ELM for feature selection and classification of cocaine dependent patients on structural MRI data

M. Termenón[1], M. Graña[1], A. Barrós-Loscertales[2], C. Ávila[2]

[1]Computational Intelligence Group
Basque Country University, UPV-EHU

[2]Dpto. Psicología Básica, Clínica y Psicobiología
Universitat Jaume I, Castellón de la Plana, Spain

Outline

Introduction

- Application of Machine Learning (ML) techniques for the computer aided diagnosis (CAD) of cocaine adicted subjects.
- Aim:
  - To obtain discriminant regions in the brain of structural (T1) Magnetic Resonance Imaging (MRI) data.
  - To train and test classifiers able to discriminate cocaine dependent patients from healthy subjects.

Cocaine Adiction

- Cocaine is one of the most illegal consumed drugs.
- Its chronic abuse may cause: ischemic, hemorrhagic strokes, cerebral infarcts, depression and neuropsychological abnormalities.
- Selected regions in the brains of cocaine users show functional, neurochemical and structural abnormalities.
- These regions can be used to identify the differences between the brains of cocaine users and nonusers and then, to select an adecuate pharmacotherapy to treat this disorder.

T1 Magnetic Resonace Imaging

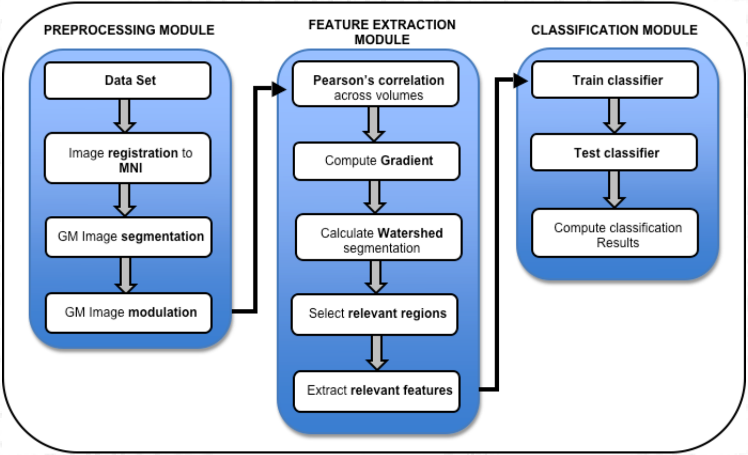
- MRI is a medical imaging technique used in radiology to visualize detailed internal structures.
- It provides good contrast between the different soft tissues of the body.

Database

- 70 cocaine-dependent patients (34.41 $\pm$ 6.62).
- 54 matched controls (33.38 $\pm$ 7,87).
- Exclusion criteria: neurological illness, prior head trauma, positive HIV status, diabetes, Hepatitis C or other medical illness and psychiatric disorders.
- Groups were matched on the basis of age and level of education.
- Patients were recruited from the Addiction Treatment Service of San Agustín in Castellón, Spain.

Outline

Pipeline


PREPROCESSING MODULE
Data Set
Image registration to MNI
GM Image segmentation
GM Image modulation
FEATURE EXTRACTION MODULE
Pearson's correlation across volumes
Compute Gradient
Calculate Watershed segmentation
Select relevant regions
Extract relevant features
CLASSIFICATION MODULE
Train classifier
Test classifier
Compute classification Results

Outline

1. Introduction
2. Methods
   - Preprocessing Module
   - Feature Extraction Module
3. Validation Experiments
4. Classification Results and Features Location
5. Summary and Conclusions

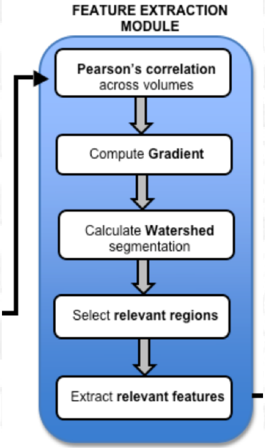
Preprocessing Module

- Appropriate data preprocessing, ensuring anatomical correspondence of voxels intersubjects, is of paramount importance.
- We perform the preprocessing on Statistical Parametric Mapping (SPM) software running on Matlab.
- Several steps:
  - Reorientation.
  - Tissue segmentation.
  - Bias correction
  - Spatial normalization to MNI152 template.
    - Linear registration step
    - Non-linear shape registration
  - GM images modulation to restore tissue volume changes.

