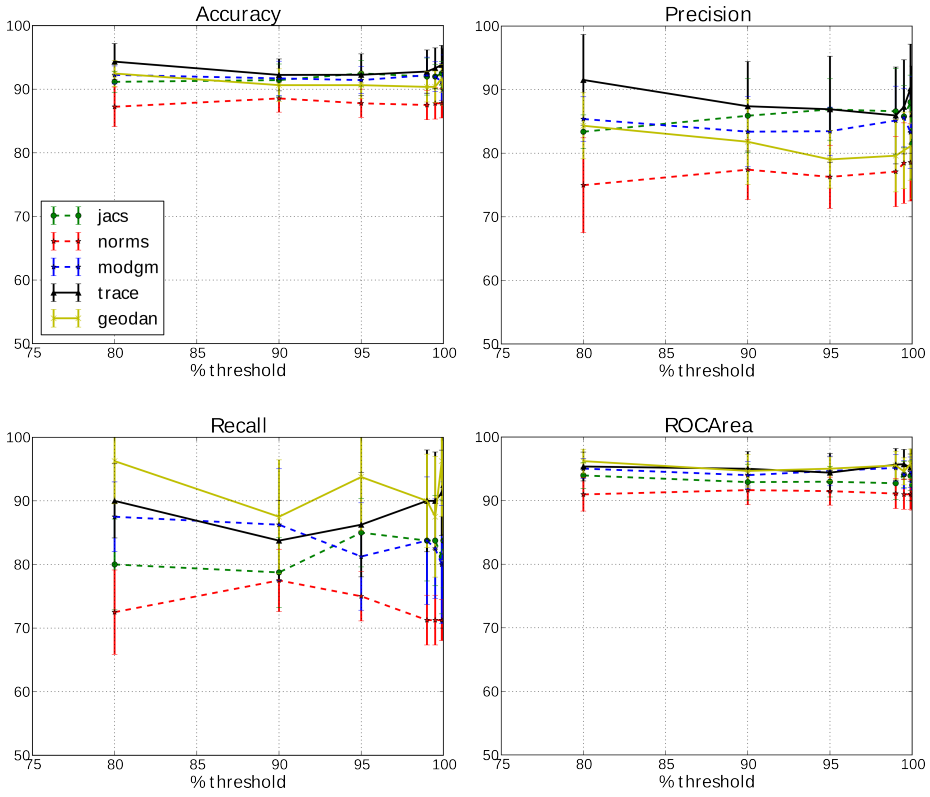


Figure 5: Mean and standard deviation of Accuracy, Precision, Recall and ROC area over the 10-fold cross-tests results of linear SVM trained minimizing error rate, features selected using **Welch's t-test**. Note that the cross-validation population subsets were all the same for all the experiments.

		Accuracy	Precision	Recall	AUC
	jacs	88.10 (0.00)	50.76 (1.70)	58.33 (3.73)	92.32 (1.53)
	norms	88.10 (2.75)	84.14 (6.12)	58.33 (8.98)	94.79 (2.53)
PC	modgm	**92.07 (1.12)**	**95.83 (5.89)**	**86.67 (4.71)**	**96.67 (0.44)**
	trace	89.43 (2.70)	79.55 (5.77)	65.00 (5.00)	94.27 (0.43)
	geodan	88.15 (2.25)	87.74 (3.96)	70.00 (0.00)	93.49 (1.04)
	jacs	91.27 (1.22)	86.44 (1.89)	**81.67 (7.45)**	95.36 (1.54)
	norms	89.89 (1.77)	82.62 (6.94)	79.67 (3.73)	93.12 (1.48)
BD	modgm	89.74 (1.77)	82.94 (4.14)	78.00 (5.00)	**97.37 (0.29)**
	trace	89.52 (1.78)	82.43 (4.15)	80.33 (7.45)	95.67 (1.08)
	geodan	**92.09 (2.60)**	**88.09 (5.33)**	80.00 (4.00)	95.37 (0.99)
	jacs	92.15 (3.07)	87.36 (6.67)	85.373 (6.87)	93.67 (0.39)
	norms	88.50 (0.89)	76.40 (3.62)	75.00 (4.00)	92.95 (0.87)
WT	modgm	91.43 (0.89)	**85.98 (1.55)**	83.33 (3.73)	94.54 (0.50)
	trace	**92.83 (0.91)**	85.62 (0.85)	86.33 (3.73)	94.32 (0.55)
	geodan	89.92 (1.78)	79.42 (4.85)	**94.67 (7.45)**	**95.00 (0.52)**

Table 5: Mean (standard deviation) of Accuracy, Precision, Sensitivity, and ROC area of the full 10-fold cross-validation classification results for features selected with the 95% percentile. Note that the cross-validation population subsets were all the same for all the experiments.

the Pearson's correlation, and that results are very close in some instances, in fact there are not statistically significant differences. The results compare well with other approaches to AD prediction from MRI data [32, 5, 30] which report accuracies below those reported here for smaller datasets.

