Sparse methods for Alzheimer's Disease classification

Miguel Ferro Passinhas Medeiros Rodrigues
miguel.passinhas@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

May 2018

Abstract

The diagnosis of Alzheimer's disease (AD) using image biomarkers establishes a complex computational challenge. The high dimensionality of this problem is a consequence of the lack of catalogued patients in the databases compared to the number of variables from the images. Nowadays, most of the decisions are still made by physicians without the support of automatic diagnosis tools. The development of a framework able to help and expedite the medical specialists decisions is the main objective of the present study. In this tool, a series of sparse methods capable of dealing with high dimensional problems were used, inducing sparsity in the models created to automatically detect a patient with AD. The two sparse methods used are both based on $L1-norm$ regularization parameter. The Lasso method, where the voxels from PET scans are used as an input, and the Group Lasso, where the spatial information between voxels, based on brain regions, is explored. The groups of voxels used in this study were based on the Atlas Oxford-Harvard brain segmentation and alternative regions based on Gaussian Mixture Models using cognitive normal patients. The experimental results with the Lasso achieved on the ADNI database were consistent when compared with the state-of-art. Furthermore, they improved the results with the introduction of groups by the Group Lasso when the Atlas and clustered regions were used. However, regions selected manually by experts remain to obtain better accuracy results, stating the need to improve how the regions are selected.

Keywords: Alzheimer's Disease (AD), Sparse learning methods, Positron emission tomography (PET), Classification, Regions of interest (ROI)

1. Introduction

1.1. Motivation

Alzheimer's Disease (AD) is not a recent concern, although, over the past few years, this area has seen further technological advances, which have led to improvements on stability of detection parameters and of new techniques for the AD diagnosis. In order to improve the sensitivity of the Alzheimer's Disease diagnosis, several studies in the neuroimage field have been developed. These studies, where the present master thesis is framed, are crucial to support the medical professionals in their decisions and increase the speed and the possibility of early diagnosis. The work now presented is developed in order to probe the progress of computer technology in applying machine learning, in particular sparse learning methods to a set of neuroimage scans.

2. Biomarkers

The study and diagnosis of AD are supported by two main biomarkers. The cerebral spinal fluid (CSF), a fluid biomarker present in the extracellular space of the brain, and the imaging biomarkers like MRI and FDG-PET. The biomarkers studies based on CSF rely on a quantitative interpretation in comparison with the standards and can measure different levels of proteins and other biochemical changes inside the brain, such as Amyloid Beta and Tau protein. This biomarker has been shown, in several studies present in a review article from ADNI [14], to be able to identify early stages of atrophy in different brain regions through the analysis of Amyloid Beta peptide ($A\beta - 42$) and Tau protein ($P - tau$) levels. These characteristics from CSF biomarker can be used together with the structure detail possible from the imaging biomarkers increasing the information available for the diagnostic and study the evolution of AD.

The imaging biomarkers provide information about structural and functional details of the brain using several types of scans. In this work the main focus will be the three-dimension fluorodeoxyglucose positron emission tomography (FDG-PET) scan. This biomarker uses fluorodeoxyglucose (FDG) injected into the patient bloodstream as a radioactive tracer to measure the metabolism from different regions of the brain. The more active brain regions are, the more tracer is accumulated making those regions appear brighter in the images. In Fig-

ure 1 is possible to observe the difference in a PET image from a normal cognitive patient and a patient with AD, where warmer colors indicate regions with more activity.

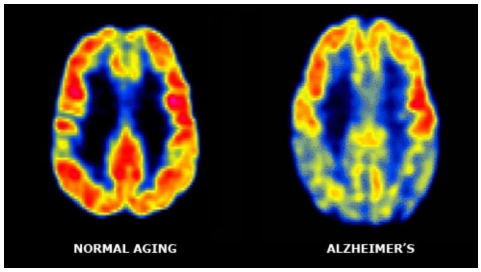


Figure 1: PET image showing the metabolic differences from a normal cognitive elderly person and a Alzheimer's disease patient, adapted from [4]. With red being a high FDG and blue low.

