# 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], A. López de Munain[4,5], J. Villanúa[2]

[1]Computational Intelligence Group, UPV-EHU

[2]Osatek, Hospital Donostia

[3]Neuroscience Department UPV-EHU

[4]Neurology Department, Donostia Hospital

[5]Centro de Investigación Biomédica en Red sobre Enfermedades
Neurodegenerativas (CIBERNED)

# Outline

# Outline

# Alzheimer's Disease (AD)

- Neurodegenerative disorder and one of the most common cause of dementia in old people.

- Still incurable and terminal.

- Although noninvasive approaches for antemortem diagnosis of AD are under development, definitive diagnosis requires a postmortem study of the brain tissue.

- T1 weighted MRI scans (sMRI) promises to aid diagnosis and treatment monitoring of AD, offering the potential for easily obtainable surrogate markers of diagnostic status and disease progression.

## Objective

- Detection of patients with very mild to mild Alzheimer's disease.

# Our approach

Using sMRI and standard classifiers:

- Feature extraction based on Voxel-based Morphometry (VBM) analysis
- Backpropagation (BP)
- Radial Basis Function Networks (RBFN)
- Learning Vector Quantization Networks (LVQ)
- Probabilistic Neural Network (PNN)

# Differential features of our work

- This issue has been addressed in many other works.

The differences here are:

- Freely available database with good quality images and well-documented.
- The number of subjects selected for this study is relatively high.

# Voxel-based Morphometry (VBM)

- Morphometry analyses allow a measurement of structural differences within or across groups throughout the entire brain.
- VBM measures differences in local concentrations of brain tissue, through a voxel-wise comparison of multiple brain images.
- The most popular brain morphometry analysis.

# VBM Preprocessing Pipeline



Voxel-Based Morphometry Process Pipeline
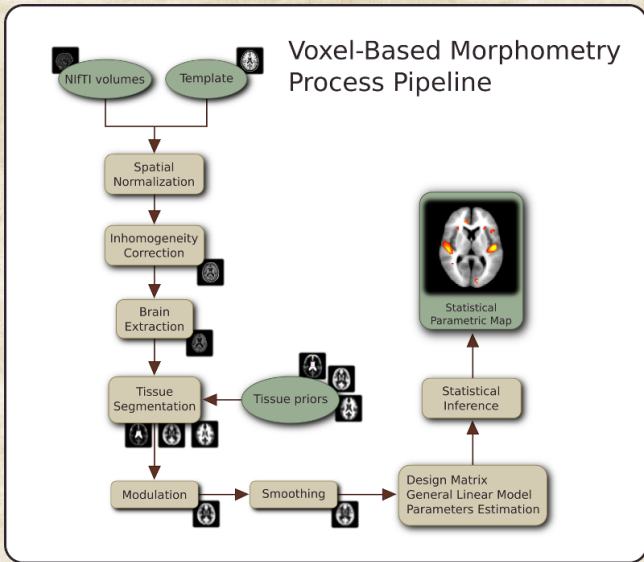
# VBM and the General Linear Model (GLM)

- After preprocessing we fit the data to a linear statistical model, each grey matter voxel independently.

- Use the estimated model parameter values to look for a specific effect we are interested in:

  - Identifying and characterizing structural differences in GM among populations.

# VBM and GLM

- The GLM equation expresses the observed response variable in terms of a linear combination of regressors.

$$Y = X\beta + \varepsilon$$

- Y: observation vector (Mx1)
- X: design matrix (MxL). Each column corresponds to an effect that the user has built into the experiment or that may confound the results.
- $\beta$: regressor or covariate vector (Lx1). Unknown parameters
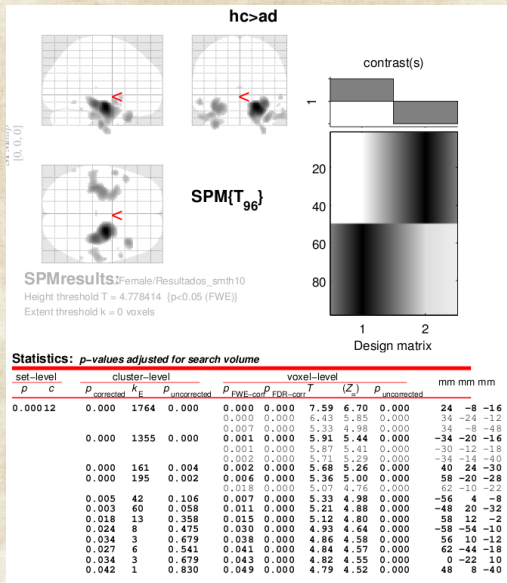- $\varepsilon$: vector of error terms (Mx1)

# VBM (Statistical Inference)

- On the results of GLM a t-test is computed at each voxel.

