# Finding Coherent Regions in PET Images for the Diagnosis of Alzheimer's Disease

Helena Aidos, João Duarte and Ana Fred
*Instituto de Telecomunicações, Instituto Superior Técnico, Lisbon, Portugal*

Keywords:     Support Vector Machines, ROI, Feature Extraction, Image Segmentation, Mutual Information.

Abstract:     Alzheimer's disease is a type of dementia that mainly affects elderly people, with unknown causes and no effective treatment up to date. The diagnosis of this disease in an earlier stage is crucial to improve patients' life quality. Current techniques focus on the analysis of neuroimages, like FDG-PET or MRI, to find changes in the brain activity. While high accuracies can be obtained by combining the analysis of several types of neuroimages, they are expensive and not always available for medical analysis. Achieving similar results using only 3-D FDG-PET scans is therefore of huge importance. While directly applying classifiers to the FDG-PET scan voxel intensities can lead to good prediction accuracies, it results in a problem that suffers from the curse of dimensionality. This paper thus proposes a methodology to identify regions of interest by segmenting 3-D FDG-PET scans and extracting features that represent each of those regions of interest, reducing the dimensionality of the space. Experimental results show that the proposed methodology outperforms the one using voxel intensities despite only a small number of features is needed to achieve that result.

## 1 INTRODUCTION

One of the most common forms of dementia is Alzheimer's disease (AD), a progressive brain disorder that has no known cause or cure. It is a disease that slowly leads to memory loss, confusion, impaired judgment, personality changes, disorientation and the inability to communicate. An early detection is very important for an effective treatment, especially in the Mild Cognitive Impairment (MCI) stage, to slow down the progress of the symptoms and to improve patients' life quality. MCI is a condition where a person has mild changes in thinking abilities, but it does not affect daily life activities. People with MCI are more likely to develop AD, even though recent studies suggest that a person with MCI may revert back to normal cognition on its own (Alzheimer's Association, 2013).

Neuroimages allow the identification of brain changes and have been used for automated diagnosis of AD and MCI (Silveira and Marques, 2010; Ye et al., 2012). Due to the high variability of the pattern of brain degeneration in AD and MCI, the analysis of brain images is a very difficult task. Moreover, attempts are being made to develop tools to automatically analyze the images and, consequently, diagnosis AD and MCI conditions (Morgado et al., 2013;

Ramírez et al., 2013).

Most of the techniques developed have focused on analyzing small parts of the brain like hippocampus (Gerardin et al., 2009) or the gray matter volume (Fan et al., 2008). However, these techniques have some limitations by the fact that the brain atrophy affects many and different regions in different stages of the disease. Therefore, researchers are focusing their techniques in analyzing the pattern of the entire brain. However, this leads to the "curse of dimensionality" because a brain image, like the fluorodeoxyglucose positron emission tomography (FDG-PET), contains thousands of voxels (or features). Dimensionality reduction and feature selection techniques are therefore fundamental for achieving high accuracy predictors for the diagnosis of Alzheimer's disease.

Some techniques are based in the segmentation of the brain into Regions of Interest (ROIs), which are associated with atrophy caused by the disease. Then, voxel intensities from each ROI are used as features (Zhang et al., 2011; Mikhno et al., 2012). Some other dimensionality reduction techniques from Machine Learning field (Lopez et al., 2009; Segovia et al., 2012), and feature selection techniques (Bicacro et al., 2012; Chaves et al., 2009) have been applied to the diagnosis of AD.

In this paper, we propose a methodology to au-

tomatically extract features that represent interesting regions of the brain and, consequently, reducing the dimensionality of the space. One of the advantages of this methodology is that brain images, like FDG-PET, do not need to be pre-processed in order to remove the background and the scalp. This is due to the choice of the clustering algorithm, which is a variant of the DB-SCAN (density-based spatial clustering of applications with noise) called XMT-DBSCAN (Tran et al., 2012). Another advantage is that the space we obtain is approximately $100\times$ smaller when compared to the original one, consisting of voxel intensities. This happens because each region (cluster) obtained by the clustering algorithm is represented by a feature, which is a weighted mean of the voxel intensities of that region.