2.1. Brain Image Segmentation

The imaging biomarker is used to visualize volume and analyze anatomical structures. The visualization of these properties is only possible with a proper image segmentation, which is one of the most critical phases in medical image analysis. Within its large scope of applications, image segmentation can be used for surgery and radiation therapy planning, automatic labeling of anatomical structures and in many other fields of research. The main purpose of this technique is to find or extract regions of interest in a given image. In the particular case of this work, the main task of image segmentation is to identify and group similar anatomic regions in the brain using FDG-PET images. However, this can be a difficult task due to the low contrast, to the similar intensity values between features and to the existence of blurry images making it harder to distinguish the region contours, even manually. To overcome this problem, a variety of different segmentation methods that can be grouped into five different categories according to the feature they are based on: manual, intensity-based, atlas-based, surface-based (a template shape for a given brain structure) and hybrid segmentation.

The manual segmentation refers to a method where a human operator, e.g. an expert physician, labels by hand every region in the image turning this method highly expensive and inefficient. The intensity-based method includes all methods that use the pixels or voxels (tiny volumetric cube unit in a 3D image) intensity as the main feature to determine the different regions. This can be a simple intensity threshold or an unsupervised and automatic method such as clustering.

Besides the two previous approaches, one of the most efficient methods is the atlas-based. It is important to have in mind that an Atlas is a reference map containing the label information and the spatial coordinates of each brain structure, one example of that can be seen in Figure 2. To handle this method the first step is to create the atlas reference map. One of the ways to do that is by calculating a probabilistic model based on a training dataset that contains several images previously labeled. After the Atlas map is created, this method is straightforward and efficient to apply in images aligned to it because the reference map only needs one step to segment the new images.

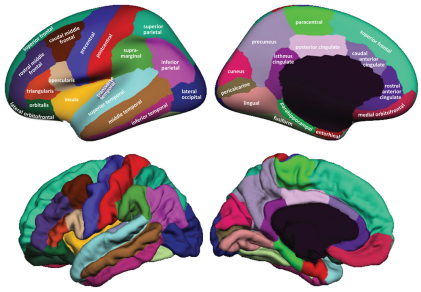


Figure 2: Atlas regions and surfaces example based on the Mindboggle 101 database from [7]

3. Background

The present chapter will feature a description of the methods adopted to perform brain segmentation. The sparse learning techniques used to solve computational problems of AD classification, through the use of a database of PET images.

3.1. Data Normalization

Two types of normalization were used, the first one is a reference cluster normalization and was specifically developed in [15] for the AD classification, the second one is a standardization applied to all the features in the images. This image normalization improves the algorithm's performance and prediction.

The reference cluster normalization approach was first introduced to remove the bias caused by the cerebral global mean normalization (CGM). The normalization used in the image database used in this thesis is based on Yakushev [15] reference cluster method. Yakushev's method only needs two iterations to complete the normalization. In the first one, a voxel-based analysis is performed with CGM. This analysis leads to appearance of some hypermetabolic regions in areas that are known to not be affected by the AD, this happens because the CGM is significantly decreased in patients with AD. In the next step, a statistical t-map is created with the results obtained from the first iteration and only the hypermetabolic voxels with a thresh-

old of $t - values > 2$ ($p < 0.05$) are included in the reference cluster. In the final step, all the data is normalized in respect to the reference cluster created. Firstly, the mean value from the reference cluster voxels of each image is computed and then all the voxels of that image are multiplied by the mean value correspondent. This method has shown improvements in classification results in AD vs CN and CN vs MCI [15].

After the Yakushev normalization, a standardization is required to center the data to be able to use the logistic regression for classification. The features are rescaled to have the characteristics of a standard normal distribution with mean zero and standard deviation equal to one. This normalization is calculated for each feature using the z-score formula, equation (1).

$$Z_n = \frac{X_n - \mu}{\sigma} \tag{1}$$

Where X_n is the feature from the n sample to be normalized, μ the mean and σ the standard deviation of the feature from all the samples.