6.2. Location of discriminant voxel sites

To have a qualitative validation of the process, we have localized the voxel sites selected by the feature selection process on two probabilistic atlases (MNI structural atlas and the Harvard-Oxford cortical and subcortical atlas). Voxel sites localized are those appearing in 90% of the feature selection processes performed independently in each cross-validation fold. The regions from the different experiments are all within similar brain areas: frontal and parietal lobes, cerebellum, temporal and occipital lobes, frontal pole, lateral occipital cortex, superior division, precentral gyrus, postcentral gyrus and hippocampus. All these findings agree with the imaging biomarkers reported in the literature [10, 6, 8, 27].

In figures 6 and 7, we show some slices of the MNI152 template with an overlay with the 95% percentile voxel site localizations across the feature selection processes that were performed in the cross-validation folds over the modulated GM data and the Jacobian map data, respectively. Figures show the findings using the three different significance measures.

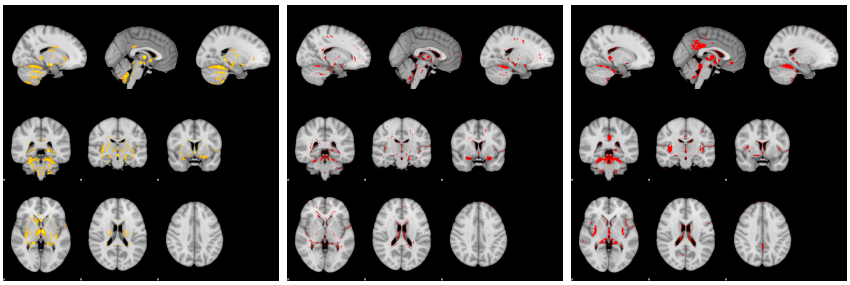


Figure 6: Discriminant voxel sites selected applying a 95% percentile on the empirical distribution of the Pearson correlation (left), Bhattacharyya distance (middle) and Welch t-test (right) computed on the modulated GM (*modgm*).

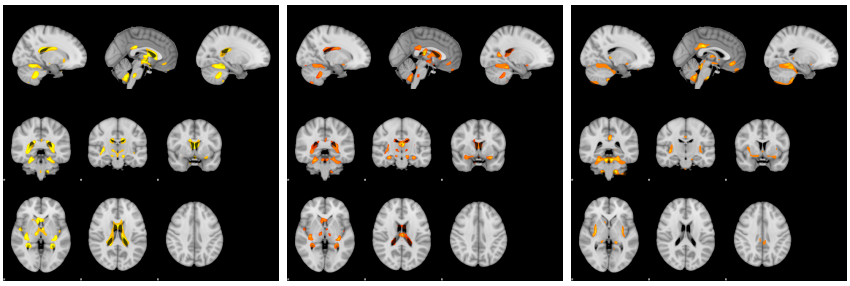


Figure 7: Discriminant voxel sites selected applying a 95% percentile on the empirical distribution of the Pearson correlation (left), Bhattacharyya distance (middle) and Welch t-test (right) computed on the Jacobian maps (*jacs*).

7. Conclusions

In this paper we report new results on Computer Aided Diagnosis (CAD) based on features selected from deformation fields, and modulated GM, obtained from non-linear registration processes. Specifically, five scalar measures computed using the Jacobian matrix of the displacement fields have been used as raw data. For feature selection, three different voxel site significance measures have been computed. Classification has been performed using conventional linear SVM algorithm. The average localization of discriminant sites is consistent with the literature regarding imaging biomarkers of AD. Classification results are competitive with other reported approaches to the CAD of AD. The difference of classification results between deformation measures is not statistically significative if we take into account the standard deviation from the cross-validation experiments to perform a two-sided t-test, neither was the voxel site significance measure. The result extends to the diverse performance measures computed, thought their respective ranges of values change between them. These results are unexpected, as the usual assumption is that the Jacobian maps are characterizing the anatomical deformations to detect atrophy and other effects of AD and other neurodegenerative diseases.