- The t-test values constitute a Statistical Parametric Map (SPM).

- The decision threshold for the test is set using Random Field Theory to account for spatial dependencies.

# Subjects

- The set of subjects used consists in 98 women selected from the Open Access Series of Imaging Studies (OASIS) database

|  | Very mild to mild AD | Normal |
|---|---|---|
| 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) |

- We find many subjects with high MMSE and low CDR.

# Feature Extraction

- The clusters (regions) detected as result of VBM were used as a mask on the grey matter (GM) segmentation images to select the potentially most discriminant voxels.

- Two sets of featues were extracted:
  1. Mean and standard deviation of grey matter probability voxels within each cluster (MSD)
  2. All grey matter voxels within clusters in a vector (VV)

# General Experiment Pipeline
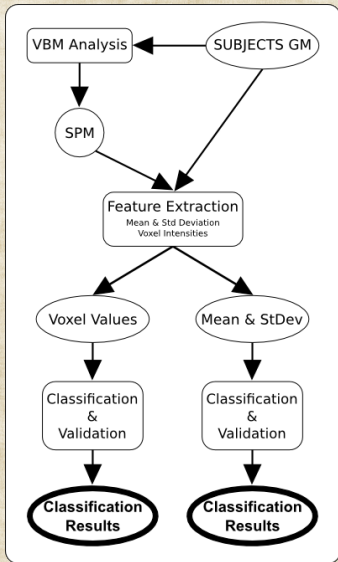
# Classifiers

- Many supervised classification algorithms were used:
    - Multi-layer perceptron using Backward propagation of errors or Backpropagation (MLP-BP)
    - Radial Basis Function Networks (RBFN)
    - Probabilistic Neural Networks (PNN)
    - Learning Vector Quantization (LVQ)
    - Support Vector Machines (SVM) linear and RBF kernel
    - Classification using AdaBoost and Support Vector Machines (SVM) as weaklearners.
    - Combination of multiple SVM with and without AdaBoost

# Multi-layer Perceptron with Backpropagation (ML-BP)

- A non-linear generalization of the squared error gradient descent learning rule for updating the weights of the artificial neurons in a single-layer perceptron.

- We have used the resilient backpropagation, which uses only the derivative sign to perform the weight updating.

# Radial Basis Function Networks (RBFN)

- Are ANN that use radial basis functions as activation functions.

- RBFNs consist of a two layer neural network, where each hidden unit implements a radial activated function.

- Training consists of the unsupervised training of the hidden units followed by the supervised training of the output units' weights.

# Probabilistic Neural Networks (PNN)

- A PNN is a special type of ANN that uses a kernel-based approximation to form an estimate of the probability density function of categories in a classification problem.

- The distance is computed from the point being evaluated to each of the other points and a RBF is applied to the distance to compute the weight for each point.

- The most common RBF function used is the Gaussian function, where a spread value must be set.

- We performed a search for the best sigma value in the range (0, 1).

# Learning Vector Quantization (LVQ)

- LVQ provides a method for training <span style="color:red">competitive layers in a supervised manner</span>.

- The system is composed of an unsupervisedly trained competitive layer which performs a partitioning of the input space.

- The supervisedly trained output layer provides the labeling of the input data according to its belonging to an input region (crisp clustering) or to its degree of membership (soft clustering).

- The basic versions proposed by Kohonen are known as the LVQ1 and LVQ2.

# Adaptive Boosting (AdaBoost)

- Meta-algorithm for machine learning that can be used in conjunction with many other learning algorithms to improve their performance.

- Adaptive in the sense that subsequent classifiers built are adjusted in favor of those instances misclassified by previous classifiers.