3.2. Fixed Brain Segmentation

The probabilistic Atlas Harvard-Oxford [6], shown in Figure 3, which is composed by 48 cortical and 21 subcortical areas, was created by the Oxford Centre for Functional MRI of the Brain (FMRIB) with data structure and segmentations provided by the Harvard Center for Morphometric Analysis (CMA). This atlas was created from images with 21 healthy male and 16 healthy female subjects with ages between 18 and 50 years old that were individually labeled and then combined to form a probability map for each label.

Figure 3: The combination of Atlas Subcortical and Cortical areas

In order to use both subcortical and cortical atlases at the same time in the developed group sparse method a combined atlas comprising a total of 67 areas was created. This combined atlas, shown in Figure 3, was obtained by removing the left and right cortex from the subcortical atlas and merging it with the cortical atlas and finally removing the remaining overlapped voxels from the subcortical atlas White Matter areas.

The manually created ROI used in this work was created with prior medical knowledge by consulting the expert doctor Durval Campos Costa from Champalimaud Clinical Centre. Dr. Durval Campos Costa manually identified ten regions as being the most important regions of the brain to identify AD in a imaging test, resulting in a total of approximately 159000 voxels from the brain.

3.3. Clustering

In most cases, using predefined ROIs to perform brain segmentation is a suitable approach. However, the regions do not always adjust to the database for several reasons such as different type of MRI or PET scanners, or the difference between subjects. In the attempt to improve the brain segmentation, the features and voxels were re-arranged in groups that have the same characteristics, that is, using clustering algorithms. The probability based clustering using Gaussian mixture models (GMM) with the Expectation-Maximization (EM) algorithm [12], will be discussed. Considering the N observations from $X = \left\{x^{(1)}, ..., x^{(N)}\right\}$ without labels, it is assumed that GMM can give a density estimation for X with $p(x)$ with k Gaussian mixture models, equation (2).

$$p(x|\lambda) = \sum_{n=1}^{k} \alpha_n f_n(x|z_n, \theta_n) \tag{2}$$

where $f_n(x|z_n, \theta_n)$ denotes the multivariate Gaussian density n with parameter vector $\theta_n = \{\mu_n, \Sigma_n\}$, α_n is the mixing proportion for f_n with $\sum_{n=1}^{k} \alpha_n = 1$, z is a vector with k variables indicating the mixture component that generated x and λ is the set of parameters that describes the Gaussian mixture models with $\lambda = \{\alpha, \theta\}$. For each component n the multivariate Gaussian density function is given by equation (3) with mean μ_n and variance Σ_n.

$$f_n(x|\theta_n) = \frac{1}{(2\pi)^{d/2}|\Sigma_n|^{1/2}} e^{-\frac{1}{2}(x-\mu_n)^T \Sigma_n^{-1}(x-\mu_n)} \tag{3}$$

To estimate all the parameters of the Gaussian mixture model , the Expectation Maximization method is used. This is an iterative algorithm type with two main steps. The first one, the E-Step tries to guess the values of z, while the M-Step computes the new parameter values of the models taking into account the assignment in E-Step.

1. E-Step, for each i, n, with $0 < i < N$ and $0 < n < k$

$$w_n^{(i)} = p(z^{(i)} = n|x^i, \alpha_n, \mu_n, \Sigma_n) = \frac{f_n(x^{(i)}|z^{(i)}=n, \mu_n, \Sigma_n) \cdot \alpha_n}{\sum_{j=1}^{k} f_j(x^{(i)}|z^{(i)}=l, \mu_l, \Sigma_l) \cdot \alpha_l} \tag{4}$$

2. M-Step, update the parameters with new w (4).

$$\alpha_n = \sum_{i=1}^{N} w_n^{(i)} \tag{5}$$

$$\mu_n = \frac{\sum_{i=1}^{N} w_n^{(i)} x^{(i)}}{\sum_{i=1}^{N} w_n^{(i)}} \tag{6}$$

$$\Sigma_n = \frac{\sum_{i=1}^{N} w_n^{(i)} (x^{(i)} - \mu_n)(x^{(i)} - \mu_n)^T}{\sum_{i=1}^{N} w_n^{(i)}} \tag{7}$$

3. Check if there is convergence: compute the log-likelihood, equation (8), and compare with the previous iteration $(t-1)$,

$$L(t) = \log l(\Theta) = \sum_{i=1}^{N} \log f_n(x^{(i)}|\Theta) \tag{8}$$

$$\Delta = L(t-1) - L(t) < \epsilon \tag{9}$$

given a threshold, ϵ, if $\Delta < \epsilon$ equation (9) stop, otherwise repeat.