This paper is organized as follows: section 2 explains each step of the proposed methodology and section 3 presents the dataset used in this paper as well the results obtained for the proposed methodology and for the classification task using the voxel intensity. Conclusions are drawn in Section 4.

# 2 THE PROPOSED METHODOLOGY

In order to analyze the FDG-PET scans for each task: AD versus CN (Cognitive Normal), MCI versus CN and AD versus MCI, we propose the methodology shown in Figure 1. We start by segmenting each 3-D image (a FDG-PET scan from a subject), followed by a construction of a probability matrix indicating the degree of belonging of each voxel to a region found by the clustering/segmentation algorithm. Then, we perform a feature extraction step using the voxel intensities and the probability matrix, obtaining a feature space representation for each problem. Finally, feature selection is applied and the subjects are classified, using support vector machines.

## 2.1 Step 1: Image Segmentation

Over the years, several 3-D segmentation methods have been developed such as region growing, watershed, among others (Arbeláez et al., 2011; Tripathi et al., 2012); watershed algorithm (Beucher and Lantuejoul, 1979) is the most widely used. However, watershed tends to over-segment the 3-D images when the data is dense and non-homogeneous, or generate under-segmentation results in the case of dense regions with irregular shapes of objects. Since our FDG-PET scans are noisy images that have regions with different sizes, densities and irregular shapes,

we propose to use a version of the DBSCAN algorithm, namely the XMT-DBSCAN (Tran et al., 2012), to segment the 3-D images.

XMT-DBSCAN is an extension of the original DBSCAN but has a few differences. Firstly, the local density of a voxel (a pixel in DBSCAN) is computed in the sub-window with size $ws = (2w+1) \times (2w+1) \times (2w+1)$ centered in the voxel, instead of the ball with radius $eps$. In our methodology, the local density is computed as

$$density(v_{ijk}) = \frac{\sum_{all-elements} I^w_{v_{ijk}} \odot K^w}{a_k}, \quad (1)$$

where $\odot$ is the element-wise product of two equally sized data cubes, $K^w$ is a cubic Gaussian kernel with standard deviation equal to $ws/(4\sqrt{2\log(2)})$, $I^w_{v_{ijk}}$ is the sub-window from the intensity image, and $a_k$ is the number of non-zero values in $K^w$.

The identification of the voxels as core points, border points and noise is similar to the original DBSCAN. Another modification to the original DB-SCAN is in the definition of density-reachable chain (Ester et al., 1996), which is modified to contain only core voxels. This means that labeling the border points is made in a post-processing step, at the end of the algorithm, when all core points are identified.

## 2.2 Step 2: Coherence Matrix

After segmenting each 3-D image, we obtain a partition into regions (clusters) and we need to find some consensual information for each population (AD, CN or MCI). In that sense, we construct a block coherence matrix $\mathbf{C}$, with as many blocks as the squared number of subjects of a population. The idea is to perform a pairwise comparison between the partitions obtained by XMT-DBSCAN for each subject of a population. Therefore,

$$\mathbf{C}(\mu(l,i),\mu(p,j)) = \frac{|C^l_i \cap C^p_j|}{\sqrt{|C^l_i| \cdot |C^p_j|}}, \quad (2)$$

where $\mu(l,i)$ is the indexation function for the coherence matrix $\mathbf{C}$, $|C^l_i \cap C^p_j|$ is the number of voxels belonging to both $C^l_i$ and $C^p_j$, with $C^l_i$ the region/cluster $i$ from subject $l$ and $C^p_j$ the region/cluster $j$ from subject $p$. The indexation function is given by

