Supervised classification using deformation-based features for Alzheimer's disease detection on the OASIS cross-sectional database

Alexandre Savio*, Manuel Graña

Grupo de Inteligencia Computacional (GIC), Universidad del País Vasco (UPV/EHU), San Sebastián, Spain

Abstract. In the last 10 years, detection of Alzheimer's disease based on brain T1-weighted Magnetic Resonance Imaging (MRI) have been a highly sought goal in the neuroscientific community. However, the methods were assessed on different datasets and not always publicly available ones making reproducibility and validation impossible. Here, we evaluated five deformation-based features (Jacobian determinant and trace, modulated grey matter (GM), deformation vector norm and geodesic anisotropy) and three feature selection processes (Pearson correlation, Bhattacharyya distance and Welch's t-test) using the 416 subjects (316 controls and 100 patients) from the OASIS database. Our objective was to correctly discriminate between controls (CN) and Alzheimer's disease patients (AD). For this task we used a fully and non-fully independent stratified 10-fold cross-validation and linear SVM. Our best mean detection result was 79.43% of precision, 96.67% of sensitivity and 97.37 of area under the ROC curve. We also show discriminant voxel site locations found with each measure.

1 Introduction

Alzheimer's Disease (AD) is a neurodegenerative disorder which is one of the most common cause of dementia in old people. Due to the socioeconomic importance of the disease in occidental countries there is a strong international effort focus 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 [23]. Therefore, the development of automated detection procedures based in MRI and other medical imaging techniques [7] is of high interest. In the early stages of AD brain atrophy may be subtle and spatially distributed over many brain regions [16,8,5,2], including the entorhinal cortex, the hippocampus, lateral and inferior temporal structures, anterior and posterior cingulate.

* This work has been partially supported by the "Ayudas para la Formación de Personal Investigador" fellowship from the Gobierno del País Vasco.

A similar study [27] obtained 83% of accuracy using similar approaches to detect subjects with mild cognitive impairment. Although their results can not be reproduced, this work confirms that the approach that we follow is a promising area of research. Recently [6], a study with 509 subjects drawn from the public available ADNI database achieved high accuracies up to 81% of sensitivity and 95% of specificity using different feature selection approaches based on tissue segmentations, cortical and hippocampal measures and a hold out method for validation. In [30] they combined regional segmentation and a manual hippocampus delineation features from AddNeuroMed were used for detection of AD patients with a sensitivity of 90% and a specificity of 94% using orthogonal partial least squares (OLPS).

Machine learning methods have become very popular to classify functional or structural brain images to discriminate them into two classes: normal or a specific neurodegenerative disorder [3]. There are different ways to extract features from MRI for SVM classification: based on morphometric methods [24,7,13], based on ROIs (region of interest) [21] or GM voxels in automated segmentation images [17]. Work has also been reported on the selection of a small set of the most informative features for classification, such as the SVM-Recursive Feature Elimination [9], the selection based on statistical tests [21], the wavelet decomposition of the RAVENS maps [19], among others. It is important to note that these techniques are aimed to help clinicians with more statistical evidence for the diagnosis, it is not intended to substitute any other existing diagnosis procedure.

In this paper we use deformation-based features to guide the feature extraction process analyzing displacement vectors resulting from non-linear registration procedures with high number of degrees of freedom. Each subject in the dataset is registered to a standard template and scalar measurements are calculated from the resulting displacement fields. These features are then used in a filter selection process to localize the most discriminant voxel sites of the dataset. Using these findings, sets for each measure and distance are created for further cross-validation classification tests. We report the results of Support Vector Machine (SVM) with linear kernel performing the classification task.

Section Materials and Methods gives a description of the subjects selected for the study, the image processing, feature extraction details and the classifier system. Section Results gives our classification performance results and section Conclusions gives the conclusions of this work and further research suggestions.

2 Materials and Methods

The procedure we followed in this work was: (1) nonlinearly register each subject to the MNI template, (2) calculate 5 measures using the displacement fields and GM volume estimation, (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.