Future works are driven towards the application of other classification algorithms to the feature databases extracted for this work. Extension of the works to bigger datasets is also desirable.

Acknowlegdments

We thank the Washington University ADRC for making the OASIS MRI data available. Research partially funded by the Ministerio de Ciencia e Innovación of the Spanish Government, and the Basque Government funds for the research group and predoctoral grant.

[1] Alzheimer's Association. 2011 alzheimer's disease facts and figures. *Alzheimer's and Dementia*, 7(2):208–244, 2011.

[2] P G Batchelor, M Moakher, D Atkinson, F Calamante, and A Connelly. A rigorous framework for diffusion tensor calculus. *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 53(1): 221–225, January 2005. ISSN 0740-3194. doi: 10.1002/mrm.20334. URL `http://www.ncbi.nlm.nih.gov/pubmed/15690523`. PMID: 15690523.

[3] A. Bhattacharyya. On a measure of divergence between two multinomial populations. *Sankhya: The Indian Journal of Statistics (1933-1960)*, 7 (4):401–406, July 1946. ISSN 0036-4452. URL `http://www.jstor.org/stable/25047882`. Copyright 1946 Indian Statistical Institute.

[4] R. Chaves, J. Ramírez, J.M. Górriz, and C.G. Puntonet. Association rule-based feature selection method for alzheimer?s disease diagnosis. *Expert*

Systems with Applications, 39(14):11766–11774, October 2012. ISSN 0957-4174. doi: 10.1016/j.eswa.2012.04.075. URL http://www.sciencedirect.com/science/article/pii/S0957417412006720.

[5] D. Chyzhyk, M. Graña, A. Savio, and J. Maiora. Hybrid dendritic computing with kernel-LICA applied to alzheimer's disease detection in MRI. *Neurocomputing*, 75(1):72–77, January 2012. ISSN 0925-2312. doi: 10.1016/j.neucom.2011.02.024. URL http://dx.doi.org/10.1016/j.neucom.2011.02.024.

[6] G. Chételat, B. Desgranges, V. De La Sayette, F. Viader, F. Eustache, and J.-C. Baron. Mapping gray matter loss with voxel-based morphometry in mild cognitive impairment. *Neuroreport*, 13(15):1939–1943, October 2002. ISSN 0959-4965. URL http://www.ncbi.nlm.nih.gov/pubmed/12395096. PMID: 12395096.

[7] J. Cohen. *Statistical power analysis for the behavioral sciences*. Routledge, 1988. ISBN 9780805802832.

[8] A. Convit, J. de Asis, M. J. de Leon, C. Y. Tarshish, S. De Santi, and H. Rusinek. Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to alzheimer's disease. *Neurobiology of Aging*, 21(1):19–26, February 2000. ISSN 0197-4580. URL http://www.ncbi.nlm.nih.gov/pubmed/10794844. PMID: 10794844.

[9] C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S. M. Resnick. Detection of prodromal alzheimer's disease via pattern classification of MRI. *Neurobiology of aging*, 29(4):514–523, April 2008. ISSN 0197-4580. doi: 10.1016/j.neurobiolaging.2006.11.010. PMID: 17174012 PMCID: 2323584.

[10] B. C. Dickerson, I. Goncharova, M. P. Sullivan, C. Forchetti, R. S. Wilson, D. A. Bennett, L. A. Beckett, and L. deToledo-Morrell. MRI-derived entorhinal and hippocampal atrophy in incipient and very mild alzheimer's disease. *Neurobiology of Aging*, 22(5):747–754, October 2001. ISSN 0197-4580. URL http://www.ncbi.nlm.nih.gov/pubmed/11705634. PMID: 11705634.

[11] A. Gallix, J.M. Górriz, J. Ramírez, I.A. Illán, and E.W. Lang. On the empirical mode decomposition applied to the analysis of brain SPECT images. *Expert Systems with Applications*, (0). ISSN 0957-4174. doi: 10.1016/j.eswa.2012.05.058. URL http://www.sciencedirect.com/science/article/pii/S0957417412007749.

[12] E. Gerardin, G. Chételat, M. Chupin, R. Cuingnet, B. Desgranges, H.-S. Kim, M. Niethammer, B. Dubois, S. Lehéricy, L. Garnero, F. Eustache, and O. Colliot. Multidimensional classification of hippocampal shape features discriminates alzheimer's disease and mild cognitive impairment from normal aging. *NeuroImage*, 47(4):1476–1486, October 2009.

