Neurocognitive disorder detection based on Feature Vectors extracted from VBM analysis of structural MRI

A. Savio[1], M.T. García-Sebastián[1], D. Chyzyk[1], C. Hernandez[1], M. Graña[1], A. Sistiaga[3,5,6], A. López de Munain[4,5,6], J. Villanúa[2]

December 22, 2010

[1]Grupo de Inteligencia Computacional, www.ehu.es/ccwintco
[2]Osatek S.A., Hospital Donostia San Sebastián, Spain.
[3]Neuroscience Department, Universidad del Pais Vasco UPV-EHU, San Sebastian, Spain
[4]Neurology Service, Donostia Hospital, San Sebastian, Spain
[5]Area de Neurociencias, Instituto Biodonostia
[6]Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Instituto Carlos III, Spain

Abstract

Dementia is a growing concern due to the aging process of the western societies. Non-invasive detection is therefore a high priority research endeavor. In this paper we report results of classification systems applied to the feature vectors obtained by a feature extraction method computed on Structural Magnetic Resonance Imaging (sMRI) volumes for the detection of two neurological disorders with cognitive impairment: Myotonic Dystrophy of type 1 (MD1) and Alzheimer Disease (AD). The feature extraction process is based on the voxel clusters detected by Voxel Based Morphometry (VBM) analysis of sMRI upon a set of patient and control subjects. This feature extraction process is specific for each kind of disease and is grounded on the findings obtained by medical experts. The 10-fold cross-validation results of several statistical and neural network based classification algorithms trained and tested on these features show high specificity and moderate sensitivity of the classifiers, suggesting that the approach is better suited for rejecting than for detecting early stages of the diseases.

1 Introduction

Nowadays, there is much research effort devoted to the development of ways to provide automatized diagnostic support tools that may help the clinicians

to perform their work faster with additional assesment data, to meet the ever increasing demands of primary attention of a rising population of patients with neurological disorders. The present paper will be focused on the application of statistical and Computational Intelligence algorithms for the automatic detection of two very specific pathologies, Alzheimer's Disease (AD) and Myotonic Dystrophy of type 1 (MD1), from the analysis of structural (T1 weighted) Magnetic Resonance Imaging (sMRI) data. The AD is a primary dementia while MD1 is a muscular dystrophy, but both of them show cognitive impairment. The prevalence of MD1 in our local region of Gipuzkoa (Spain) is high[17, 18]. We describe a feature extraction method based on Voxel Based Morphometry (VBM). These features will be the input for several Artificial Neural Network (ANN) and Support Vector Machine (SVM) classification systems.

We have found in the literature several sMRI feature extraction proposals for classification: some based on morphometric methods [16, 20, 37, 65], some based on ROIs/VOIs (regions-of-interest/volumes-of-interest) [45, 43, 24], and some on gray matter (GM) voxels in automated segmentation images [39]. There are also studies aiming to explore the improvement obtained in the SVM classifier by adding covariates such as demographic or genotype information [64]. Work has also been reported on the selection of the most informative features for classification, such as the SVM-Recursive Feature Elimination [20], the selection based on statistical tests [45, 53] or the wavelet decomposition of the RAVENS maps [43], among others.

Our approach uses the VBM detected clusters as a mask to select the potentially most discriminating voxels on the Grey Matter (GM) segmentation volumes. Feature vectors for classification are either the GM segmentation voxel values or some summary statistics of each cluster. We both consider the feature vector computed from all the VBM clusters and the combination of the individual classifiers built from the clusters independently. We test classification systems built using the standard SVM, with linear and non-linear (RBF) kernels, and some ANN architectures: Learned Vector Quantization (LVQ), Multi-Layer Perceptron (MLP), Radial Basis Function (RBF), Probabilistic Neural Networks (PNN). We have also tested combinations of SVM classifiers trained on independent VBM clusters and an Adaptive Boosting (AdaBoost) strategy tailored to the SVM [44]. As a general result, the diverse systems showed a moderate accuracy due to a moderate sensitivity, and high specificity. Best results we obtained with an LVQ approach and an AdaBoost on SVM classifiers. Some of the results have been published separately in conference proceedings [25, 56, 55].

Section 2 gives some background medical information on the studied dementia. Section 3 gives a description of the subjects selected for the study, the image processing and in Section 4 we describe feature extraction details and the classifier systems. Section 5 gives classification performance results obtained in the diverse computational experiments performed on the data. Finally, section 6 gives the conclusions of this work and suggestions for further research .

2 Medical background

2.1 Alzheimer's Disease

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. This is one of the reasons why Magnetic Resonance Imaging (MRI) based early diagnosis is a current research hot topic in the neurosciences. The pharmaceutical companies have already recognized that imaging techniques especially MRI and Positron Emission Tomography (PET) provide "surrogate" information concerning the pattern and rate of neurodegeneration, which can be used to monitor the effects of treatments which slow the progression of neurodegeneration. Therefore, there is high interest in the development of automated detection procedures based on MRI and other medical imaging techniques.

Besides MRI, other medical imaging methods are being studied for AD diagnosis. There are studies applying Support Vector Machine (SVM) either with linear [65, 39, 64] or nonlinear [20, 43] kernels, to discriminate AD patients from controls based on Positron Emission Tomography (PET), Single-Photon Emission Tomography (SPECT) functional volumes [24, 46, 38, 53] or studies that combine structural and functional information such as [19], where sMRI and PET volumes are used.

Many of the classification studies on the detection of AD were done with both men and women. However, it has been demonstrated that brains of women are different from men's to the extent that it is possible to discriminate the gender via MRI analysis [42]. Moreover, it has been shown that VBM is sensitive to the gender differences. For these reasons, we have been very cautious in this study. We have selected a set of 98 MRI women's brain volumes. It must be noted that this is a large number of subjects compared with the other studies referred above.

2.2 Myotonic Dystrophy Type 1

Myotonic Dystrophy type 1 (MD1) is a slowly progressive myopathy characterized by varying multisystemic involvement, affecting skeletal and smooth muscles, the heart (arrhythmia, electrical conductivity defects), the endocrine system (hyperinsulinemia) and eyes (cataract) [34]. It is transmitted in an autosomal dominant manner and it is due to an unstable pathological expansion of (CTG)n repeats [6]. Epidemiologically, MD1 is the most frequent neuromuscular disorder with a reported prevalence between 69 to 90 cases per million [50]. However, the prevalence is significantly higher in Gipuzkoa (North of Spain), reaching 300 cases per million inhabitants [17].

Previous neuroimaging studies using MRI of MD1 patients have found corti-

cal atrophy, increased ventricular size with periventricular hypodensity, subcortical white matter involvement and calcification of the basal ganglia [12, 15, 35]. Many brain disorders have subtle morphological abnormalities of the brain not easily detected on routine examination of MR images [52]; VBM methods have become popular for detecting these abnormalities [3]. In MD1, volumetric studies have identified atrophy of the GM along different cortical areas (mainly in prefrontal area) but these previous studies are methodologically limited regarding to the sample size and the data correction level [2, 51]. Besides, these studies have analyzed the association between brain volumes and some clinical as well as genetic data, but none but none of them have measured the correlation with neuropsychological data.

3 Materials

In this section we will describe the main characteristics of the sMRI data used for the computational experiments. The AD experiments were performed on a subset of the publicly available OASIS database, while the DM1 experiments were performed on a database that has been developed in the Donostia Hospital along the past years of clinical practice with this disease. The data have been obtained from diverse MRI scanners and under different imaging protocols, therefore they can not be combined at the present state of our research to produce AD versus MD1 discriminant systems.

3.1 MD1 Subjects

The MD1 patients analysed in this work were selected from those attending in the outpatient consultancies at the Neurology Department of the Donostia Hospital (San Sebastian), a tertiary public hospital which covers a population of 650,000 inhabitants (almost all of Guipuzcoa province). All patients were explored by a neurologist and had previously participated in another study in which we assessed them neuropsychologically [59]. The patient-selection criteria were as follows: Inclusion criteria for MD1 patients: Between 18 and 65 years old and molecular confirmation of the clinical diagnosis. Exclusion criteria: A history of a major psychiatric or somatic disorder (in accordance with DSM-IV criteria), acquired brain damage or alcohol or drug abuse, the presence of corporal paramagnetic body devices (pacemaker, etc.) that impedes a MRI study and the presence of cerebral anomalies which could affect the volumetric analysis. An age and sex matched healthy control subject (CS) was included for each MD1 patient. This control group consisted of unaffected family members and healthy volunteers with none of these pathologies. All patients were informed of the objectives and details of the study and signed an informed consent. The study was approved by the hospital's ethics committee.