$$\mu(l,i) = i + \sum_{j=1}^{l-1} m_j,$$

with $m_j$ the number of clusters in the partition of subject $j$, i.e., $\mu(l,i)$ gives the index corresponding to cluster $i$ of subject $l$, where each partition of a subject has $m_j$ clusters. Figure 2 shows an example of a
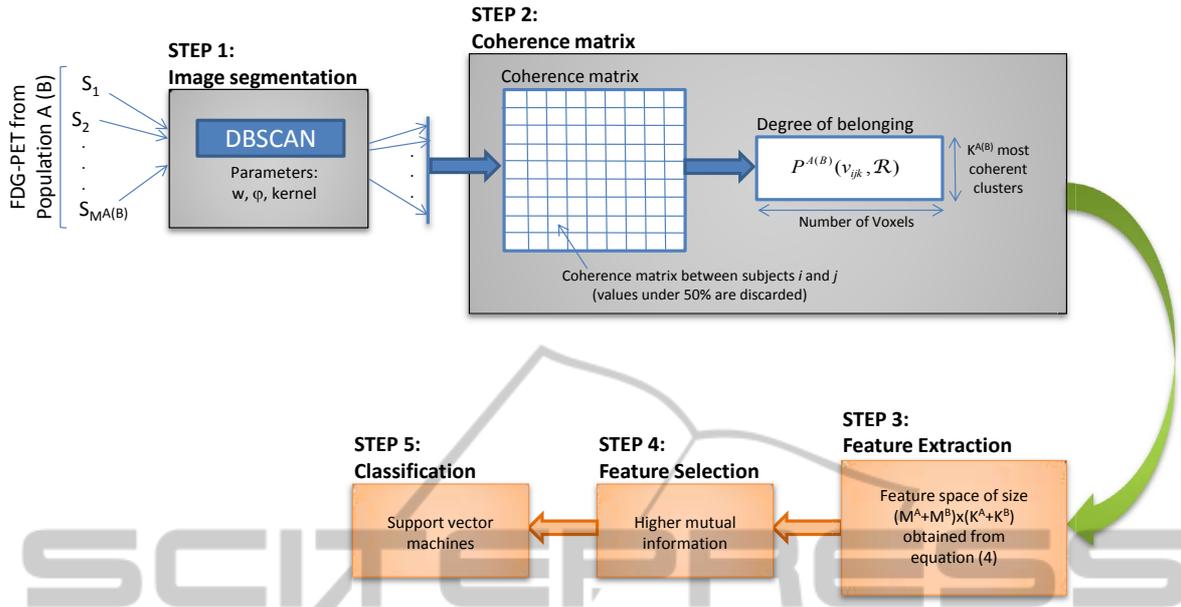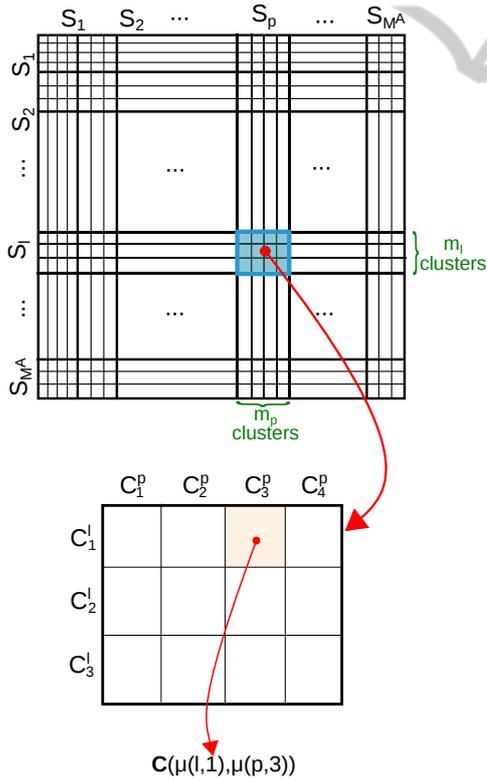
Figure 1: The proposed methodology.



Figure 2: Coherence matrix.

coherence matrix.

The matrix $\mathbf{C}$ shows the degree of overlapping of each pair of clusters. Since we want a region that is common in most of the subjects, we consider that val-

ues under 50% of overlapping are discarded.