ISSN 1053-8119. doi: 10.1016/j.neuroimage.2009.05.036. URL http://www.sciencedirect.com/science/article/pii/S1053811909005485.

[13] M. Graña, M. Termenon, A. Savio, A. Gonzalez-Pinto, J. Echeveste, J.M. Pérez, and A. Besga. Computer aided diagnosis system for alzheimer disease using brain diffusion tensor imaging features selected by pearson's correlation. *Neuroscience Letters*, 502(3):225–229, September 2011. ISSN 0304-3940. doi: 10.1016/j.neulet.2011.07.049. URL http://www.sciencedirect.com/science/article/pii/S0304394011011268.

[14] D. Horínek, A. Varjassyová, and J. Hort. Magnetic resonance analysis of amygdalar volume in alzheimer's disease. *Current Opinion in Psychiatry*, 20(3):273–277, 2007.

[15] Peter J. Huber. *Robust Statistics*. Wiley Series in Probability and Statistics. January 2005. ISBN 9780471418054, 9780471725251. URL http://onlinelibrary.wiley.com/book/10.1002/0471725250.

[16] T. Joachims. A support vector method for multivariate performance measures. *Proceedings of the 22nd International Conference on Machine Learning*, pages 377—384, 2005. URL http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.140.1854.

[17] T. Kailath. The divergence and bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15(1):52–60, February 1967. ISSN 0018-9332. doi: 10.1109/TCOM.1967.1089532.

[18] J. A. Kaye, T. Swihart, D. Howieson, A. Dame, M. M. Moore, T. Karnos, R. Camicioli, M. Ball, B. Oken, and G. Sexton. Volume loss of the hippocampus and temporal lobe in healthy elderly persons destined to develop dementia. *Neurology*, 48(5):1297–1304, May 1997. ISSN 0028-3878. URL http://www.ncbi.nlm.nih.gov/pubmed/9153461. PMID: 9153461.

[19] R. J. Killiany, T. Gomez-Isla, M. Moss, R. Kikinis, T. Sandor, F. Jolesz, R. Tanzi, K. Jones, B. T. Hyman, and M. S. Albert. Use of structural magnetic resonance imaging to predict who will get alzheimer's disease. *Annals of Neurology*, 47(4):430–439, April 2000. ISSN 0364-5134. URL http://www.ncbi.nlm.nih.gov/pubmed/10762153. PMID: 10762153.

[20] Arno Klein, Jesper Andersson, Babak A. Ardekani, John Ashburner, Brian Avants, Ming-Chang Chiang, Gary E. Christensen, D. Louis Collins, James Gee, Pierre Hellier, Joo Hyun Song, Mark Jenkinson, Claude Lepage, Daniel Rueckert, Paul Thompson, Tom Vercauteren, Roger P. Woods, J. John Mann, and Ramin V. Parsey. Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *NeuroImage*, 46(3):786 – 802, 2009. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2008.12.037. URL http://www.sciencedirect.com/science/article/pii/S1053811908012974.

[21] S. Kloppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack, J. Ashburner, and R. S. J. Frackowiak. Automatic classification of MR scans in alzheimer's disease. *Brain*, 131(3): 681–689, March 2008. doi: 10.1093/brain/awm319. URL `http://brain.oxfordjournals.org/cgi/content/abstract/131/3/681`.

[22] N. Kriegeskorte, W. K. Simmons, P. S. F. Bellgowan, and C. I. Baker. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci*, 12(5):535–540, May 2009. ISSN 1097-6256. doi: 10.1038/nn.2303. URL `http://dx.doi.org/10.1038/nn.2303`.

[23] N. Lepore, C. Brun, Y. Y. Chou, M. C. Chiang, R. A. Dutton, K. M. Hayashi, E. Luders, O. L. Lopez, H. J. Aizenstein, A. W. Toga, J. T. Becker, and P. M. Thompson. Generalized tensor-based morphometry of HIV/AIDS using multivariate statistics on deformation tensors. *IEEE Transactions on Medical Imaging*, 27(1):129–141, January 2008. ISSN 0278-0062. doi: 10.1109/TMI.2007.906091. URL `http://www.ncbi.nlm.nih.gov/pubmed/18270068`. PMID: 18270068.

[24] O. Luaces, F. Taboada, G. M. Albaiceta, L. A. Domínguez, P. Enríquez, and A. Bahamonde. Predicting the probability of survival in intensive care unit patients from a small number of variables and training examples. *Artificial Intelligence in Medicine*, 45(1):63–76, January 2009. ISSN 1873-2860. doi: 10.1016/j.artmed.2008.11.005. URL `http://www.ncbi.nlm.nih.gov/pubmed/19185475`. PMID: 19185475.