		MD1	CS
Socio-demographic characteristics			
Number of subjects		30	30
Age	Mean (SD)	44.0 (11.6)	44.2 (11.7)
	Min-Max	24-62	22-62
Sex n (%)	Male	14 (47%)	14 (47%)
	Female	16 (53%)	16 (53%)
Educational level n (%)	Primary	18 (60%)	5 (21%)
	Secondary	7 (23%)	9 (37%)
	Higher	5 (17%)	10 (41%)
Clinical and molecular characteristics			
Muscle weakness (MIRS[1])	Mean (SD)	2.9 (1.2)	–
	Min-Max	1-5	
Molecular defect (CTG)	Mean (SD)	635 (472)	–
	Min-Max	65-1833	
White matter lesions n (%)	Yes	16 (53%)	5 (18%)
	No	14 (47%)	22 (82%)

Table 1: Summary of subject demographics and MD1 status. [1]Muscular Impairment Rating Scale.

3.1.1 Imaging Protocol for MD1 data

MR scanning was performed on a 1.5 Tesla scanner (Achieva Nova, Philips). The current results are based on a high-resolution volumetric "turbo field echo" (TFE) series (sagital 3D T1 weighted acquisition, TR = 7.2, TE = 3.3, flip angle = 8, matrix = 256 x 232, slice thickness 1mm, voxel dimensions of 1mm x 1mm x 1mm, NSA = 1, nº slices 160, slice thickness= 1, gap= 0, total scan duration 5´34¨). In addition to 3DT1 weighted images, conventional axial dual T2 turbo spin echo images (TR = 1800, TE = 20, flip angle = 90º , FOV= 230, matrix = 256 x 154, slice thickness= 5, gap= 1, nº acquisition =2, nº slices= 22) and fluid attenuation inversion recovery images (FLAIR) in coronal plane (TR = 10000, TI = 2800, TE= 140 , FOV= 230, matrix = 256 x 159, slice thickness= 5, gap= 1, nº acquisition =2, nº slices= 25) were acquired to evaluate the presence of white matter lesions (WMLs). All the scans were acquired on the same MR scanner and no hardware or software upgrades of the equipment were carried out within the study period.

3.2 OASIS subjects

Ninety eight right-handed women (aged 65-96 yr) were selected from the Open Access Series of Imaging Studies (OASIS) database [48]. OASIS data set has a cross-sectional collection of 416 subjects covering the adult life span aged 18 to 96 including individuals with early-stage Alzheimer's Disease. We have ruled out a set of 200 subjects whose demographic, clinical or derived anatomic volumes information was incomplete. For the present study there are 49 subjects

	Very mild to mild AD	CS
No. of subjects	49	49
Age	78.08 (66-96)	77.77 (65-94)
Education	2.63 (1-5)	2.87 (1-5)
Socioeconomic status	2.94 (1-5)	2.88 (1-5)
CDR (0.5 / 1 / 2)	31 / 17 / 1	0
MMSE	24 (15-30)	28.96 (26-30)

Table 2: Summary of subject demographics and dementia status. Education codes correspond to the following levels of education: 1 less than high school grad., 2: high school grad., 3: some college, 4: college grad., 5: beyond college. Categories of socioeconomic status: from 1 (biggest status) to 5 (lowest status). MMSE score ranges from 0 (worst) to 30 (best).

who have been diagnosed with very mild to mild AD and 49 non-demented. A summary of subject demographics and dementia status is shown in table 3.2.

3.2.1 OASIS Imaging Protocol

The OASIS database has been built following a strict imaging protocol, to avoid variations due to imaging protocol which would pose big image normalization problems. Multiple (three or four) high-resolution structural T1-weighted magnetization-prepared rapid gradient echo (MP-RAGE) images were acquired [21] on a 1.5-T Vision scanner (Siemens, Erlangen, Germany) in a single imaging session. Image parameters: TR= 9.7 msec., TE= 4.0 msec., Flip angle= 10, TI= 20 msec., TD= 200 msec., 128 sagittal 1.25 mm slices without gaps and pixels resolution of 256×256 (1×1mm).

4 Methods

In this section we will describe the different computational processes applied to the data. First, we summarize the Voxel Based Morphometry (VBM) process, and we give details of its independent application to the AD and MD1 data. Second, we describe the feature extraction process. Finally, we give summary descriptions of the diverse approaches used to build the classifiers, both single and combinations of classifiers.

4.1 Voxel-based Morphometry (VBM)

Morphometry analysis has become a common tool for computational brain anatomy studies. It allows a comprehensive measurement of structural differences within a group or across groups, not just in specific structures, but throughout the entire brain. Voxel Based Morphometry (VBM) is a computational approach to neuroanatomy that measures differences in local concentrations of brain tissue, through a voxel-wise comparison of multiple brain images

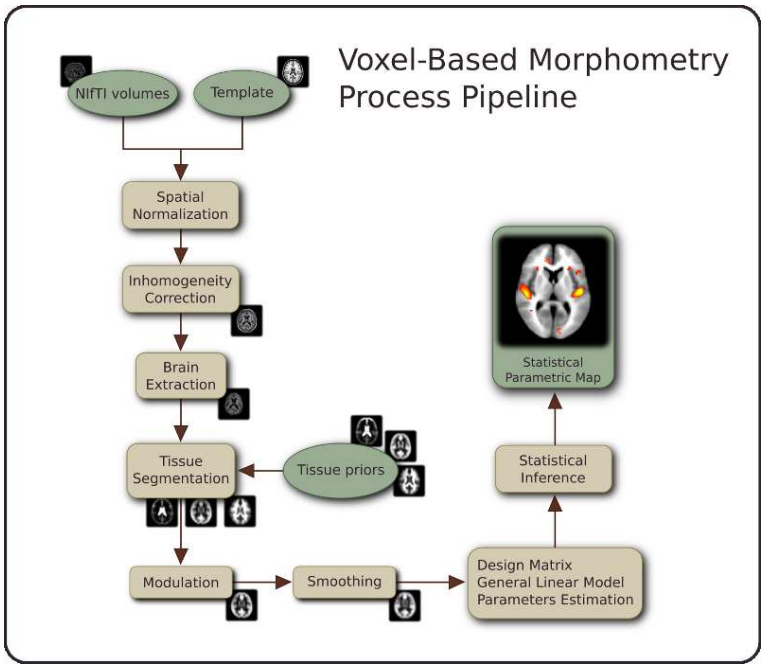

Figure 1: The processing pipeline of the Voxel Based Morphometry (VBM) on structural MRI volumes.

[3]. For instance, VBM has been applied to study volumetric atrophy of the grey matter (GM) in areas of neocortex of AD patients vs. control subjects [9, 57, 23]. The processing pipeline of VBM is illustrated in figure 1. The procedure involves the spatial normalization of subject images into a standard space, segmentation of tissue classes using *a priori* probability maps, smoothing to correct noise and small variations, and voxel-wise statistical tests. Smoothing is done by convolution with a Gaussian kernel whose the Full-Width at Half-Maximum (FWHM) is tuned to the data of ech case study. Statistical analysis is based on the General Linear Model (GLM) to describe the data in terms of experimental and confounding effects, and residual variability, applied to each voxel independently. Statistical inference is used to test hypotheses that are expressed as linear functions of the GLM estimated regression parameters. These linear scalar functions are called contrasts in SPM terminology. The value of this contrast at each voxel constitutes a Statistical Parametric Map (SPM), which is thresholded according to the Random Field theory.

4.1.1 Image processing and VBM for OASIS

We have used the average MRI volume for each subject, provided in the OASIS data set. These images are already registered and re-sampled into a 1-mm isotropic image in atlas space and the bias field has been already corrected [48].