4. Sparse Learning Methods

Sparse methods are a powerful tool to deal with high dimensional problems and performing feature selection. These methods were applied in our dataset after the data normalization and brain segmentation are done. This step is responsible to reduce the feature space and also to build the classifier. Firstly, without any type of brain segmentation a voxel-based approach was used in the regularization method known as Lasso. In this algorithm, all the internal brain voxels intensity are considered as input and it is not taken into account any type of spatial information between voxels. In the second type tested, Group Lasso, the spatial information is explored by using groups of features that were obtained as a result of the brain segmentations . The Least Absolute Shrinkage and Selection Operator known as Lasso, is a well-known and powerful method capable of doing regularization and feature selection using the L_1-norm. This method uses a penalization term which is added to the cost function in order to regularize the coefficients and perform feature selection by zeroing some variables. It is controlled by the factor λ, where the bigger is, the fewer features are selected to the model. In this thesis a ratio of λ_{max} was used to measure the regularization, where the λ_{max} is the maximum value of λ where above it shall obtain the zero solution model. With the value of $\lambda = 1$ being a fully regularized model and $\lambda = 0$ a model without any regularization.

4.1. Sparse Logistic Regression

The Lasso was applied with logistic regression to estimate the parameters for the classifier and to perform the classification. The final objective is to train a logistic regression classifier capable of distinguishing between two labels $y^{(i)} \in \{-1, +1\}$ obtained by using N training samples given by $X = \left\{(x^{(i)}, y^{(i)}), ..., (x^{(N)}, y^{(N)})\right\}$. The logistic regression classifier calculates the conditional probability of a given sample be a associated to the label y, in equation (10).

$$p(y^{(i)}|\omega, x^{(i)}) = \frac{1}{1 + \exp(-y(\omega^T x^{(i)} + c))} \tag{10}$$

Where ω is the weight vector to be calculated and c is the value of the intercept. The $\left(\omega^T x^i + c\right) = 0$ is the separating hyperplane in the feature space where $p(y|\omega, x) = 0.5$. For values above this hyperplane, $\left(\omega^T x^{(i)} + c\right) \geq 0$ we have $p(y|\omega, x) \geq 0.5$ and the label assigned is the positive value $y = 1$, otherwise we have $y = -1$ [8].

To build the classifier, the logistic loss function from equation 11 is used. This loss function is calculated over N samples with the estimated weight w vector and c scalar, and give a measure of the degree of fit to the estimated parameters.

$$f(w, c) = \sum_{i=1}^{N} \log(1 + \exp(-y^{(i)}(\omega^T x^{(i)} + c))) \tag{11}$$

The loss function is minimized in equation 12 and a penalizing term Ω is added to induce sparsity [8], which in this work is given by the Lasso and Group Lasso.

$$\min_{w,c} f(w, c) + \Omega \tag{12}$$

4.2. Lasso

The Lasso penalizing term with $L_1 - norm$ regularization is given by $\Omega = \|w\|_1$ and estimates the parameters w and c minimizing the regularized logistic regression loss function in equation (13) [9]. How the regularization parameter λ is chosen in this work is explored in the next section.

$$\min_{w,c} \sum_{i=1}^{N} \log(1 + \exp(-y^{(i)}(\omega^T x^{(i)} + c))) + \lambda \|w\|_1 \tag{13}$$

4.3. Group Lasso

This approach of Lasso considers how the features are grouped, they can be grouped by the fixed structures or by the calculated brain regions. In group lasso the parameter models are penalized with the $L_{2,1} - norm$ given by the equation (14) and results given by equation (15).