In this study we used all the subjects of the first Open Access Series of Imaging Studies (OASIS) [22]. These subjects were selected from a larger database of individuals who had participated in MRI studies at Washington University, they were all right-handed and older adults had a recent clinical evaluation. Older subjects with and without dementia were obtained from the longitudinal pool of the Washington University Alzheimer Disease Research Center (ADRC). 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). Further demographic and image acquisition details can be found in [22]. The database includes at least 3 raw anatomical MP-RAGE images from each subject as well as post-processed images: (a) corrected for interscan head movement and rigidly aligned to the Talairach and Tournoux space [26], (b) transformed to a template with a 12-parameter affine registration and merged into a 1-mm isotropic image, (c) skull-stripped and corrected for intensity inhomogeneity and (d) segmented by tissue type. To carry out our experiment we used the volumes from (c).

The spatial normalization of each subject of the database has been performed with FSL FNIRT [25]. A four step registration process with increasing resolution and a scaled conjugate gradient minimization method has been performed using the default parameters and the MNI152 brain template. A visual check has been performed for all images in every processing step carried out in this experiment.

2.1 Voxel measures

Five measures were tested in this study: deformation Jacobian determinant and trace, displacement norm, geodesic anisotropy of the deformation tensor and modulated GM.

In Tensor-based Morphometry (TBM) a template T is nonlinearly registered to a subject S, and a displacement vector $\overrightarrow{u}(\overrightarrow{r})$ is obtained such that $T(\overrightarrow{r}-\overrightarrow{u})$ corresponds with $S(\overrightarrow{r})$, where $\overrightarrow{r}$ denotes the voxel location. The Jacobian matrix in this case describes the velocity of the deformation procedure in the neighboring area of each voxel and it is defined by

$$\mathbf{J}_i = \begin{pmatrix} \partial(x-u_x)/\partial x & \partial(x-u_x)/\partial y & \partial(x-u_x)/\partial z \\ \partial(y-u_y)/\partial x & \partial(y-u_y)/\partial y & \partial(y-u_y)/\partial z \\ \partial(z-u_z)/\partial x & \partial(z-u_z)/\partial y & \partial(z-u_z)/\partial z \end{pmatrix}. \quad (1)$$

For all the subjects these matrices are compared in a General Linear Model procedure to find group difference effects. Here instead we use the Jacobian matrices to find relevant features for disease detection in a supervised classification algorithm.

Using the deformation Jacobian matrices we define five measures which will be used as features in our feature extraction procedures.

The determinant of the Jacobian matrix is the most commonly used scalar measure of deformation for TBM analyses [20]. The determinant of the Jacobian matrix $\mathbf{J}_i$ is commonly used to analyze the distortion necessary to deform the images into agreement. A value $\det(\mathbf{J}_i) > 1$ implies that the neighborhood

adjacent to the displacement vector in voxel i was stretched to match the template (i.e., local volumetric expansion), while $\det(\mathbf{J}_i) < 1$ is associated with local shrinkage. These features are denoted as *jacs* in the following section.

We also used the trace of the Jacobian matrix of the displacement field, denoted as *trace* in the next sections.

The displacement vector magnitudes defined as $m_i = \sqrt{x_i^2 + y_i^2 + z_i^2}$ is denoted as *norms* in the sections bellow.

A modulated GM volume has been calculated for each subject, we used FSL FAST [25] to obtain an estimation of the GM partial volume map and then modulated it by multiplying to the Jacobian determinant. We denoted these features as *modgm* further on.

Geodesic anisotropy (GA) measures the geodesic distance between the Green strain tensors defined as $\mathbf{S}_i = \left(\mathbf{J}_i^T \mathbf{J}_i\right)^{1/2}$ [20] on the tangent plane at the origin of the symmetric positive-definite tensor manifold [11]. We denoted these features as *geodan* in the following sections.

2.2 Feature Selection

We localized discriminant voxel sites using three methods: Pearson's correlation (PC), Bhattacharyya distance (BD) and Welch's t-test (WT).