The Statistical Parametric Mapping software (SPM8) [1] was used to compute the VBM which gives us the spatial mask to obtain the classification features. Images were reoriented into a right-handed coordinate system to work with SPM8. The tissue segmentation step does not need to perform bias correction. We performed the modulation normalization for GM, because we are interested in this tissue for this study. We performed a spatial smoothing before performing the voxel-wise statistics, setting the FWHM of the Gaussian kernel to 10mm isotropic. A GM mask was created from the average of the GM segmentation volumes of the subjects under study. Thresholding the average GM segmentation, we obtain a binary mask that includes all voxels with probability greater than 0.1 in the average GM segmentation volume. This interpretation is not completely true, since the data is modulated, but it is close enough for the mask to be reasonable. We designed the statistical analysis as a two-sample t-test in which the first group corresponds with AD subjects. In SPM software jargon: the contrast has been set to [-1 1], a right-tailed (groupN > groupAD), correction FWE, p-value=0.05. The VBM detected clusters are used for the feature extraction for the classification procedures. Statistical significance was determined using an extent threshold of 0 adjacent voxels for two sample comparisons. The clusters of significant voxels detected by the VBM analysis are displayed in figure 2, they agree with the findings reported in the literature [9, 57, 23].

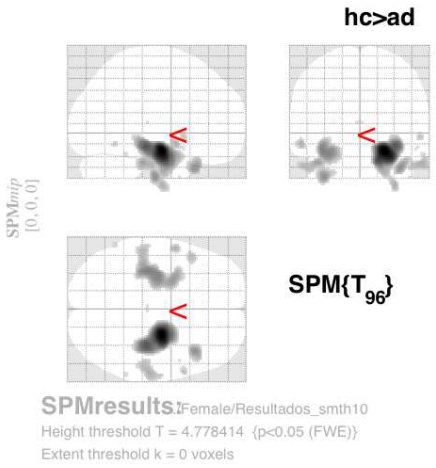

Figure 2: SPM results: clusters of significant voxels with increased gray matter density in the controls relative to the patient subjects, detected by the VBM process.

4.1.2 Image processing and VBM for MD1

The Statistical Parametric Mapping software (SPM8) [1] was used to compute the VBM which gives us the spatial mask to obtain the classification features.

Images were reoriented into a right-handed coordinate system to work with SPM8. The tissue segmentation step does not need to perform bias correction. We performed the modulation normalization for GM, because we are interested in this tissue for this study. We performed a spatial smoothing before performing the voxel-wise statistics, setting the FWHM of the isotropic Gaussian kernel to 8mm, 9mm, 10mm, 11mm and 12mm . For all comparisons we have applied a significance probability threshold of $p<0.05$ FWE corrected for multiple comparisons at a voxel-level. Statistical significance was determined using spatial extent thresholds of 0, 100 and 200 adjacent voxels for two sample comparisons.

4.2 Feature extraction

We have tested two different feature vector extraction processes, based on the voxel location clusters detected as a result of the VBM analysis. The process is illustrated in figure 3. The VBM detected clusters are used as masks to determine the voxel positions where the features are extracted. These masks are applied to the GM density volumes result of the segmentation step in the VBM analysis.

1. The first feature extraction process computes the mean and standard deviation of the GM voxel values of each voxel location cluster, we denote these features as MSD in the result tables given below.

2. The second feature extraction process computes a very high dimensional vector with all the GM segmentation values for the voxel locations included in each VBM detected cluster. The voxel values were ordered in this feature vector according to the coordinate lexicographical ordering. We denote these features as VV in the result tables below.

4.3 Neural Network and Statistical Classification Algorithms

We deal with two class classification problems, given a collection of training/testing input feature vectors $X = \{\mathbf{x}_i \in \mathbb{R}^n,\ i = 1, \ldots, l\}$ and the corresponding labels $\{y_i \in \{-1, 1\},\ i = 1, \ldots, l\}$, which sometimes can be better denoted in aggregated form as a binary vector $\mathbf{y} \in \{-1, 1\}^l$. The algorithms described below build some classifier systems based on this data. The simplest algorithm is the 1-NN which involves no adaptation and uses all the training data samples. The classification rule is of the form:

$$c(\mathbf{x}) = y_{i^*} \text{ where } i^* = \arg \min_{i=1,\ldots,l} \{\|\mathbf{x} - \mathbf{x}_i\|\},$$

that is, the assigned class is that of the closest training vector. To validate their generalization power we use ten fold cross-validation.

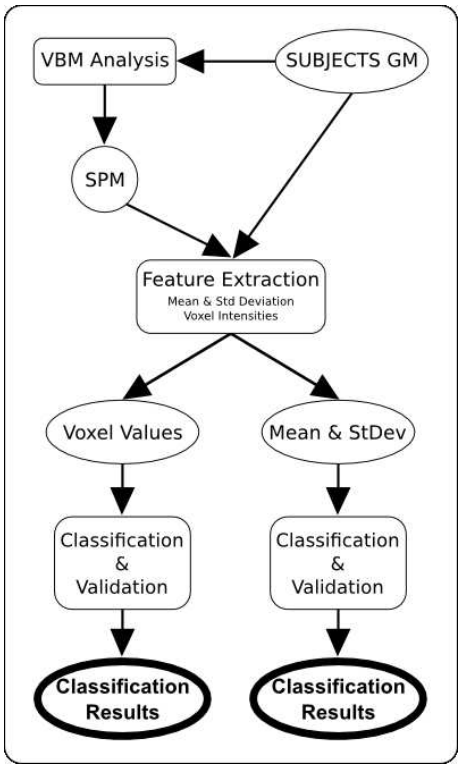

Figure 3: Flow diagram of the feature extraction process from the subjects' GM segmentation volumes.

4.3.1 Support Vector Machines

The Support Vector Machines (SVMs) have attracted attention from the pattern recognition community [24, 62] owing to a number of theoretical and computational merits derived from [63]. SVM separates a given set of binary labelled 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 linearised higher- dimensional space [63]. 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 [63].

The Support Vector Machine (SVM)[63] algorithm used for this study is included in the libSVM (`http://www.csie.ntu.edu.tw/~cjlin/libsvm/`) software package. The implementation is described in detail in [13]. 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 two different kernels were tested: the

linear and the radial basis function (RBF) kernel. The linear kernel function is defined as $K(\mathbf{x}_i, \mathbf{x}_j) = 1+\mathbf{x}_i^T\mathbf{x}_j$, this kernel shows good performance for linearly separable data. The RBF kernel is defined as $K(\mathbf{x}_i, \mathbf{x}_j) = exp(-\frac{||\mathbf{x}_i-\mathbf{x}_j||^2}{2\sigma^2})$. This kernel is best suited to deal with data that have a class-conditional probability distribution function approaching the Gaussian distribution [8]. The RBF kernel is largely used in the literature because it corresponds to the mapping into an infinite dimension feature space, and it can be tuned by its variance parameter σ.

4.3.2 Multi Layer Perceptron trained with Backpropagation

Backward propagation of errors, or backpropagation (BP), [54, 36, 33] is a nonlinear generalization of the squared error gradient descent learning rule for updating the weights of artificial neurons in a single-layer perceptron, generalized to feed-forward networks, also called Multi-Layer Perceptron (MLP). Backpropagation requires that the activation function used by the artificial neurons (or "nodes") is differentiable with its derivative being a simple function of itself. The backpropagation of the error allows to compute the gradient of the error function relative to the hidden units. It is analytically derived using the chain rule of calculus. During on-line learning, the weights of the network are updated at each input data item presentation. We have used the resilient backpropagation, which uses only the derivative sign to perform the weight updating.