$$\Omega = \lambda \sum_{i=1}^{k} w_i^g \|w_{G_i}\|_2 \tag{14}$$

$$\min_{w,c} \sum_{i=1}^{N} \log(1 + \exp(-y^{(i)}(\omega^T x^{(i)} + c))) + \lambda \sum_{i=1}^{k} w_i^g \|w_{G_i}\|_2 \quad (15)$$

Where k is the number of total groups, w_i^g is the weight for i group, and w_{G_i} through w_{G_k} are divided into k non-overlapping groups. In this work the weights for the groups were of $w_i^g = \sqrt{d_i}$, where d_i is dimension of the group i, according to the study from [10] this penalization gave better results and is capable to penalize larger groups in a heavier way.

5. Implementation

5.1. Database description

The data used in this thesis was provided by the Alzheimers Disease Neuroimaging Initiative (ADNI) group. Since 2004, the ADNI mission has been helping the scientific community through the validation of biomarkers use, including blood tests, cerebrospinal fluid tests and MRI/PET for AD's diagnosis and studies. The three main purposes of this group are the ADs early detection and the identification of biomarkers to track the disease; the application of new diagnosis methods at the initial stages of the disease, so that any intervention can me more effective; and building a shareable database with several researchers around the world.

The PET biomarkers are three-dimensional images representing the brain volume and have a dimension of $121 \times 145 \times 121$ voxels, in a total of 2122945 voxels, where each voxel represents the intensity value over a small volume. Axes orientation in the brain can be seen in the Figure 4, where, the x axis varies from 0 to 121, increasing from left side to the right side of the brain, the y axis varies from 0 to 145, increasing from back (posterior) to front (anterior) side of the brain and the z axis varies from 0 to 121, increasing from inferior to superior side of the brain.

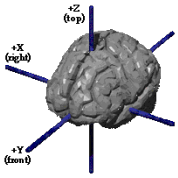


Figure 4: Brain 3D axes orientation used in the PET images. Adapted from [2]

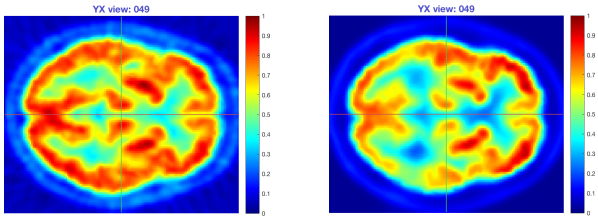
5.1.1 Participants

In the database used for this work there are a total of 133 PET images labeled with Alzheimer's disease (AD) or cognitive normal (CN) being all from different subjects. The database description is built with the patients number, gender, average age and average score in two cognitive tests, Neuropsychiatric Inventory Questionnaire (NPIQ) and Functional Assessment Questionnaire (FAQ) (in both the higher score the greater impairment). This description can be seen in Table 1.

Table 1: PET database from ADNI used in the tests

Type	AD	CN
Subjects	58	75
Gender	34 M & 24 F	49 M & 26 F
Age (years)	76.12 ± 6.59	76.03 ± 4.68
NPIQ (max 36)	3.90 ± 3.45	0.39 ± 0.80
FAQ (max 30)	13.55 ± 6.83	0.24 ± 0.91

In Figure 5 subject labeled with AD and other CN with the respective color grading where bigger value of intensity means more brain activity is represented. Both images represented are slices from plane YX with $z = 49$, looking to them, it is possible to conclude that in the AD case the voxels have lower intensity when compared with the CN subject, and in some particular areas the AD subject exhibit values near zero intensity where the CN slice has normal values, which means, AD subject have already lost brain activity in that region.



(a) CN subject (b) AD subject

Figure 5: Two different subject PET images: one CN and another with AD. Both slices in the YX plane with $z = 49$

5.2. Data Normalization

To improve algorithm's performance and prediction the reference cluster normalization, [15] [15], was used. As previously referred this normalization is based on the appearance of hypermetabolic regions, with the cerebral global mean normalization (CGM), in the conserved regions. With that and resorting a statistical parametric mapping (SPM), the reference cluster with highest t-values voxels is selected between the two groups (AD vs CN).