Considering each voxel site independently, we build two vectors $\mathbf{p}_j$ and $\mathbf{q}_j$ (only one in case of PC) composed of the voxel intensities at the j-th voxel site across control subjects and patients (all the subjects in PC) in the training set (fully independent cross-validation) or in the whole dataset (non-fully independent cross-validation). Once the distance measure is obtained for each voxel site we calculate a mask that indicates voxel sites above a threshold of robust range of 95%. This mask is used to extract independent sets of feature vectors from the subjects for classification.

Pearson's product-moment correlation coefficient (PC) PC [4] (typically denoted by r) is a measure of the correlation (linear dependence) between two variables $\mathbf{x}$ and $\mathbf{y}$, resulting in a value between $+1$ and -1 inclusive. It is widely used in statistics as a measure of the strength of linear dependence between two variables. It was developed by Karl Pearson from a similar but slightly different idea introduced by Francis Galton [12]. The correlation coefficient is sometimes called "Pearson's r."

Then for each voxel site we calculated $\mathrm{abs}(r)$ as a discriminant metric for feature selection.

Bhattacharyya distance (BD) BD [1] is a measure of similarity between two discrete or continuous probability distributions. Assuming that we have two univariate Gaussian distributions (controls and patients) for each voxel site, the absolute Bhattacharyya distance [15] was calculated between both groups.

Wech's t-test (WT) WT [29] is an adaptation of Student's t-test for the case when comparing two samples that possibly have unequal variances. We computed the absolute value of the WT between both groups for feature selection.

A 10-fold cross-validation was carried out to calculate the results, we stratified training and test set in order to have proportional number of controls and patients in each random disjoint set.

2.3 Support Vector Machines

The Support Vector Machines (SVMs) have attracted attention from the pattern recognition community [6] owing to a number of theoretical and computational merits derived from [28]. SVM separates a given set of binary labeled training data with a hyperplane that is maximally distant from the two classes (known as the maximal margin hyperplane). The objective is to build a discriminating function using training data that will correctly classify new examples $(\mathbf{x}, y)$. When no linear separation of the training data is possible, SVMs can work effectively in combination with kernel techniques using the kernel trick, so that the hyperplane defining the SVMs corresponds to a nonlinear decision boundary in the input space that is mapped to a linearized higher- dimensional space [28]. In this way the decision function can be expressed in terms of the support vectors only:

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

where $K(.,.)$ is a kernel function, α_i is a weight constant derived from the SVM process and the $\mathbf{s}_i$ are the support vectors [28].

The Support Vector Machine (SVM) [28] algorithm used for this study is included in the SVM-Perf (`http://svmlight.joachims.org/svm_perf.html`) software package. The implementation is described in detail in [14]. Given training vectors $\mathbf{x}_i \in \mathbb{R}^n, i = 1, \ldots, l$ of the subject features of the two classes, and a vector $\mathbf{y} \in \mathbb{R}^l$ such that $y_i \in \{-1, 1\}$ labels each subject with its class, in our case, for example, patients were labeled as -1 and control subject as 1. To construct a classifier, the SVM algorithm tries to maximize the classification margin. To this end it solves the following optimization problem:

$$\min_{w,b,\xi} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{l} \boldsymbol{\xi}_i$$

subject to $y_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b) \geq (1 - \xi_i),\ \xi_i \geq 0, i = 1, 2, \ldots, n$. The dual optimization problem is

$$\min_{\alpha} \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{Q}\boldsymbol{\alpha} - \mathbf{e}^T\boldsymbol{\alpha},$$

subject to $\mathbf{y}^T\boldsymbol{\alpha} = 0,\ 0 \leq \alpha_i \leq C,\ i = 1, \ldots, l$, where $\mathbf{e}$ is the vector of all ones, $C > 0$ is the upper bound on the error, $\mathbf{Q}$ is an $l \times l$ positive semi-definite matrix,

$Q_{ij} \equiv y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$, and $K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ is the kernel function that describes the behavior of the support vectors. Here, the training vectors $\mathbf{x}_i$ are mapped into a higher (maybe infinite) dimensional space by the function $\phi(\mathbf{x}_i)$. C is a regularization parameter used to balance the model complexity and the training error.