We restrict our presentation of BP to train the weights of the MLP for the current two class problem. Let the instantaneous error E_p be defined as:

$$E_p(\mathbf{w}) = \frac{1}{2}\left(y_p - z_K(\mathbf{x}_p)\right)^2, \tag{1}$$

where y_p is the p-th desired output y_p, and $z_K(x_p)$ is the network output when the p-th training exemplar x_p is inputted to the MLP composed of K layers, whose weights are aggregated in the vector $\mathbf{w}$. The output of the j-th node in layer k is given by:

$$z_{k,j}(\mathbf{x}_p) = f\left(\sum_{i=0}^{N_{k-1}} w_{k,j,i} z_{k-1,i}(\mathbf{x}_p)\right), \tag{2}$$

where $z_{k,j}$ is the output of node j in layer k, N_k is the number of nodes in layer k, $w_{k,j,i}$ is the weight which connects the i-th node in layer $k-1$ to the j-th node in layer k, and $f(\cdot)$ is the sigmoid nonlinear function, which has a simple derivative:

$$f'(\alpha) = \frac{df(\alpha)}{d\alpha} = f(\alpha)(1 - f(\alpha)). \tag{3}$$

The convention is that $z_{0,j}(\mathbf{x}_p) = \mathbf{x}_{p,j}$. Let the total error E_T be defined as follows:

$$E_T(\mathbf{w}) = \sum_{p=1}^{l} E_p(\mathbf{w}), \tag{4}$$

where l is the cardinality of X. Note that E_T is a function of both the training set and the weights in the network. The backpropagation learning rule is defined as follows:

$$\Delta w(t) = -\eta \frac{\partial E_p(\mathbf{w})}{\partial w} + \alpha \Delta w(t-1), \tag{5}$$

where $0 < \eta < 1$, which is the learning rate, the momentum factor α is also a small positive number, and w represents any single weight in the network. In the above equation, $\Delta w(t)$ is the change in the weight computed at time t. The momentum term is sometimes used ($\alpha \neq 0$) to improve the smooth convergence of the algorithm. The algorithm defined by equation (5) is often termed as *instantaneous backpropagation* because it computes the gradient based on a single training vector. Another variation is *batch backpropagation*, which computes the weight update using the gradient based on the total error E_T.

To implement this algorithm we must give an expression for the partial derivative of E_p with respect to each weight in the network. For an arbitrary weight in layer k this can be written using the Chain Rule:

$$\frac{\partial E_p(\mathbf{w})}{\partial w_{k,j,j}} = \frac{\partial E_p(\mathbf{w})}{\partial z_{k,j}(\mathbf{x}_p)} \frac{\partial z_{k,j}(\mathbf{x}_p)}{\partial w_{k,j,i}}. \tag{6}$$

Because the derivative of the activation function follows equation 3, we get:

$$\frac{\partial z_{k,j}(\mathbf{x}_p)}{\partial w_{k,j,i}} = z_{k,j}(\mathbf{x}_p)(1 - z_{k,j}(\mathbf{x}_p)) z_{k-1,j}(\mathbf{x}_p), \tag{7}$$

and

$$\frac{\partial E_p(\mathbf{w})}{\partial z_{k,j}(\mathbf{x}_p)} = \sum_{m=1}^{N_{k+1}} \frac{\partial E_p(\mathbf{w})}{\partial z_{k+1,m}(\mathbf{x}_p)} z_{k+1,m}(\mathbf{x}_p)(1 - z_{k+1,m}(\mathbf{x}_p)) w_{k+1,m,j},$$

which at the output layer corresponds to the output error :

$$\frac{\partial E_p(\mathbf{w})}{\partial z_K(\mathbf{x}_p)} = z_L(\mathbf{x}_p) - y_p. \tag{8}$$

4.3.3 Radial Basis Function Networks

Radial Basis Function networks (RBF) [14, 36] are a type of ANN that use radial basis functions as activation functions. RBFs consist of a two layer neural network, where each hidden unit implements a radial activated function. The output units compute a weighted sum of hidden unit outputs. Training consists of the unsupervised training of the hidden units followed by the supervised training of the output units' weights. RBFs have their origin in the solution of

a multivariate interpolation problem [7]. Arbitrary function $g(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ can be approximated by a map defined by a RBF network with a single hidden layer of K units:

$$\hat{g}_{\boldsymbol{\theta}}(\mathbf{x}) = \sum_{j=1}^{K} w_j \phi(\sigma_j, \|\mathbf{x} - \mathbf{c}_j\|), \tag{9}$$

where $\boldsymbol{\theta}$ is the vector of RBF parameters including $w_j, \sigma_j \in \mathbb{R}$, and $\mathbf{c}_j \in \mathbb{R}^n$; let us denote $\mathbf{w} = (w_1, w_2, \ldots, w_p)^T$, then the vector of RBF parameters can be expressed as $\boldsymbol{\theta}^T = \left(\mathbf{w}^T, \sigma_1, \mathbf{c}_1^T, \ldots, \sigma_K, \mathbf{c}_K^T\right)$. Each RBF is defined by its center $\mathbf{c}_j \in \mathbb{R}^n$ and width $\sigma_j \in \mathbb{R}$, and the contribution of each RBF to the network output is weighted by w_j. The RBF function $\phi(\cdot)$ is a nonlinear function that monotically decreases as $\mathbf{x}$ moves away from its center $\mathbf{c}_j$. The most common RBF used is the isotropic Gaussian:

$$\hat{g}_{\boldsymbol{\theta}}(\mathbf{x}) = \sum_{j=1}^{p} w_j \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_j\|^2}{2\sigma_j^2}\right).$$

The network can be thought as the composition of two functions $\hat{g}_{\boldsymbol{\theta}}(\mathbf{x}) = W \circ \Phi(\mathbf{x})$, the first one implemented by the RBF units $\Phi : \mathbb{R}^n \to \mathbb{R}^K$ performs a data space transformation which can be a dimensionality reduction or not, depending on whether $K > n$. The second function corresponds to a single layer linear Perceptron $W : \mathbb{R}^K \to \mathbb{R}$ giving the map of the RBF transformed data into the class labels. Training is accordingly decomposed into two phases. First a clustering algorithm is used to estimate the Gaussian RBF parameters (centres and variances). Afterwards, linear supervised training is used to estimate the weights from the hidden RBF to the output. In order to obtain a binary class label output, a hard limiter function is applied to the continuous output of the RBF network.

4.3.4 Probabilistic Neural Networks

A Probabilistic Neural Network (PNN) [61] uses a kernel-based approximation to form an estimate of the probability density function of categories in a classification problem. In fact, it is a generalization of the Parzen windows distribution estimation, and a filtered version of the 1-NN classifier. The distance of the input feature vector $\mathbf{x}$ to the stored patterns is filtered by a RBF function. Let us denote the data sample partition as $X = X_1 \cup X_{-1}$, where $X_1 = \left\{\mathbf{x}_1^1, \ldots, \mathbf{x}_{n_1}^1\right\}$ and $X_1 = \left\{\mathbf{x}_1^{-1}, \ldots, \mathbf{x}_{n_{-1}}^{-1}\right\}$. That is, superscripts denote the class of the feature vector and $n_1 + n_{-1} = n$. Each pattern $\mathbf{x}_j^i$ of training data sample is interpreted as the weight of the j-th neuron of the i-th class. Therefore the response of the neuron is computed as the probability of the input feature vector according to a Normal distibution centered at the stored pattern:

$$\Phi_{i,j}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left[-\frac{\left\|\mathbf{x} - \mathbf{x}_j^i\right\|}{2\sigma^2}\right], \tag{10}$$

Therefore the output of the neuron is inside $[0, 1]$. The tuning of a PNN network depends on selecting the optimal sigma value of the spread σ of the RBF functions, which can be different for each class. In this paper an exhaustive search for the optimal spread value in the range (0, 1) for each training set has been done. The output of the PNN is an estimation of the likelihood of the input pattern $\mathbf{x}$ being from class $i \in \{-1, 1\}$ by averaging the output of all neurons that belong to the same class:

$$p_i(x) = \frac{1}{n_i} \sum_{j=1}^{n_i} \Phi_{i,j}(\mathbf{x}). \tag{11}$$

The decision rule based on the output of all the output layer neurons is simply:

$$\hat{y}(\mathbf{x}) = \arg\max_i \{p_i(\mathbf{x})\}, \quad i \in \{-1, 1\}. \tag{12}$$

where $\hat{y}(\mathbf{x})$ denotes the estimated class of the pattern $\mathbf{x}$. If the a priori probabilities for each class are the same, and the losses associated with making an incorrect decision for each class are the same, the decision layer unit classifies the pattern $\mathbf{x}$ in accordance with the optimal Bayes' rule.