We start by searching the most coherent cluster in matrix $\mathbf{C}$ and obtain a region $\mathcal{R}$ corresponding to the union of all clusters with an overlapping over 50% to the most coherent cluster found. Inside the region $\mathcal{R}$, we compute the probability (for a certain population) of each voxel belong to $\mathcal{R}$ as

$$P^A(v_{ijk}, \mathcal{R}) = \frac{\sum_{C_k \in \mathcal{R}} \mathbf{1}_{\{v_{ijk} \in C_k\}}}{\sum_{C_k \in \mathcal{R}} \mathbf{1}_{\{C_k \in \mathcal{R}\}}}, \qquad (3)$$

where $C_k$ is the $k$-th cluster of region $\mathcal{R}$, $\mathbf{1}_{\{v_{ijk} \in C_k\}}$ is 1 if $v_{ijk} \in C_k$, and 0 otherwise; $v_{ijk}$ is a voxel in the 3-D image and $A \in \{AD, CN, MCI\}$. The numerator of the previous equation is a count of the number of clusters in $\mathcal{R}$ where the voxel belongs, and the denominator is just the number of clusters in $\mathcal{R}$. This process is repeated until no coherent clusters are left in matrix $\mathbf{C}$. Therefore, $P^A$ is $K \times N$ matrix, with $K$ the number of regions and $N$ the number of voxels in the 3-D image.

## 2.3 Step 3: Feature Extraction

So far we have found regions containing relevant information for each population. Now we want to discriminate AD vs CN, CN vs MCI and AD vs MCI. This means that we will construct a feature space for each of these problems using the voxels intensities from two populations and the regions found in step 2 corresponding to the same two populations.

Consider that $M^A$ is the number of subjects from population $A$ and $M^B$ the number of subjects from population $B$. Also, $K^A$ and $K^B$ are the number of
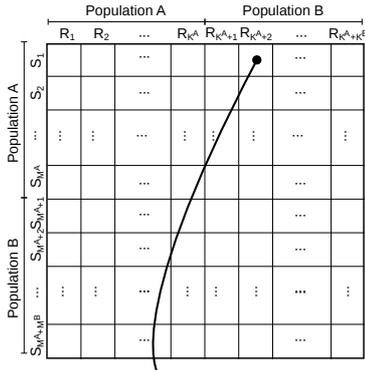
$$\mathbf{F}(\alpha(1, \text{population A}), \beta(2, \text{population B}))$$

Figure 3: New feature space representation obtained by feature extraction.

regions found in step 2 for population $A$ and $B$, respectively. We want to construct a feature space $\mathbf{F}$ with $M^A + M^B$ samples and dimension $K^A + K^B$, in the following way

$$\mathbf{F}(\alpha(p,r), \beta(q,s)) = \frac{\sum_{v_{ijk}} I(v_{ijk} \in S_p^r) \cdot P^s(v_{ijk}, \mathcal{R}_q)}{\sum_{v_{ijk}} P^s(v_{ijk}, \mathcal{R}_q)}, \quad (4)$$

with $r,s \in \{\text{population } A, \text{population } B\}$. $I(v_{ijk} \in S_p^r)$ is the intensity of voxel $v_{ijk}$ from subject $p$ in population $r$ and $P^s(v_{ijk}, \mathcal{R}_q)$ is the probability that voxel $v_{ijk}$ belongs to region $\mathcal{R}_q$ in population $s$. $\alpha(p,r)$ and $\beta(q,s)$ are indexation functions given by

$$\alpha(p,r) = \begin{cases} p & \text{if } r = \text{ population A} \\ M^A + p & \text{if } r = \text{ population B} \end{cases}$$

and

$$\beta(q,s) = \begin{cases} q & \text{if } s = \text{ population A} \\ K^A + q & \text{if } s = \text{ population B} \end{cases}$$

respectively. $\alpha(p,r)$ is the indexation for subjects and $\beta(q,s)$ the indexation for regions, as illustrated in figure 3.