The kernel function chosen results in different kinds of SVM with different performance levels, and the choice of the appropriate kernel for a specific application is a difficult task. In this study we only tested the linear kernel for simplification. The linear kernel function is defined as $K(\mathbf{x}_i, \mathbf{x}_j) = 1 + \mathbf{x}_i^T \mathbf{x}_j$.

3 Results

In this section we present the linear kernel SVMs performance using the full 10-fold cross-validation. In table 1 we report accuracy $((TP + TN)/N)$, precision $(TP/(TP + FP))$, sensitivity $(TP/(TP + FN))$, F1-score $(Precision \times Sensitivity)$ and area under the ROC curve (AUC) [10,14] for each combination of deformation measure and feature selection method. As the accuracy is not a good measure for non balanced datasets, i.e., has no equal number of patients and controls, we will take into account the rest of performance measures.

The results obtained with *jacs* and *modgm* are significantly worse than the other measures in all distances. The best performance is the *trace* using the Bhatacharyya distance with 97.37 of AUC, although the experiments with the same measure and the other distances give very good results as well. The next measures in performance are *norms* and finally *geodan*.

		Accuracy	Precision	Sensitivity	F1-Score	AUC
	jacs	76.02 (0.91)	50.76 (1.70)	58.33 (3.73)	54.22 (2.11)	80.32 (1.53)
	norms	88.10 (2.75)	87.14 (6.12)	58.33 (8.98)	69.67 (8.48)	94.79 (2.53)
PC	modgm	88.10 (0.00)	81.74 (3.96)	65.00 (5.00)	72.14 (1.55)	94.27 (0.43)
	trace	89.43 (2.70)	74.55 (5.77)	**86.67 (4.71)**	80.07 (4.81)	**96.67 (0.44)**
	geodan	**92.07 (1.12)**	**95.83 (5.89)**	70.00 (0.00)	**80.83 (2.15)**	93.49 (1.04)
	jacs	79.27 (1.22)	56.44 (1.89)	65.00 (5.00)	60.39 (3.25)	82.36 (1.54)
	norms	88.89 (1.77)	**88.09 (5.33)**	61.67 (3.73)	72.55 (4.38)	94.12 (1.48)
BD	modgm	81.74 (1.77)	67.94 (4.14)	43.33 (7.45)	52.74 (6.93)	91.67 (1.08)
	trace	**93.09 (2.60)**	79.43 (4.15)	**96.67 (7.45)**	**87.14 (5.21)**	**97.37 (0.29)**
	geodan	84.92 (1.78)	72.62 (6.94)	60.00 (0.00)	65.57 (2.74)	90.37 (0.99)
	jacs	84.15 (3.07)	67.36 (6.67)	68.33 (6.87)	67.76 (6.23)	91.67 (0.39)
	norms	88.50 (0.89)	**79.40 (3.62)**	70.00 (0.00)	74.36 (1.53)	94.95 (0.87)
WT	modgm	82.93 (0.89)	65.98 (1.55)	58.33 (3.73)	61.89 (2.83)	93.54 (0.50)
	trace	**89.83 (0.91)**	74.62 (0.85)	**88.33 (3.73)**	**80.88 (2.10)**	**95.00 (0.52)**
	geodan	84.92 (1.78)	69.42 (4.85)	66.67 (7.45)	67.64 (4.61)	94.32 (0.55)

Table 1. Mean (standard deviation) of Accuracy, Precision, Sensitivity, F1-Score and ROC area of the fully independent 10-fold cross-validation classification results. Note that the cross-validation population subsets were all the same for all the experiments.

3.1 Comparing full and partial cross-validation

One of our concerns was the effect on the results of the fully independent cross-validation method against non-fully independent cross-validation method. In the first, the voxel selection mask was created using the whole dataset, leading to a double dipped analysis often used in research studies and sometimes difficult to detect in the articles [18]. In table 2 we show the classification results for the non-fully independent cross-validation method, the improvements are significant in most cases.