4.3.5 Learning Vector Quantization

Learning vector quantization (LVQ) [40, 60] Learning Vector Quantization (LVQ) as introduced by Kohonen [41] represents every class $c \in \{-1, 1\}$ by a set $W(c) = \{\mathbf{w}_i \in \mathbb{R}^n; i = 1, \ldots, N_c\}$ of weight vectors (prototypes) which tesselate the input feature space. Let us denote W the union of all prototypes, regardless of class. If we denote c_i the class the weight vector $\mathbf{w}_i \in W$ is associated with, the decision rule that classifies a feature vector $\mathbf{x}$ is as follows:

$$c(\mathbf{x}) = c_{i^*}$$

where

$$i^* = \arg\min_i \{\|\mathbf{x} - \mathbf{w}_i\|\}.$$

The training algorithm of LVQ aims at minimizing the classification error on the given training set, i.e., $E = \sum_j (y_j - c(\mathbf{x}_j))^2$, modifying the weight vectors on the presentation of input feature vectors. The heuristic weight updating rule is as follows:

$$\Delta\mathbf{w}_{i^*} = \begin{cases} \epsilon \cdot (\mathbf{x}_j - \mathbf{w}_{i^*}) & \text{if } c_{i^*} = y_j \\ -\epsilon \cdot (\mathbf{x}_j - \mathbf{w}_{i^*}) & \text{otherwise} \end{cases}, \tag{13}$$

that is, the input's closest weight is adapted either toward the input if their classes match, or away from it if not. This rule is highly unstable, therefore, the practical approach consists in performing an initial clustering of each class data samples to obtain an initical weight configuration using equation 13 to perform the fine tuning of the classification boundaries. This equation corresponds to

a LVQ1 approach. The LVQ2 approach involves determining the two input vector's closest weights. They are moved toward or away the input according to the matching of their classes.

4.3.6 Combination of independent SVMs trained per VBM cluster

We have considered also the construction of independent SVM classifiers for each VBM detected cluster, meaning that only the corresponding features are used for train/test, and the combination of their responses in two ways:

1. By a simple majority voting, using the cluster with greatest statistical significance to resolve ties. This can be viewed as a simplified combination of classifiers. We denote this system Indep-SVM in the results reported below.

2. We have defined a combination of classifiers weighted by the individual training errors, where the classifier weights are computed using the AdaBoost SVM algorithm [44], assuming an uniform weighting of the data samples. We present this approach in Algorithm 1. We denote AB-SVM this approach in the tables below.

4.3.7 Adaptive Boosting

Adaptive Boosting (AdaBoost)[58, 22] is a meta-algorithm for machine learning that can be used in conjunction with many other learning algorithms to improve their performance. AdaBoost is adaptive in the sense that subsequent classifiers built are tweaked in favor of those instances misclassified by previous classifiers. AdaBoost is sensitive to noisy data and outliers. Otherwise, it is less susceptible to the over-fitting problem than most learning algorithms.

AdaBoost calls a weak classifier repeatedly in a series of rounds $t = 1, ..., T$. For each call a distribution of weights W_t is updated and indicates the importance of examples in the data set for the classification. On each round, the weights of each incorrectly classified example are increased (or alternatively, the weights of each correctly classified example are decreased), so that the new classifier focuses more on those examples.

Following these ideas, we have also tested a combination of SVM classifiers following the the Diverse-AdaBoost-SVM [44], shown here as Algorithm 2. In this approach we built a sequence of SVM classifiers of increasing variance parameter. The results of the classifiers are weighted according to their statistical error to obtain the response to the test inputs in the 10-fold validation process. We denote DAB-SVM this approach in the tables below.

5 Computational Experiments Results

We evaluate the performance of the classifiers built with the diverse strategy using a 10-fold cross-validation methodology, illustrated in figure 4. We have

Algorithm 1 Combining the independent SVM trained per cluster

1. **Input:** as many sets of training samples with labels as clusters in the statistical parametric map $T_k = \{(x_1, y_1), \ldots, (x_N, y_N)\}, k = 1..C$, where N is the number of samples of each cluster.

2. **Initialize:** the weights of training samples: $w_i^k = 1/N$, for all $i = 1, ..., N$

3. **For each *k* cluster do**

   (a) Search the best γ for the RBF kernel for the training set T_k, we denote it as γ_k.

   (b) Train the SVM with T_k and γ_k, we denote the classifier as h_k.

   (c) Classify the same training T_k set with h_k.

   (d) Calculate the training error of h_k: $\epsilon_k = \sum_{i=1}^{N} w_i^k, \quad y_i \neq h_k(x_i)$.

   (e) Compute the weight of the cluster classifier h_k: $\alpha_k = \frac{1}{2}\ln(\frac{\epsilon_k}{1-\epsilon_k})$.

4. **Output:** for each test data x its classification is $f(x) = sign(\sum_{k=1}^{C} \alpha_k h_k(x))$.

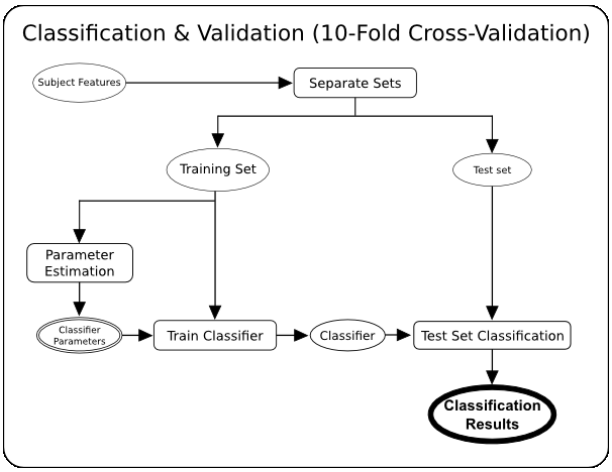

Figure 4: Flow chart of the 10-fold cross validation procedure followed in the experiments reported in this paper.

Algorithm 2 Diverse AdaBoostSVM

1. **Input:** a set of training samples with labels $\{(x_1, y_1), \ldots, (x_N, y_N)\}$; the initial σ, σ_{ini}; the minimal σ, σ_{min}; the step of σ, σ_{step}; the threshold on diversity DIV.
2. **Initialize:** the weights of training samples: $w_i^t = 1/N$, for all $i = 1, ..., N$
3. **Do while** ($\sigma > \sigma_{ini}$)
   (a) Calculate gamma: $\gamma = \left(2\sigma^2\right)^{-1}$.
   (b) Use σ to train a component classifier h_t on the weighted training set.
   (c) Calculate the training error of h_t: $\epsilon_t = \sum_{i=1}^{N} w_i^t, \quad y_i \neq h_t(x_i)$.
   (d) Calculate the diversity of h_t: $D_t = \sum_{i=1}^{N} d_t(x_i)$, where $d_t(x_i) = \begin{cases} 0 & if\ h_t(x_i) = y_i \\ 1 & if\ h_t(x_i) \neq y_i \end{cases}$
   (e) Calculate the diversity of weighted component classifiers and the current classifier: $D = \sum_{t=1}^{T} \sum_{i=1}^{N} d_t(x_i)$.
   (f) If $\epsilon_t > 0.5$ or $D < DIV$: decrease σ by σ_{step} and go to (a).
   (g) Set weight of the component classifier h_t: $\alpha_t = \frac{1}{2} \ln(\frac{\epsilon_t}{1-\epsilon_t})$.
   (h) Update the weights of training samples: $w_i^{t+1} = w_i^t exp(-\alpha y_i h_t(x_i))$.
   (i) Normalize the weights of training samples: $w_i^{t+1} = w_i^{t+1}(\sum_{i=1}^{N} w_i^{t+1})^{-1}$.
4. **Output:** $f(x) = sign(\sum_{k=1}^{C} \alpha_k h_k(x))$.

enconded the classes as follows: 0 for patients, 1 for controls. Positives correspond to class 0. To quantify the results we measured the Accuracy defined as the ratio of the number of test volumes correctly classified to the total of tested volumes. We also quantified the sensitivity and specificity of each test defined as $Sensitivity = \frac{TP}{TP+FN}$ and $Specificity = \frac{TN}{TN+FP}$, where true positives (TP) is the number of patient volumes correctly classified; true negatives (TN) is the number of control volumes correctly classified; false positives (FP) is the number of control volumes classified as diseased patients and false negatives (FN) is the number of diseased patient volumes classified as control subjects. As the image assessment is an additional finding meant to support other diagnostic information sources, there is a specific need for high sensitivity and specificity systems. Thus, these performance measure were preferred above others, like the F-measure or the Area under ROC, which are better suited to explore and discuss the parameter sensitivity of the classification building algorithm [47].