Equation (4) is equivalent to compute a weighted mean of the intensity of a subject, where some voxels contribute more than others, obtaining a feature space for each classification task.

## 2.4 Step 4: Feature Selection

Typically, the number of voxels in a FDG-PET image is very high and some of those voxels are unimportant for the task in hand. So, it is very important to reduce the dimensionality of the space through feature selection. We use mutual information (MI) to rank the features and choose the ones with higher value.

Table 1: Clinical and demographic characteristics of each group. Age and MMSE (Mini Mental State Exam) values are means ($\pm$ standard deviations).

| Attributes | AD | MCI | CN |
|---|---|---|---|
| Number of subjects | 59 | 59 | 59 |
| Age | 78.26 | 77.71 | 77.38 |
| | ($\pm$6.62) | ($\pm$6.88) | ($\pm$4.87) |
| Sex (% of males) | 57.63 | 67.80 | 64.41 |
| MMSE | 19.60 | 25.68 | 29.20 |
| | ($\pm$5.06) | ($\pm$2.97) | ($\pm$0.92) |

Consider that $x_i$ is the $i$-th element of a vector representing a feature $\mathbf{x}$, and $y$ a target value or label. The MI between the random variable $x_i$ and $y$ is given by

$$MI(i) = \sum_{x_i} \sum_{y} P(x_i, y) \log \frac{P(x_i, y)}{P(x_i)P(y)}. \quad (5)$$

The probability density functions for MI were estimated through the use of histograms.

## 2.5 Step 5: Classification

After selecting the most relevant features for each of the three diagnostic problems, we classify subjects through the support vector machine (SVM) algorithm with a linear kernel (Cortes and Vapnik, 1995). The SVM algorithm is a popular classifier in several areas, including diagnosis of neurological diseases like Alzheimer.

# 3 EXPERIMENTS

## 3.1 Dataset

In this study, we used FDG-PET images for AD, MCI and CN subjects, retrieved from the ADNI database. The subjects were chosen to obey a certain criteria: the Clinical Dementia Rating (CDR) should be 0.5 or higher for AD patients, 0.5 for MCI patients and 0 for normal controls. This selection results in a dataset composed by 59, 142 and 84 subjects for AD, MCI and CN, respectively. Since our task is classification using the SVM algorithm, we decided to balanced the classes by using a random sub-sampling technique. Thus, 59 subjects from each MCI and CN groups were selected randomly. Table 1 summarizes some clinical and demographic information in each group.

The FDG-PET images have been pre-processed to minimize differences between images: each image was co-registered, averaged, reoriented (the anterior-posterior axis of each subject was parallel to the AC-PC line), normalized in its intensity, and smoothed

to uniform standardized resolution. A more detailed description of the pre-processing is available in the ADNI project webpage[1].

The complete $64 \times 64 \times 30$ FDG-PET images were used, which means that no background or extra-cranial voxels were excluded. We left those voxels because the image segmentation step will automatically discard them and only the relevant voxels will be labeled.

## 3.2 Experimental Setup

The FDG-PET image of the brain of each individual needs to be segmented with XMT-DBSCAN, the segmentation algorithm proposed in the methodology. In section 2, we state that XMT-DBSCAN has two parameters: window size $w$ and $\varphi$ which is a threshold to identify core and border voxels (see (Ester et al., 1996) for more details). We set $w$ to 2 and 3, and $\varphi$ takes values from $\{0.3, 0.5, 0.7\}$. The first part of our experiments consists in the analysis of the influence of these parameters in the results.

In the feature selection step we discretized the probability density functions through histograms with 8 bins and, after ranking the features according to the MI, we choose the ones with higher value. We consider several number of features selected by the MI, according to table 2.