		Accuracy	Precision	Sensitivity	F1-Score	AUC
	jacs	75.20 (1.68)	49.48 (2.85)	60.00 (0.00)	54.19 (1.70)	79.78 (1.23)
	norms	87.70 (2.14)	85.06 (2.35)	58.33 (8.98)	68.91 (7.45)	93.59 (3.16)
PC	modgm	88.10 (0.00)	77.78 (0.00)	70.00 (0.00)	73.68 (0.00)	95.37 (0.61)
	trace	89.02 (1.86)	72.65 (3.40)	**88.33 (3.73)**	**79.71 (3.38)**	**96.34 (0.61)**
	geodan	**90.48 (1.37)**	**87.96 (6.45)**	70.00 (0.00)	77.86 (2.51)	93.12 (0.94)
	jacs	79.68 (1.15)	57.07 (1.78)	66.67 (4.71)	61.47 (3.06)	81.13 (1.50)
	norms	90.08 (0.89)	**87.20 (0.67)**	68.33 (3.73)	76.58 (2.68)	94.48 (1.68)
BD	modgm	86.91 (1.82)	81.54 (5.98)	58.33 (3.73)	67.98 (4.34)	95.94 (0.31)
	trace	**92.27 (2.60)**	77.44 (4.62)	**96.67 (7.45)**	**85.88 (5.06)**	**97.26 (0.25)**
	geodan	84.92 (1.77)	79.16 (6.36)	50.00 (5.77)	61.11 (5.20)	89.84 (1.03)
	jacs	86.59 (1.86)	72.96 (4.24)	71.67 (3.73)	72.28 (3.70)	92.63 (1.13)
	norms	86.51 (1.13)	**75.93 (1.31)**	63.33 (4.71)	69.01 (3.30)	94.74 (0.49)
WT	modgm	85.71 (0.00)	72.50 (2.50)	65.00 (5.00)	68.34 (1.66)	94.84 (0.56)
	trace	**90.65 (0.91)**	75.32 (0.72)	**91.67 (3.73)**	**82.68 (1.92)**	**95.59 (0.24)**
	geodan	88.10 (0.00)	77.78 (0.00)	70.00 (0.00)	73.68 (0.00)	95.00 (0.48)

Table 2. Mean (standard deviation) of Accuracy, Precision, Sensitivity, F1-Score and ROC area of the non-fully independent 10-fold cross-validation classification results. Note that the cross-validation population subsets were all the same for all the experiments.

3.2 Location of discriminant voxels

The average location of the voxel sites that remained after the threshold are listed in this section. In 1 we show a few slices of the MNI152 template with an overlay with the average voxel sites across the full cross-validation training sets that survived the 95% range threshold. We also checked the locations of the thresholded voxel sites in the training sets separately using probabilistic atlases (MNI structural atlas and the Harvard-Oxford cortical and subcortical atlas). 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.

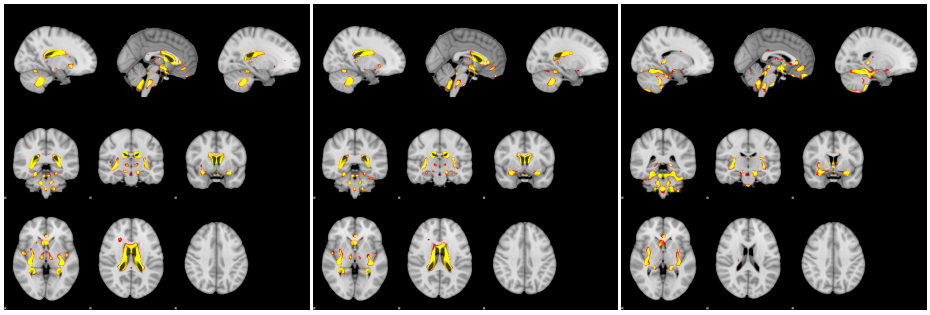

Fig. 1. The average voxel sites across the full cross-validation training sets voxel sites which passed the 95% robust range threshold in the PC (left), BD (middle) and WT (right) of the trace measure.