[25] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner. Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience*, 19(9):1498–1507, September 2007. ISSN 0898-929X. doi: 10.1162/jocn.2007.19.9.1498. URL `http://www.ncbi.nlm.nih.gov/pubmed/17714011`. PMID: 17714011.

[26] F.J. Martínez-Murcia, J.M. Górriz, J. Ramírez, C.G. Puntonet, and D. Salas-González. Computer aided diagnosis tool for alzheimer?s disease based on Mann?Whitney?Wilcoxon U-Test. *Expert Systems with Applications*, 39(10):9676–9685, August 2012. ISSN 0957-4174. doi: 10.1016/j.eswa.2012.02.153. URL `http://www.sciencedirect.com/science/article/pii/S0957417412004162`.

[27] D. Medina, L. DeToledo-Morrell, F. Urresta, J. D. E. Gabrieli, M. Moseley, D. Fleischman, D. A. Bennett, S. Leurgans, D. A. Turner, and G. T. Stebbins. White matter changes in mild cognitive impairment and AD: a diffusion tensor imaging study. *Neurobiology of Aging*, 27(5):663–672, May 2006. ISSN 1558-1497. doi: 10.1016/j.neurobiolaging.2005.03.026. URL `http://www.ncbi.nlm.nih.gov/pubmed/16005548`. PMID: 16005548.

[28] P. T. Nelson, H. Braak, and W. R. Markesbery. Neuropathology and cognitive impairment in alzheimer disease: a complex but coherent relationship.

Journal of Neuropathology and Experimental Neurology, 68(1):1–14, January 2009. ISSN 0022-3069. doi: 10.1097/NEN.0b013e3181919a48. URL http://www.ncbi.nlm.nih.gov/pubmed/19104448. PMID: 19104448.

[29] F. Nielsen and S. Boltz. The Burbea-Rao and bhattacharyya centroids. *arXiv:1004.5049*, April 2010. URL http://arxiv.org/abs/1004.5049.

[30] C. Plant, S. J. Teipel, A. Oswald, C. Böhm, T. Meindl, J. Mourao-Miranda, A. W. Bokde, H. Hampel, and M. Ewers. Automated detection of brain atrophy patterns based on MRI for the prediction of alzheimer's disease. *NeuroImage*, 50(1):162–174, March 2010. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2009.11.046. URL http://www.sciencedirect.com/science/article/pii/S1053811909012312.

[31] A. Savio. Supervised classification using deformation-based features for alzheimer's disease detection on the oasis cross-sectional database. In Manuel Graña et al., editor, *Advances in Knowledge-Based and Intelligent information and Engineering Systems 2012*. IOS Press, 2012.

[32] A. Savio, M.T. García-Sebastián, D. Chyzyk, C. Hernandez, M. Graña, A. Sistiaga, A. López de Munain, and J. Villanúa. Neurocognitive disorder detection based on feature vectors extracted from VBM analysis of structural MRI. *Computers in Biology and Medicine*, 41(8):600–610, August 2011. ISSN 0010-4825. doi: 10.1016/j.compbiomed.2011.05.010. URL http://www.sciencedirect.com/science/article/pii/S0010482511001065.

[33] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. J. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J. M. Brady, and P. M. Matthews. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23 Suppl 1:S208–219, 2004. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2004.07.051. URL http://www.ncbi.nlm.nih.gov/pubmed/15501092. PMID: 15501092.

[34] J. Talairach and P. Tournoux. *Co-Planar Stereotaxic Atlas of the Human Brain: 3-D Proportional System: An Approach to Cerebral Imaging*. Thieme, January 1988. ISBN 0865772932.

[35] S. J. Teipel, C. Born, M. Ewers, A.L. W. Bokde, M. F. Reiser, H.-J. Möller, and H. Hampel. Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. *NeuroImage*, 38(1):13–24, October 2007. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2007.07.008. URL http://www.sciencedirect.com/science/article/B6WNP-4P718MM-4/2/a63c5d0de3f88f1da150b390cd97e0dc.

[36] V. N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, September 1998. ISBN 0471030031.

[37] B. L. Welch. The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, 34(1/2):28–35, January 1947. ISSN 0006-3444. doi: 10.2307/2332510. URL `http://www.jstor.org/stable/2332510`. Copyright 1947 Biometrika Trust.