The final step of the proposed methodology consists in classifying subjects using a linear SVM. We set the cost of misclassification in SVM as $\{2^{-16}, 2^{-14}, 2^{-12}, 2^{-10}, 2^{-8}, 2^{-6}, 2^{-4}, 2^{-2}, 2^0, 2^2, 2^4\}$ and performed a $20 \times 10$ nested cross-validation procedure (Varma and Simon, 2006).

We compare the proposed methodology with the one consisting of the voxel intensities, called MI-SVM. In that strategy, we first need to pre-process the FDG-PET images to remove the background and the scalp. Afterwards, steps 4 and 5 of the proposed methodology are applied. The number of selected features used to classify the subjects are shown in table 2.

## 3.3 Results

Firstly, we want to study the influence of the two parameters ($w$ and $\varphi$) of the image segmentation step in the classifier. Figure 4 shows the accuracy of the classifier for the considered parameters values.

In AD vs CN problem we see that the best result is higher than 89% and it is given when we consider a $\varphi = 0.7$ and $w = 3$ in the XMT-DBSCAN algorithm. Also, this best value is obtained with a lower
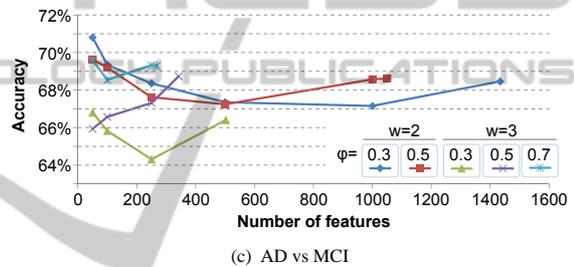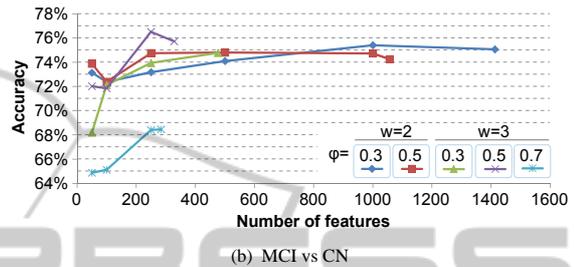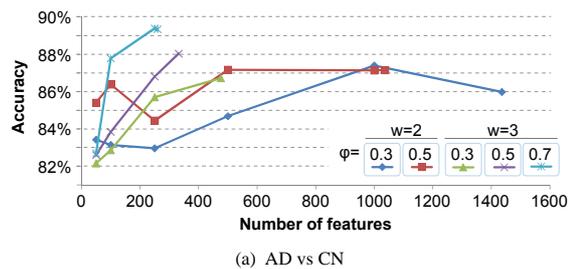
---

[1]http://adni.loni.usc.edu/methods/pet-analysis/pre-processing/



(a) AD vs CN



(b) MCI vs CN



(c) AD vs MCI

Figure 4: Average accuracy of $20 \times 10$ nested cross-validation of the proposed methodology for several different parameters consider in the image segmentation step.

number of features, around 250 features, which corresponds to the all space for those parameters. Moreover, with $w = 2$ and $\varphi = 0.3$ we have the lowest accuracy for different number of features and the maximum is when we have a space with 1000 features with an accuracy of approximately 87%.

In MCI vs CN problem the worst result is for $w = 3$ and $\varphi = 0.7$, opposite of what we see in AD vs CN. Now the best result is higher than 76% and it is given by $w = 2$ and $\varphi = 0.5$, which means that we need a small sub-window to distinguish between MCI subjects and CN subjects. Again, we only need around 250 features for the better accuracy.

In the case of AD vs MCI, the worst results are for $w = 3$ and $\varphi = 0.3$ and it is approximately 64%, but the best result is obtained using only 50 features and a small window and density in the XMT-DBSCAN algorithm ($w = 2$ and $\varphi = 0.3$). For those parameters, we notice that if we increase the number of features, the accuracy decreases.