Outline

1. Introduction
2. Methods
   - Preprocessing Module
   - Feature Extraction Module
3. Validation Experiments
4. Classification Results and Features Location
5. Summary and Conclusions

Feature Extraction Module



Pearson's Correlation across volumes

- Voxel-wise Pearson's correlation with the indicator variable specifying the subject class label (0 healthy control, 1 patient).
- At the j-th voxel site is computed as follows:

$$r_{\mathbf{v}_j,\mathbf{y}} = \frac{n\sum_i v_{ij}y_i - \sum_i v_{ij}\sum_i y_i}{\sqrt{n\sum_i v_{ij}^2 - (\sum_i v_{ij})^2}\sqrt{n\sum_i y_i^2 - (\sum_i y_i)^2}}, \qquad (1)$$

where v_{ij} is the value of the j-th voxel site in the i-th MRI volume in the dataset and y_i is the class label value of that i-th volume.

- Computing this correlation coefficient for all voxels, we obtain the whole brain volume of correlation values (VCV).

Watershed segmentation

- Watershed transformation is computed on the gradient magnitude of the original image.
- Before the gradient, we apply a Gaussian smoothing step trying to reduce the number of ROIs given by the watershed.
- We compute the 3D spatial gradient of the previously computed VCV as:

$$\nabla F = \frac{\partial F}{\partial x}\hat{i} + \frac{\partial F}{\partial y}\hat{j} + \frac{\partial F}{\partial z}\hat{k}, \tag{2}$$

where each partial derivative is computed by differences along its corresponding axis direction.

Watershed segmentation

- Watershed transformation is a mathematical morphology technique for image segmentation.
- It is computed on the VCV volumes corresponding to the GM segmentation mask.

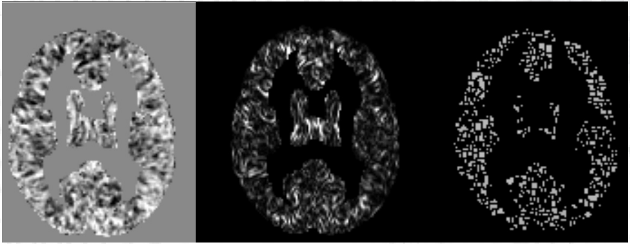


Figure: Left, the VCV of the GM computed on a training set; middle, gradient of the VCV; right, ROI watershed segmentation.

Region Selection Process

- Two different approaches:
  1. Wrapper approach using ELMs as classifiers to determine regions relevance.
  2. Application of different percentiles of the correlation coefficients distribution to select most correlated regions.

1. ELM wrapper ROI selection process

- We sort the ROI obtained from watershed segmentation in descending order of ROI's mean correlation values.
- We start training the classifier with the first ROI (most correlated one) and computing its $F - score$ measure on the test data.
- We add the second region to the data, training again the ELM and computing the new $F - score$.
- If the $F - score$ value increases, the ROI is added definitively to the feature vector, otherwise it is discarded. We repeat this process with all the regions.

2. Different percentiles

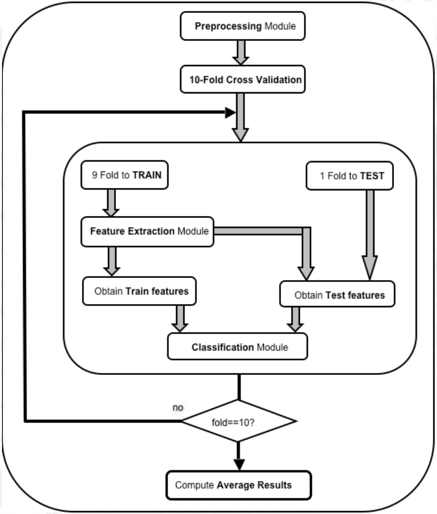
- We compute the empirical distribution of VCV's ROI average absolute values.
- We apply six different percentiles (from 90.00% to 99.95%) on this distribution to select the most discriminant regions.

Feature extraction process

- Three different procedures to extract the feature vector values:
  1. Collecting the intensity values of all the voxels that compose each region (feature vectors size = sum of the sizes of the ROIs).
  2. Computing the mean value of the voxel intensities in the ROI (feature vectors size = number of selected ROIs).
  3. Computing the median value of the voxel intensities in the ROI (feature vectors size = number of selected ROIs).

Outline

Avoiding circularity



Performance measures

$F-score$ (also called, F_1 $score$ or $F-measure$) is a measure of a test's accuracy. It is defined as the harmonic mean between $precision$ and $recall$:

$$F-score = 2 \cdot \frac{precision \cdot recall}{precision + recall},$$

where:

- $precision$ is defined as the positive predictive value, $precision = \frac{TP}{TP+FP}$,
- $recall$ is referred as the true positive rate, $recall = \frac{TP}{TP+FN}$.
- Two other scores, $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$ and $Specificity = \frac{TN}{TN+FP}$;
- where TP, FP, TN, FN are true positives, false positives, true negatives and false negatives, respectively.