5.3. Tests Design

The tests performed in this thesis can be separated into two distinct types, one uses voxel as a feature for the Lasso algorithm to build the classifier, and the other uses group of voxels as features for the Group Lasso algorithm, taking into account the spa-

tial information of the brain. This spatial information is provided by a series of pre-labeled regions containing the relation between the anatomical areas and their location in a three dimensional matrix with pre-labeled voxels.

The labeled regions can be obtained directly from an expert radiologist/physician or can be a result of several segmented brain labeled images joined together. Both of these types are referred in this thesis as "fixed areas".

To complement this type of "fixed areas" an alternative type of groups obtained using unsupervised learning was developed and tested. This new groups were developed based on Gaussian Mixture Models and Expectation Maximization algorithm which cluster the voxels in several groups taking into account voxel intensity and their spatial location. This method has the advantage to better fit the segmentation in the used database. In the flowchart of Figure 6 the types of tests executed and how they group together is represented.

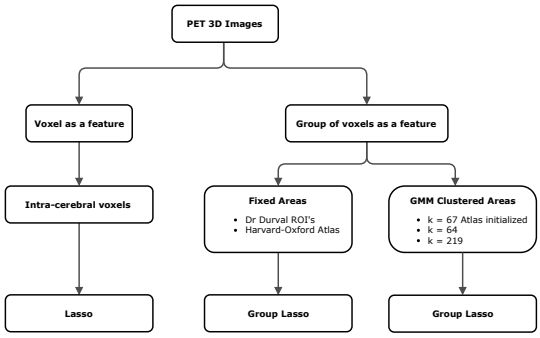


Figure 6: Tests by type of feature used and areas.

5.3.1 Classifier Performance

To evaluate the classifier in the tests described before a k-fold cross-validation evaluation with $k = 10$, will be used. This technique try to access how the results will be in a generalized database. It is one of the most used methods to evaluate the accuracy of a predictive model. The first step of the cross-validation is the partition of data in k equal sized parts and then perform k tests using $k - 1$ parts of the data for training the model and leaving the remaining part of data in each "fold" to test the model. At the end, the final score for the k cross-validation test is the average of the all k tests. This technique is represented in the schematic of the Figure 7.

The performance of the classifier will be evaluated using tree different metrics: accuracy, sensitivity and specificity. Accuracy in equation 16 gives the proportion of correct labeled subjects. While sensitivity (true positive rate), equation 17, and

Figure 7: K-fold cross-validation scheme. Adapted from [1]

specificity (true negative rate), equation 18, give the accuracy over the positive cases and negative cases, respectively.

$$Accuracy = \frac{TruePositive + TrueNegative}{TotalPopulation} \quad (16)$$

$$Sensitivity = \frac{TruePositive}{TruePositive + FalseNegative} \quad (17)$$

$$Specificity = \frac{TrueNegative}{TrueNegative + FalsePositive} \quad (18)$$

5.4. Brain Image Segmentation

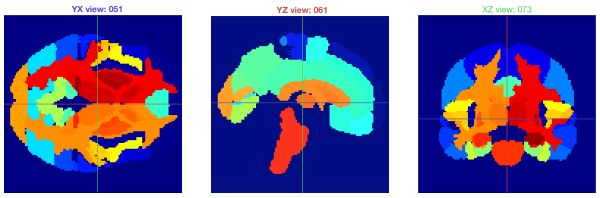
To be able to incorporate spatial information of the brain with Group Lasso, two types of brain segmentation were used. In the next sections, details on each type of regions used and how they were obtained in the case of the developed ones will be presented.

5.4.1 Fixed Areas

The tests performed with fixed areas resorted to the Atlas from Harvard-Oxford and the manually selected regions of interest by the specialist in the area from Champalimuad Foundation, Dr. Durval Campos Costa. This atlas was created by the Oxford Centre FMRIB in collaboration with Harvard CMA and resulted in a structure with a total of 67 labeled regions separated in two different subatlas the subcortical and cortical areas containing 21 and 48 labeled regions respectively. The merged Atlas is represented in Figure 8 with three plane views, where each color is associated to a distinct structure. Regions symmetry and how well they are structured and delimited are also observable in the same figure. Note that in the next sections, to analyze the test results and the feature selection this Atlas will be also used as a reference to locate voxels in anatomical areas.