The two parameters we are discussing affects not only the number of features of the space, but also the

Table 2: Number of features used to tested the feature selection step. The maximum number of features used corresponds to the complete feature space, as stated by columns 2-4, depending on the problem.
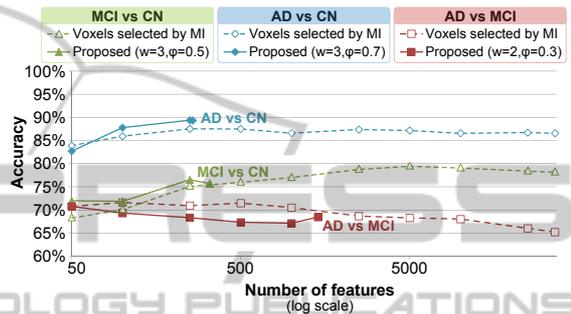
| Parameter | Max. features | | | |
|---|---|---|---|---|
| Space | AD vs CN | MCI vs CN | AD vs MCI | Number of selected features |
| $w = 2, \varphi = 0.3$ | 1436 | 1413 | 1433 | 50, 100, 250, 500, 1000, Max. features |
| $w = 2, \varphi = 0.5$ | 1037 | 1057 | 1050 | 50, 100, 250, 500, 1000, Max. features |
| $w = 3, \varphi = 0.3$ | 476 | 476 | 502 | 50, 100, 250, Max. features |
| $w = 3, \varphi = 0.5$ | 332 | 328 | 342 | 50, 100, 250, Max. features |
| $w = 3, \varphi = 0.7$ | 260 | 284 | 268 | 50, 100, 250, Max. features |
| voxel intensity | | 36209 | | 50, 100, 250, 500, 1000, 2500, 5000, 10000, 25000, Max. features |

accuracy of the classifier. From figure 4, we notice that for distinguish between AD and CN subjects we need to create large regions with high density (intensity). This makes sense, since FDG-PET scans of the brain measures the glucose used, and patients with Alzheimer's disease had a big decrease in brain metabolism of glucose compared to a normal patient. Moreover, if we want to distinguish between MCI patients and CN or AD patients, we need to decrease the size of clusters, which leads to an increase of number of regions/features and look for differences in more specific locations of the brain. This happens because MCI is a transition stage: some MCI patients may convert to Alzheimer others just remain stable over time or even remit.

Figure 5 compares the best result obtained with the proposed methodology for each problem with the methodology using the voxel intensity. For AD vs CN our methodology outperforms MI-SVM with only a few features. Even if we use the all space, MI-SVM is always worst than our methodology. Something similar happens for MCI vs CN, using few features (around 250 features) our methodology outperforms MI-SVM. However, if we increase the number of features until 5000, MI-SVM can predict better than our methodology. In AD vs MCI, our best performance is for the lowest number of features (50 features) and it is comparable to MI-SVM; after that we perform worst than MI-SVM. Notice that, MI-SVM remains almost constant until 1000 features and then the performance decreases. Moreover, our methodology starts to increase after the 1000 features. This may indicate that we need more features to discriminate MCI subjects from AD subjects, which means we need to decrease the size of the regions corresponding to a decreasing in the parameters $w$ and/or $\varphi$.



Figure 5: Average accuracy of $20 \times 10$ nested cross-validation of the best curve from each problem in figure 4 compared to voxels intensities chosen through mutual information and classified using SVM.

with Alzheimer's disease from the ones with mild cognitive impairment and from normal ones. The proposed methodology has several stages: starts with a segmentation of the FDG-PET image, followed by a grouping of clusters to form regions with relevant information. Those regions form a feature space and the most important ones are selected by ranking their mutual information with the target output. Finally a classifier is used to identify the subjects.

For number of features under 100, the proposed methodology outperforms another strategy consisting in ranking the mutual information of features with the target output, where the features are only the voxels intensities. Moreover, by comparing using all the space in both strategies, our methodology outperforms the other strategy, using a small number of features. Another advantage of this methodology is that the complete FDG-PET image was used, since the segmentation algorithm can identify background and extracranial voxels, which means we do not need to pre-process the images to remove those voxels.