5.1 Results on AD detection on the OASIS data

We report the average accuracy, sensitiviy and specificity of the 10-fold cross-validation of the systems developed for AD detection computed over the OASIS data. For all the classifiers, we have determined the optimal values of the classifier parameters via independent grid searches performed at each cross-validation fold. For the SVM we searched for the optimal C and γ values. For the MLP-BP we look for the optimal number of hidden units. For the RBF and PNN, the spread of the radial basis functions was determined. For the LVQ the number of hidden units was determined. For the Diverse AdaBoost SVM (DAB-SVM) the parameters of the AdaBoost were set as follows: $\sigma_{min} = 0.1$, $\sigma_{ini} = 100$, $\sigma_{step} = 0.1$. The DIV value is set as as 0.6.

The results on the MSD features (24 values from each volume) are presented in table 3. The best accuracy results were obtained by the non linear SVM with RBF kernels, specially when embedded in the AdaBoost process (0.85). The LVQ approach give almost comparable results. Overall, it must be noted that the specificity is systematically higher than the sensitivity for all the classification strategies tested. Detailed examination of the results reveals that the decrease of sensitivity is due to the bad recognition results on the AD subjects at early stages of the disease. As the trend is common to all the classification algorithms tested, this lack of sensitivity must be attributed to the feature vector computed from the VBM detected clusters.

The average results of 10-fold cross-validation tests computed on the VV features, are presented in table 4. For this definition of the feature vector, the number of features is 3611. Again the specificity is systematically greater than the classifier's sensitivity at all cases, suggesting that the feature vector definition must be improved to cope with the early detection cases. We obtain a best accuracy result with the AdaBoost applied to the SVM with RBF kernel trained independently at each VBM detected cluster (rbf-AB-SVM). Despite this result, which is the best accuracy found, the remaining classifier results are significatively worse for the VV features than for the MSD features. This must

Classif.	Accuracy	Sensitivity	Specificity
linear SVM	0.78	0.72	0.88
rbf SVM	**0.81**	0.75	0.89
MLP-BP	0.78	0.69	0.88
RBF	0.66	0.65	0.68
PNN	0.78	0.62	0.94
LVQ1	**0.81**	0.72	0.90
LVQ2	**0.83**	0.74	0.92
Indep-linear-SVM	0.74	0.51	0.97
Indep-rbf-SVM	0.75	0.56	0.95
linear-AB-SVM	0.71	0.54	0.88
rbf-AB-SVM	0.79	0.78	0.80
rbf-DAB-SVM	**0.85**	0.78	0.92

Table 3: Results over the MSD features computed from the OASIS data for AD detection

be accounted to fact that the VV dimensionality is several orders of magnitude greater than the MSD and the proposed classifiers are unable to deal with that and the the consequent data scarcity. Note that the best result in table 4 corresponds to the independent training of SVM for each detected cluster. This "divide and conquer" strategy seems to produce good results in this case.

5.2 MD1 Results

In the computational experiments done on the MD1 data, we focus on the SVM classifier and explore the effect of the various VBM parameters on the resulting classification accuracy, sensitivity and specificity. In the tables below we show results for various FWHM and spatial cluster extent threshold (Size-Thr) value settings in the SPM processing to obtain the VBM clusters of pixels with significant differences. Note that the increase in the FWHM parameter naturally implies an increase in the number of features considered. The increase of the Size-Thr parameter implies that more clusters are rejected because of their small size, and the resulting feature vector has lower dimensionality. The results on the MSD features are presented in table 5, while the results on the VV features are presented in table 6. Note that there are few instances that reach the 80% accuracy. Notice also in both tables that the specificity is systematically greater than the sensitivity. As was said in the AD study above, this trend forces to attribute the lack of sensitivity to the feature extraction process. In order to deal with the milder dementia cases, the process must be improved. It must be also taken into account that until now there have not been findings in sMRI associated with the MD1, meaning that the works reported here are the among the first published results in this sense. Another interesting effect that can be appreciated in the tables 5 and 6 is that growing number of features, obtained with stronger smoothing does not imply growing classification

Classif.	Accuracy	Sensitivity	Specificity
linear SVM	0.73	0.72	0.75
rbf SVM	0.76	0.77	0.76
MLP-BP	0.78	0.72	0.84
RBF	0.72	0.65	0.80
PNN	0.74	0.68	0.81
LVQ1	0.79	0.76	0.82
LVQ2	0.77	0.76	0.78
Indep-linear-SVM	0.77	0.74	0.80
Indep-rbf-SVM	0.78	0.76	0.82
linear-AB-SVM	0.73	0.76	0.70
rbf-AB-SVM	**0.86**	0.80	0.92
rbf-DAB-SVM	0.78	0.71	0.85

Table 4: Results over the VV features computed fom the OASIS data for AD detection

accuracy. This goes contrary to the assessments by human experts that favor stronger smoothing in order to obtain better and bigger detections of the effect.

5.3 Discussion of results.

The classifiers applied to the data features make decisions based on the whole of the feature vector. However, the locations of the feature sources, the voxel cluster found by VBM, has some clinical interpretations. For AD the voxel cluster findings were mostly located in the temporal lobe, though some voxels in the frontal lobe have also been detected. These results confirm previous findings of temporal lobe atrophic changes in AD [4, 11, 10] [...], and suggest that these abnormalities may be confined to specific sites within that lobe, rather than showing a widespread distribution. The classification results are a second level assessment of these findings, because they show that it is possible some degree of AD prediction from them.

For MD1 the most discriminant voxels were in the caudate nucleus, fronto-parietal lobe lobe and thalamus. These results involving the fronto-parietal areas, agree with previous studies [2] and with the profile suggested by our previous neuropsychological results in a large DM1 sample [59]. In fact, we have found the volume corresponding to the caudate nuclei and the thalamus to be smaller in DM1 patients than in the CS. Basal ganglia have traditionally been associated with motor processes, although there is increasing evidence to support their role in cognitive functions [49]. Basal ganglia nuclei relate to one another and to the cortex through different cortico-striatal loops, which emphasizes the functional interrelationship between the neocortex, the striatum and the thalamus. Three of the five loops in which the striatum is involved pass through the caudate and the thalamus on the way to the cortex: the oculomotor, dorsolateral prefrontal and lateral orbitalfrontal loops [31]. The

FWHM(mm)	Size-Thr	#Features	Accuracy	Sensitivity	Specificity
8	0	76	0.78	0.73	0.83
	100	8	0.77	0.67	0.87
	200	4	0.77	0.67	0.87
9	0	76	**0.80**	0.70	0.90
	100	16	0.75	0.67	0.83
	200	4	0.76	0.67	0.87
10	0	70	0.78	0.63	0.93
	100	22	0.77	0.73	0.80
	200	8	0.78	0.70	0.87
11	0	64	0.72	0.63	0.80
	100	24	0.75	0.63	0.87
	200	12	0.75	0.63	0.87
12	0	68	0.72	0.63	0.80
	100	36	0.73	0.63	0.83
	200	18	0.75	0.70	0.80

Table 5: SVM classification results (10-fold crossvalidation) for MSD features, based on t-test VBM of the data, FWE=0.05

FWHM	Threshold	Features	Accuracy	Sensitivity	Specificity
8	0	2059	**0.82**	0.83	0.80
	100	1226	0.78	0.70	0.87
	200	958	**0.80**	0.80	0.77
9	0	2826	0.78	0.73	0.83
	100	2044	0.77	0.73	0.80
	200	1182	0.75	0.67	0.83
10	0	3710	0.77	0.73	0.80
	100	3103	**0.80**	0.77	83
	200	2131	0.73	0.70	0.77
11	0	5022	0.73	0.73	0.73
	100	4278	0.78	0.73	0.83
	200	3434	0.75	0.70	0.80
12	0	6542	0.76	0.73	0.80
	100	6391	0.75	0.70	0.80
	200	5148	0.73	0.70	0.76

Table 6: SVM classification results (10-fold crossvalidation) for VV features, based on t-test VBM of the data, FWE=0.05

last two target two prefrontal cortical areas that are thought to be involved in various aspects of cognitive behaviour. In fact, the dorsolateral cortex is the structure mostly closely associated with executive functions and allows the organization of information to facilitate a response. The orbitofrontal circuit allows the integration of limbic and emotional information with behavioural responses [5].