ROI selection

We have used the mean and the median intensity values of the ROIs for this process.

1. ELM based region selection process:

- Standard ELM algorithm, with two different number of hidden nodes (100 and 1000).
- For each ROI addition step, we split the training set into training and validation sets, performing a 10 fold cv, repeated 10 times.
- We compute the mean $F-score$ and that is the value we use to decide if that region will be included as discriminant region.

#Nodes	F-score (%)	#regions
100	58 ± 1	14 ± 3
1000	74 ± 5	22 ± 8

Table: Region selection F-score.

ROI selection

2. **Percentile based selection:**
   - We have applied 6 different percentiles on the VCV empirical distribution to select the most correlated regions.
   - The number of selected ROIs grows quickly so that tuning of this approach seems to be more tricky.

	Percentil					
	99.95%	99.90%	99.50%	99.00%	95.00%	90.00%
#regions	6 ± 0	12 ± 1	62 ± 1	124 ± 1	622 ± 2	1245 ± 5

Table: Average number of regions depending on the applied percentile.

Classification

- Standard ELM was trained with different number of hidden layer nodes (100, 1000, 2000, 3000) and sigmoid function activation function.
- We repeat each cross-validation process 50 times, so reported results are the average values of the performance measures.
- We report comparison results with other classifiers, OP-ELM, linear kernel SVM and 1-NN under the same cross-validation framework.

Outline

Results without ROI selection

- As a baseline result, we consider the feature vectors composed of the mean and median intensity values of all the brain watershed ROIs.

	Mean intensity per ROI				Median intensity per ROI			
#Nodes	**Acc**	**Recall**	**Spec**	**F**	**Acc**	**Recall**	**Spec**	**F**
100	54±4	54±7	54±6	53±5	51±5	50±7	51±6	50±6
1000	63±4	63±6	64±6	63±4	64±4	63±5	66±6	63±4
2000	70±3	70±5	71±6	70±6	69±4	68±5	70±5	68±4
3000	74±4	72±5	76±5	73±4	75±3	73±4	76±5	74±3

Table: Standar ELM classification results using mean and median intensity values of all the watershed ROIs.

Results on the ROIs selected by wrapper ELM

	100				1000			
#Nodes	**Acc**	**Recall**	**Spec**	**F**	**Acc**	**Recall**	**Spec**	**F**
100	49±4	49±7	49±6	48±5	54±4	55±6	53±6	54±4
1000	47±3	48±3	47±4	47±3	62±3	68±3	55±5	64±2
2000	47±2	48±2	47±3	48±2	63±3	70±3	55±4	65±2
3000	48±2	49±2	47±3	48± 2	64±2	71±2	57±3	66±2

Table: Standard ELM Classification results on features extracted from the wrapper ELM selected ROIs.

Results on the ROIs selected as percentiles

- Feature vectors given by the intensity values of each voxel of each relevant region:

	99.95%				99.90%			
Nodes	**Acc**	**Recall**	**Spec**	**F**	**Acc**	**Recall**	**Spec**	**F**
100	54±4	54±6	54±6	54±4	56±4	55±6	56±5	55±5
1000	57±2	60±3	54±3	58±2	60±2	53±3	66±3	56±3
2000	57±2	60±2	55±2	58±2	60±2	53±3	67±3	57±3
3000	58±1	61±2	55±2	59±1	60±2	53±3	68±3	57±3
	99.50%				99.00%			
Nodes	**Acc**	**Recall**	**Spec**	**F**	**Acc**	**Recall**	**Spec**	**F**
100	64±5	65±7	63± 6	64±5	61±5	62±6	60±7	61±5
1000	79±2	78±4	79±3	78±2	80±3	80±3	81±4	80±3
2000	80±2	81±3	79±2	80±2	82±2	82±3	82±3	82±2
3000	80±2	81±3	80±2	80±2	83±2	83±3	83±3	83±2
	95.00%				90.00%			
Nodes	**Acc**	**Recall**	**Spec**	**F**	**Acc**	**Recall**	**Spec**	**F**
100	57±4	59±6	56±6	57±4	57±4	58±7	57±6	57±5
1000	84±3	84±4	84±4	84±3	83±3	82±4	84±5	82±3
2000	89±2	89±4	89±3	89±2	87±2	85±3	89±3	87±2
3000	**91±2**	**90±3**	**91±2**	**90±2**	89±2	88±3	**91±3**	89±2

Results on the ROIs selected as percentiles

- Feature vectors given by the mean or median intensity values of each ROI. (we only show results for the last 2 percentiles).