## 4 CONCLUSIONS

This paper proposes a methodology to find interesting regions in the brain to efficiently discriminate subjects

## ACKNOWLEDGEMENTS

ENB/114606/2009 and PTDC/EEI-SII/2312/2012. Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

# REFERENCES

Arbeláez, P., Maire, M., Fowlkes, C., and Malik, J. (2011). Contour detection and hierarchical image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 33(5):898–916.

Association, A. (2013). 2013 Alzheimer's disease facts and figures. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 9(2):208–245.

Beucher, S. and Lantuejoul, C. (1979). Use of watersheds in contour detection. In *Int. Work. on Image Processing: Real-time Edge and Motion Detection/Estimation*.

Bicacro, E., Silveira, M., Marques, J. S., and Costa, D. C. (2012). 3D image-based diagnosis of Alzheimer's disease: bringing medical vision into feature selection. In *Int. Symp. on Biomedical Imaging*, pages 134–137.

Chaves, R., Ramirez, J., Gorriz, J. M., Lopez, M., Alvarez, I., Salas-Gonzalez, D., Segovia, F., and Padilla, P. (2009). SPECT image classification based on NMSE feature correlation weighting and SVM. In *IEEE Nuclear Science Symp. Conf. Record*, pages 2715–2710.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20:273–297.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Int. Conf. on Knowledge Discovery and Data Mining*, pages 226–231.

Fan, Y., Batmanghelich, N., Clark, C., and Davatzikos, C. (2008). Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *NeuroImage*, 39:1731–1743.

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

Lopez, M., Ramirez, J., Gorriz, J. M., Salas-Gonzalez, D., Alvarez, I., Segovia, F., and Chaves, R. (2009). Multivariate approaches for Alzheimer's disease diagnosis using bayesian classifiers. In *IEEE Nuclear Science Symp. Conf. Record*, pages 3190–3193.

Mikhno, A., Nuevo, P. M., Devanand, D. P., Parsey, R. V., and Laine, A. F. (2012). Multimodal classification of dementia using functional data, anatomical features and 3D invariant shape descriptors. In *Int. Symp. on Biomedical Imaging*, pages 606–609.

Morgado, P., Silveira, M., and Marques, J. S. (2013). Efficient selection of non-redundant features for the diagnosis of Alzheimer's disease. In *Int. Symp. on Biomedical Imaging*, pages 640–643.

Ramírez, J., Górriz, J. M., Salas-Gonzalez, D., Romero, A., López, M., Álvarez, I., and Gómez-Río, M. (2013). Computer-aided diagnosis of Alzheimer's type dementia combining support vector machines and discriminant set of features. *Information Sciences*, 237:59–72.

Segovia, F., Górriz, J. M., Ramírez, J., Salas-Gonzalez, D., Álvarez, I., López, M., and Chaves, R. (2012). A comparative study of feature extraction methods for the diagnosis of Alzheimer's disease using the ADNI database. *Neurocomputing*, 75:64–71.

Silveira, M. and Marques, J. S. (2010). Boosting Alzheimer disease diagnosis using PET images. In *Int. Conf. on Pattern Recognition*, pages 2556–2559.

Tran, T. N., Nguyen, T. T., Willemsz, T. A., van Kessel, G., Frijlink, H. W., and van der Voort Maarschalk, K. (2012). A density-based segmentation for 3D images, an application for X-ray micro-tomography. *Analytica Chimica Acta*, 725:14–21.

Tripathi, S., Kumar, K., Singh, B. K., and Singh, R. P. (2012). Image segmentation: a review. *Int. Journal of Computer Science and Management Research*, 1(4):838–843.

Varma, S. and Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7(91).

Ye, J., Farnum, M., Yang, E., Verbeeck, R., Lobanov, V., Raghavan, N., Novak, G., DiBernardo, A., and Narayan, V. (2012). Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data. *BMC Neurology*, 12(46).

Zhang, D., Wang, Y., Zhou, L., Yuan, H., and Shen, D. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage*, 55(3):856–867.