- Sensitive to noisy data and outliers. Otherwise, less susceptible to over-fitting.

# Diverse AdaBoost

**Algorithm 2** Diverse AdaBoostSVM

1. **Input:** a set of training samples with labels $\{(x_1, y_1), \ldots, (x_N, y_N)\}$; the initial $\sigma$, $\sigma_{ini}$; the minimal $\sigma$, $\sigma_{min}$; the step of $\sigma$, $\sigma_{step}$; 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: $\gamma = (2\sigma^2)^{-1}$.

   (b) Use $\sigma$ to train a component classifier $h_t$ on the weighted training set.

   (c) Calculate the training error of $h_t$: $\epsilon_t = \sum_{i=1}^{N} w_i^t$, $y_i \neq h_t(x_i)$.

   (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 $\sigma$ by $\sigma_{step}$ 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:** $f(x) = sign(\sum_{k=1}^{C} \alpha_k h_k(x))$.

# Combination of many SVMs and Diverse-AdaBoost

- Three different AdaBoost classifiers were created:

  1. Independent SVM classifiers for each VBM detected cluster and the combination of their responses by a simple majority voting. (Indep-linear-SVM and Indep-rbf-SVM)

  2. Independent SVM classifiers for each VBM detected cluster and the combination of their responses by taking into account a given weight based on their training errors. (linear-AB-SVM and rbf-AB-SVM)

  3. Using all the voxels within the SPM; we trained SVM with different RBF kernel sigma values, weighted them taking into account its training error and then, selected the classifiers with highest diversity based on its weight. (rbf-DAB-SVM)

Classification & Validation (10-Fold Cross-Validation)
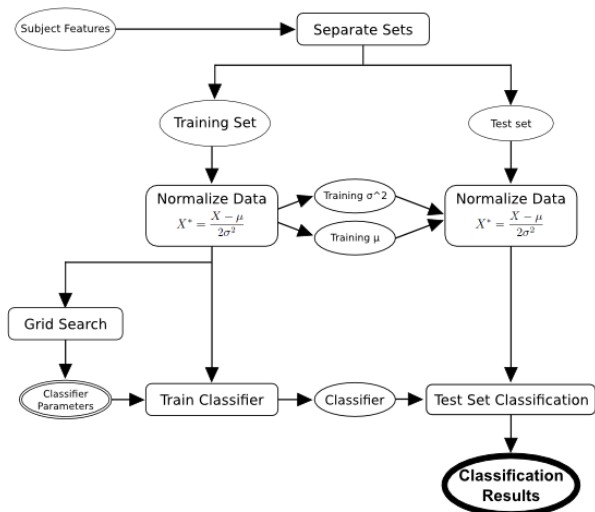
# Results

- All the results were extracted from the VBM detected clusters.

- We performed 10 times a 10-fold cross-validation for each experiment.

- For each experiment we show:
  - Size of feature vector
  - Classification accuracy
  - Sensitivity (AD patients correctly classified)
  - Specificity (controls correctly classified)

# MSD Results

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

# VV Results

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

# Outline

# 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, conduction defects),
    - the endocrine system (hyperinsulinemia)
    - and eyes (cataract).

- Its prevalence is significantly higher in Gipuzkoa (North of Spain), reaching 300 cases per million inhabitants, than the world average (69 to 90 cases per million).

## Subjects

| | | 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%) |

- SVM with RBF kernel on different data sets varying the FWHM size of the smoothing filter in the VBM process.

# MD1 MSD Results

| FWHM(mm) | Size-Thr | #Features | Accuracy | Sensitivity | Specific |
|----------|----------|-----------|----------|-------------|----------|
| 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 |

# MD1 VV Results

| 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 |

Outline

# Conclusions

- We performed feature extraction processes based on VBM analysis to classify MRI volumes of AD patients and normal subjects.

- 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 2 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.

- As we don't have post-mortem confirmation of AD subjects, the very mild demented subjects could be false positives. Post-mortem confirmation data of AD diagnosed subjects could improve the results.

# Further work

- Using other morphometry methods such as Deformation-based and Tensor-based morphometry.

- 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. 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.

Thank you for your attention.