6 Conclusions

In this work we have studied several supervised classification systems applied to discriminate patients with neurocognitive disorders (AD and MD1) from control subjects based on structural MRI (T1-weighted) data. The feature extraction processes is based on the voxel clusters detected by a VBM analysis. For the discrimination between AD patients and controls we achieve the construction of classifiers with an accuracy of 0.86 in the best case shown in table 4 in the case of OASIS females and 0.82 in case of MD1 subjects. A result of 86% of accuracy is really encouraging considering the number of subjects in the database and all the biases and errors involved in the registration, segmentation and smoothing processes performed in the pre-processing steps of the volumes in the VBM. After close examination of the results in the AD study, we have found that the subjects wrongly classified maybe the most critical ones: old control subjects classified as AD (false positives) and subjects with a very early or mild dementia classified as normal (false negatives), exactly the ones which are the target in these studies that try to perform early detection of AD. Post-mortem confirmation data of AD diagnosed subjects could improve the results. Something similar may be happening in the MD1 study. Therefore we may conclude that additional information sources and improved classification strategies are needed to reach this additional accuracy increase that would cover the most difficult cases.

Further work may be directed in the following lines:

- The consideration of features extracted on the basis of information obtained from other morphological measurement techniques, such as Deformation-based Morphometry and Tensor-based Morphometry.

- Use additional image modalities (PET, fMRI, DTI) and additional clinical data. Additional image modalities imply the mutual registration of volumes and the fusion of the diverse information sources. Additional clinical data may be used as covariates in the GLM resolution within the VBM analysis.

- Using new classification strategies, such as the uncertain classifiers, which may assign various grades to the data and provide new ways to evaluate the classifier response[47]. In the case of of pathologies with cognitive impairment, it would be more natural to try to rank the image data according to the neuropsychological scales than the binary decision that we

have been trying to implement in this paper, improving results in several ways. Also Lattice Computing approaches [28, 30, 26, 29, 27, 32]

- Future work may be addressed to the problem of three way discrimination AD vs. MD1 vs. controls. It needs an elaborated VBM using F test for the detection of clusters of voxels which can discriminate between the three classes, as well as multi-class classifiers, which in some cases requires some elaboration.

Acknowledgements

We thank the Washington University ADRC for making the OASIS database available. Research partially supported by Saiotek research projects BRAINER, S-PR07UN02, SOFLIMRI.

References

[1] http://www.fil.ion.ucl.ac.uk/spm/.

[2] G Antonini, C Mainero, A Romano, F Giubilei, V Ceschin, F Gragnani, S Morino, M Fiorelli, F Soscia, A Di Pasquale, and F Caramia. Cerebral atrophy in myotonic dystrophy: a voxel based morphometric study. *Journal of Neurology, Neurosurgery & Psychiatry*, 75(11):1611–1613, November 2004.

[3] J. Ashburner and K.J. Friston. Voxel-Based Morphometry–The methods. *NeuroImage*, 11(6):805–821, June 2000.

[4] J. C. Baron, G. Chetelat, B. Desgranges, G. Perchey, B. Landeau, V. de la Sayette, and F. Eustache. In vivo mapping of gray matter loss with Voxel-Based morphometry in mild alzheimer's disease. *Neuroimage*, 14(2):298–309, 2001.

[5] R.M. Bonelli and J.L. Cummings. Frontal-subcortical circuitry and behavior. *Dialogues Clin. Neurosci.*, 9:141–151, 2007.

[6] J.D. Brook, M.E. McCurrach, H.G. Harley, A.J. Buckler, D. Church, H. Aburatani, K. Hunter, V.P. Stanton, J.P. Thirion, and T. Hudson. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell*, 68(4):799–808, February 1992. PMID: 1310900.

[7] D.S. Broomhead and D. Lowe. Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2:321–355, 1988.

[8] C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):167, 121, 1998.

[9] G. F. Busatto, G. E. J. Garrido, O. P. Almeida, C. C. Castro, C. H. P. Camargo, C. G. Cid, C. A. Buchpiguel, S. Furuie, and C. M. Bottino. A voxel-based morphometry study of temporal lobe gray matter reductions in alzheimer's disease. *Neurobiology of Aging*, 24(2):221–231, 2003.

[10] Geraldo Busatto. Voxel-Based Morphometry in Alzheimers disease. *Expert review of neurotherapeutics*, 8(11):1691–1702, 2008.

[11] G.F. Busatto, G.E. Garrido, O.P. Almeida, C.C. Castro, C.H. Camargo, C.G. Cid, C.A. Buchpiguel, S. Furuie, and C.M. Bottino. A Voxel-Based Morphometry study of temporal lobe gray matter reductions in Alzheimer's disease. *Neurobiol Aging*, 24(2):221–231, Mar-Apr 2003.

[12] B. Censori, L. Provinciali, M. Danni, L. Chiaramoni, M. Maricotti, N. Foschi, M. Del Pesce, and U. Salvolini. Brain involvement in myotonic dystrophy: MRI features and their relationship to clinical and cognitive conditions. *Acta Neurologica Scandinavica*, 90(3):211–217, September 1994. PMID: 7847063.

[13] C.C. Chang and C.J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[14] S. Chen, C.F.N. Cowan, and P.M. Grant. Orthogonal least squares learning algorithm for radial basis function networks. *Neural Networks, IEEE Transactions on*, 2(2):302–309, 1991.

[15] M.S. Damian, G. Schilling, G. Bachmann, C. Simon, S. Stöppler, and W. Dorndorf. White matter lesions and cognitive deficits: relevance of lesion pattern? *Acta Neurologica Scandinavica*, 90(6):430–436, December 1994. PMID: 7892763.

[16] C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S.M. Resnick. Detection of prodromal alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiol Aging*, 29(4):514–523, Apr 2008.

[17] A. Lopez de Munain, A. Blanco, J. I. Emparanza, J. J. Poza, J. F. Marti Masso, A. Cobo, L. Martorell, M. Baiget, and J. M. Martinez Lage. Prevalence of myotonic dystrophy in guipuzcoa (Basque country, spain). *Neurology*, 43(8):1573, August 1993.

[18] A. Lopez de Munain, J. I. Emparanza, A. Blanco, A. Cobo, J. J. Poza, B. Basauri, and et al. Clinical manifestations of myotonic dystrophy: epidemiologic survey. *Med.Clin.(Barc.)*, 101:161–164, August 1993.

[19] Y. Fan, , S.M. Resnick, S. Wu, and C. Davatzikos. Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. *NeuroImage*, 41(2):277–285, 2008.

[20] 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.

[21] A.F. Fotenos, A.Z. Snyder, L.E. Girton, J.C. Morris, and R.L. Buckner. Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. *Neurology*, 64(6):1032–1039, March 2005.

[22] Y. Freund and R. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *European Conference on Computational Learning Theory*, pages 37, 23, 1995.

[23] G. B. Frisoni, C. Testa, A. Zorzan, F. Sabattoli, A. Beltramello, H. Soininen, and M. P. Laakso. Detection of grey matter loss in mild alzheimer's disease with voxel based morphometry. *Journal of Neurology, Neurosurgery & Psychiatry*, 73(6):657–664, 2002.

[24] G. Fung and J. Stoeckel. SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information. *Knowl. Inf. Syst.*, 11(2):243–258, 2007.

[25] M.T. García-Sebastián, A. Savio, M. Graña, and J. Villanúa. On the use of morphometry based features for Alzheimer's disease detection on MRI. In Alberto Prieto Juan M. Corchado (Editors) Joan Cabestany, Francisco Sandoval, editor, *Bio-Inspired Systems: Computational and Ambient Intelligence. / IWANN 2009 (Part I)*, volume 5517 of *LNCS*, pages 957–964. Springer-Verlag, 2009.

[26] M. Graña. A brief review of Lattice Computing. In *Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence). IEEE International Conference on*, pages 1777–1781, June 2008.