95.00%	Mean				Median			
Nodes	**Acc**	**Recall**	**Spec**	**F**	**Acc**	**Recall**	**Spec**	**F**
100	60±4	61±6	59±5	60±4	61±4	62±6	60±6	61±5
1000	84±3	84±4	84±4	84±3	85±3	85±4	86±4	85±3
2000	88±2	87±3	89±2	87±2	89±2	88±3	**90±3**	89±2
3000	89±2	88±3	89±3	89±2	**90±2**	89±2	**92±3**	**90±2**

90.00%	Mean				Median			
Nodes	**Acc**	**Recall**	**Spec**	**F**	**Acc**	**Recall**	**Spec**	**F**
100	59±5	58±6	59±7	58±5	58±4	59±7	58±6	58±5
1000	85±3	84±5	85±3	84±3	85±3	85±4	86±4	85±3
2000	89±2	88±2	**90±3**	89±2	**90±2**	88±3	**92±2**	**90±2**
3000	**92±2**	**91±3**	**92±3**	**92±2**	**92±2**	**90±2**	**94±4**	**92±2**

Comparison of ELM with other classifiers

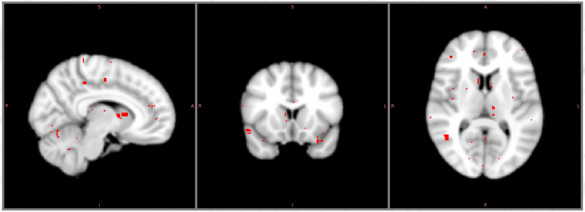
- Best results are achieved by the ELM, comparing well with the much more costly SVM classifiers.

	F-score (regions voxel values)			
Prc(%)	**ELM**	**OP-ELM**	**SVM**	**1NN**
90.00	89± 2	68±5	87±3	77±4
95.00	**90±2**	71±5	86±3	83±4
99.00	83±2	74±6	83±2	78±4
99.50	80±2	74±6	76±2	79±3
99.90	56±3	73±4	73±2	67±3
99.95	59±1	74±3	73±3	61±7

	F-score (regions mean value)			
Prc(%)	**ELM**	**OP-ELM**	**SVM**	**1NN**
90.00	**92± 2**	59±5	**91±2**	80±6
95.00	89±2	70±6	**90±2**	76±5
99.00	77±1	75±6	78±4	78±4
99.50	75±1	77±6	72±3	72±4

Location in the brain of selected ROIs

- Location of selected ROIs for feature extraction for percentile 90.00% of the VCV empirical distribution.
- Selected ROIs are located in several regions in the brain as striatum, thalamus, parahippocampal gyrus, cingulate gyrus, superior frontal gyrus and orbitofrontal cortex, among them.



Outline

Summary

- We present a procedure to discriminate patients with cocaine addiction and healthy subjects using structural MRI brain images.
- Computational pipeline involves:
  - Volume spatial normalization,
  - Computation of Pearson's correlation across volumes giving the VCV,
  - Watershed segmentation of the VCV,
  - ROI selection for feature extraction. Two methods:
    1. a wrapper approach based on ELM.
    2. applying different percentiles on the empirical distribution of the VCV.

Summary

- Using the intensity values of the voxels of all the relevant regions, we reached an accuracy and F-score measures higher than 90%.
- Using mean and median values, we achieve even betters results around 92%.
- When the number of regions is small (percentiles 99.90% and 99.95%), basic ELM results are not good enough, but OP-ELM obtains similar results than linear SVM.
- As the number of regions increases, ELM improves its performance even outperforming the rest of algorithms we are comparing with.
- Features location are related to findings in the literature about cocaine addiction, validating this approach.

Further work

- We would like to focus on:
  - regions selection process using ELMs.
  - different ways to compute the gradient before applying watershed segmentation.
- It would be also interesting to test this procedure with different neurodegenerative diseases and also with other type of MRI images as diffusion tensor MRI or functional MRI.

Thanks

- Contact:
  - Prof. Manuel Graña
  - Computational Intelligence Group
  - Universidad del País Vasco, San Sebsatián, Spain.
  - E-mail: **manuel.grana@ehu.es**
  - **http://www.ehu.es/computationalintelligence**