4 Conclusions

In this paper we report classification results on the application of a feature extraction process based on the deformation vectors obtained from non-linear registration processes. The sample is the complete cross-sectional OASIS database. From the displacement vector we calculated 5 scalar measures and tested them against 3 different methods to measure a distance between both groups. Results show that the deformation vectors can be very useful in full brain detection of AD patients and that the common measures as jacobian determinants (*jacs*) and modulated GM (*modgm*) do not provide the best discrimination results, at least in this case. Overall, taking into account the AUC, the best displacement vector measure was the *trace* and the best distance measure was the Welch's t-test. We are working on the application of non-linear SVM using RBF kernels, and extending the experimental exploration to other classifiers and combinations.

Acknowledgments We thank the Washington University ADRC for making MRI data available. This work has been partially supported by the "Ayudas para la Formación de Personal Investigador" fellowship from the Gobierno del País Vasco.

References

1. 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. Copyright 1946 Indian Statistical Institute.
2. 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. PMID: 12395096.
3. 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.

4. J. Cohen. *Statistical power analysis for the behavioral sciences*. Routledge, 1988.
5. 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. PMID: 10794844.
6. R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot. Automatic classification of patients with alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *NeuroImage*, 56(2):766–781, May 2011.
7. 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. PMID: 17174012 PMCID: 2323584.
8. 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. PMID: 11705634.
9. Y. Fan, D. Shen, and C. Davatzikos. Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*, 8(Pt 1):1–8, 2005.
10. D. Faraggi and B. Reiser. Estimation of the area under the ROC curve. *Statistics in Medicine*, 21(20):3093–3106, October 2002.
11. P. T. Fletcher and S. Joshi. Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing*, 87(2):250–262, February 2007.
12. F. Galton. Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15:246–263, January 1886. Copyright 1886 Royal Anthropological Institute of Great Britain and Ireland.
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.
14. T. Joachims. A support vector method for multivariate performance measures. *Proceedings of the 22nd International Conference on Machine Learning*, pages 377—384, 2005.
15. T. Kailath. The divergence and bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15(1):52–60, February 1967.
16. 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. PMID: 10762153.
17. 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.
18. 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.
19. Z. Lao, D. Shen, Z. Xue, B. Karacali, S. M. Resnick, and C. Davatzikos. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *NeuroImage*, 21(1):46–57, Jan 2004.

20. 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. PMID: 18270068.
21. Y. Liu, L. Teverovskiy, O. Carmichael, R. Kikinis, M.E. Shenton, C.S.. Carter, V..A. Stenger, S. Davis, H. Aizenstein, J.T.. Becker, O.L.. Lopez, and C.C.. Meltzer. Discriminative mr image feature analysis for automatic schizophrenia and alzheimer's disease classification. *Med Image Comput Comput Assist Interv. MICCAI 2004*, 7(Pt 1):393–400, 09 2004.
22. 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. PMID: 17714011.
23. 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. PMID: 19104448.
24. 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.
25. 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. PMID: 15501092.
26. 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.
27. 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.
28. V. N. Vapnik. *Statistical Learning Theory.* Wiley-Interscience, September 1998.
29. B. L. Welch. The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, 34(1/2):28–35, January 1947. Copyright 1947 Biometrika Trust.
30. E. Westman, A. Simmons, Y. Zhang, J.-S. Muehlboeck, C. Tunnard, Y. Liu, L. Collins, A. Evans, P. Mecocci, B. Vellas, M. Tsolaki, I. Kloszewska, H. Soininen, S. Lovestone, C. Spenger, and L.-O. Wahlund. Multivariate analysis of MRI data for alzheimer's disease, mild cognitive impairment and healthy controls. *NeuroImage*, 54(2):1178–1187, January 2011.