[27] M. Graña, A. Manhaes-Savio, M. Garcia-Sebastian, and E. Fernandez. A Lattice Computing approach for on-line fMRI analysis. *Image and Vision Computing*, 28(7):1155–1161, 2010. Online pattern recognition and machine learning techniques for computer-vision: Theory and applications.

[28] M. Graña, P. Sussner, and G.X. Ritter. Associative Morphological Memories for endmember determination in spectral unmixing. In *Fuzzy Systems, 2003. FUZZ '03. The 12th IEEE International Conference on*, volume 2, pages 1285–1290, May 2003.

[29] M. Graña, I. Villaverde, J.O. Maldonado, and C. Hernandez. Two Lattice Computing approaches for the unsupervised segmentation of hyperspectral images. *Neurocomputing*, 72(10-12):2111–2120, 2009.

[30] M. Graña, I. Villaverde, R. Moreno, and F.X. Albizuri. *Computational Intelligence Based on Lattice Theory*, chapter Convex Coordinates From

Lattice Independent Sets for Visual Pattern Recognition, pages 99–126. Springer-Verlag, 2007.

[31] J.A. Grahn, J.A. Parkinson, and A.M. Owen. The cognitive functions of the caudate nucleus. *Prog Neurobiol.*, 86:141–155, 2008.

[32] M. Graña, D. Chyzhyk, M. García-Sebastián, and C. Hernández. Lattice independent component analysis for functional magnetic resonance imaging. *Information Sciences*, in press, 2010.

[33] M.T. Hagan, H.B. Demuth, and M.H. Beale. *Neural Network Design*. PWS Pub. Co., Har/Dsk edition, December 1995.

[34] P. Harper. *Myotonic Dystrophy*. Oxford University Press, USA, 2 edition, October 2009.

[35] T. Hashimoto, M. Tayama, M. Miyazaki, K. Murakawa, H. Kawai, H. Nishitani, and Y. Kuroda. Neuroimaging study of myotonic dystrophy. II. MRI measurements of the brain. *Brain and Development*, 17(1):28–32, 1995.

[36] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 2 edition, July 1998.

[37] C. Huang, B. Yan, H. Jiang, and D. Wang. Combining voxel-based morphometry with artifical neural network theory in the application research of diagnosing alzheimer's disease. In *BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on*, volume 1, pages 250–254, May 2008.

[38] I.A. Illan, J.M. Gorriz, J. Ramirez, D. Salas-Gonzalez, M.M. Lopez, F. Segovia, R. Chaves, M. Gomez-Rio, and C.G. Puntonet. 18f-fdg pet imaging analysis for computer aided alzheimer's diagnosis. *Information Sciences*, 181(4):903 – 916, 2011.

[39] S. Kloppel. Automatic classification of MR scans in Alzheimer's disease. *Brain*, 131(3):681–689, 2008.

[40] T. Kohonen. *Self-organization and associative memory: 3rd edition*. Springer-Verlag New York, Inc., 1989.

[41] T. Kohonen. Learning vector quantization. In *The handbook of brain theory and neural networks*, pages 537–540. MIT Press, 1998.

[42] 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, 2004.

[43] 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.

[44] X. Li, L. Wang, and E. Sung. A study of AdaBoost with SVM based weak learners. In *Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on*, volume 1, pages 196–201 vol. 1, 2005.

[45] Y. Liu, L. Teverovskiy, O. Carmichael, R. Kikinis, M. 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. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 3216:393–401, 2004.

[46] M. Lopez, J. Ramirez, J.M. Gorriz, D. Salas-Gonzalez, I. Alvarez, F. Segovia, and C.G. Puntonet. Automatic tool for alzheimer's disease diagnosis using pca and bayesian classification rules. *Electronics Letters*, 45(8):389–391, 9 2009.

[47] O. Luaces, F. Taboada, G.M. Albaiceta, L.A. Domínguez, P. Enriquez, 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, 2009.

[48] 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.

[49] F.A. Middleton and P.L. Strick. Basal ganglia output and cognition: evidence from anatomical, behavioral, and clinical studies. *Brain Cogn.*, 42:183–200, 2000.

[50] M. L. Mostacciuolo, G. Barbujani, M. Armani, G. A. Danieli, C. Angelini, and D. C. Rao. Genetic epidemiology of myotonic dystrophy. *Genetic Epidemiology*, 4(4):289–298, 1987.

[51] M. Ota, N. Sato, Y. Ohya, Y. Aoki, K. Mizukami, T. Mori, and T. Asada. Relationship between diffusion tensor imaging and brain morphology in patients with myotonic dystrophy. *Neuroscience Letters*, 407(3):234–239, October 2006.

[52] G.S. Pell, R.S. Briellmann, C.H. P. Chan, H. Pardoe, D. F. Abbott, and G. D. Jackson. Selection of the control group for VBM analysis: influence of covariates, matching and sample size. *NeuroImage*, 41(4):1324–1335, July 2008. PMID: 18467131.

[53] J. Ramirez, J.M. Gorriz, D. Salas-Gonzalez, A. Romero, M. Lopez, I. Alvarez, and M. Gomez-Rio. Computer-aided diagnosis of alzheimer's type dementia combining support vector machines and discriminant set of features. *Information Sciences*, In Press, Corrected Proof:–, 2009.

[54] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. *Learning internal representations by error propagation*, pages 318–362. MIT Press, 1986.

[55] A. Savio, M.T. García-Sebastián, M. Graña, and J. Villanúa. Results of an adaboost approach on alzheimer's disease detection on mri. In J.R. Alvarez F. dela Paz F.J. Tolede (Eds.) J. Mira, J. M. Ferrández, editor, *Bioinspired applications in Artificial and Natural Computation, Proc. IWINAC 2009, Part II,*, volume 5602 of *LNCS*. Springer-Verlag, 2009.

[56] A. Savio, M.T. García-Sebastián, C. Hernandez, M. Graña, and J. Villanúa. Classification results of artificial neural networks for alzheimer's disease detection. In Hujun Yin (eds) Emilio Corchado, editor, *Intelligent Data Engineering and Automated Learning- IDEAL 2009*, volume 5788 of *LNCS*, pages 641–648. Springer-Verlag, 2009.

[57] R. I. Scahill, J. M. Schott, J. M. Stevens, M. N. Rossor, and N. C. Fox. Mapping the evolution of regional atrophy in alzheimer's disease: Unbiased analysis of fluid-registered serial MRI. *Proceedings of the National Academy of Sciences*, 99(7):4703, 2002.

[58] R.E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, 37(3):297–336, December 1999.

[59] A. Sistiaga, I. Urreta, M. Jodar, A. M. Cobo, J. Emparanza, D. Otaegui, J. J. Poza, J. J. Merino, H. Imaz, J. F. Marti-Masso, and A. Lopez de Munain. Cognitive/Personality pattern and triplet expansion size in adult myotonic dystrophy type 1 (DM1): CTG repeats, cognition and personality in DM1. *Psychological Medicine*, First View:1–9, 2009.

[60] P. Somervuo and T. Kohonen. Self-Organizing maps and learning vector quantization for feature sequences. *Neural Process. Lett.*, 10(2):151–159, 1999.

[61] D.F. Specht. Probabilistic neural networks. *Neural Netw.*, 3(1):109–118, 1990.

[62] D. Tao, X. Tang, X. Li, and X. Wu. Asymmetric bagging and random subspace for Support Vector Machines-based relevance feedback in image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(7):1088–1099, 2006.

[63] V. Vapnik. *Statistical learning theory*. Wiley-Interscience, 1998.

[64] P. Vemuri, J.L. Gunter, M.L. Senjem, J.L. Whitwell, K. Kantarci, D.S. Knopman, B.F. Boeve, R.C. Petersen, and C.R. Jack. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. *NeuroImage*, 39(3):1186–1197, Feb 2008.

[65] J. Zhang, B. Yan, X. Huang, P. Yang, and C. Huang. The diagnosis of Alzheimer's disease based on Voxel-Based Morphometry and Support Vector Machine. In *Proceedings of the 2008 Fourth International Conference on Natural Computation (ICNC)*, pages 197–201, Washington, DC, USA, 2008. IEEE Computer Society.