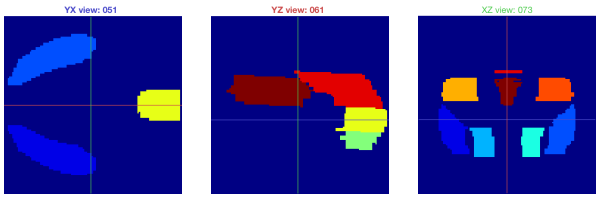
Dr. Durval's manual segmentation aims to reduce the number of voxels used and to select the

most meaningful areas for AD diagnosis. This resulted in a structure with a total of 10 distinct areas and with a total of 159082 voxels. A view of the three different planes is represented in Figure 8.



(a) YX (b) YZ (c) XZ

Figure 8: Harvard-Oxford Atlas with subcortical and cortical areas combined.



(a) YX (b) YZ (c) XZ

Figure 9: Manually selected regions of interest (ROIs) for AD diagnosis

6. Results

In this section, the classification results for all the tests will then be presented. In particular, the classification accuracy and their respectively specificity and sensitivity will be analyzed. Next, the feature selection performed in the different types of areas, the most important and most selected areas in the brain, will be explored. In the end, the best results of each method can be found summarized in the Table 3.

The tests were performed through the SLEP toolbox [9] for MATLAB in which the functions *LogisticR* and *glLogisticR* for Lasso and Group Lasso respectively, were adopted. Both algorithms were tested with the regularization input $\lambda \in [0, 1]$ as the ratio of λ_{max}, that is automatically computed by the program. In this tests the values set for λ are in a logarithm scale to be able to explore and perform more tests in the low values of regularization with $\lambda < 0.2$, because in the major part of the earlier tests performed the best models had low regularization values. Thus, the tests were performed with a $\lambda = 10^{-5:0.2:0}$. The plot with the final accuracy results for each method is represented in Figure 10 Analyzing this figure it is possible to observe that in general all the methods have a similar behavior for the different λ and with the maximum value of accuracy for each method varying between 91% and 87%. The Group Lasso Manual ROIs the method that combines the Group Lasso algorithm with the manually selected ROIs, achieved the overall highest result with an accuracy of 91.10%. The different types of Group Lasso outperformed the Lasso used with all brain voxels (considered as the baseline), with exception only for the Group Lasso Cluster $k = 219$ which was the worst method with respect to the accuracy. From Figure 10 it is also possible to conclude that methods with manually selected regions had better performance when compared to the clustered groups and the chosen atlas. This behavior was verified in both Group Lasso and Lasso, so the features selected by the algorithm taking into account the spatial relations can not outperform the usage of the Group Lasso combined with pre-selected regions by an expert in the field. With this result it is important to make clear to the reader that the manually selected regions have not only grouped the regions of interest for AD diagnosis but also made feature selection. The manual regions only use a part of the brain voxels when all the remaining groups and Atlas treated the entire brain. More precisely, in these last approaches it was used all the brain with a total of 486031 voxels when compared to the 159082 voxels in the manually selected ROIs. Finally, when we compare the Atlas groups and clustered regions in the accuracy plot it is possible to observe that they have a very similar result. In fact, the maximum value of accuracy for the Atlas, Cluster Atlas and Cluster $k = 64$ is exactly the same, differing only in the final number of voxels selected.

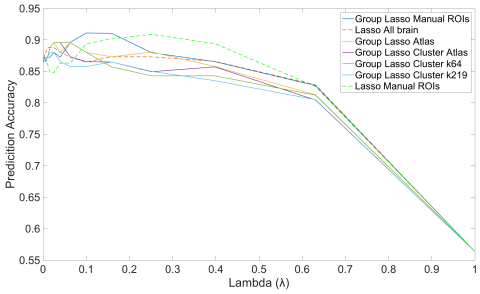


Figure 10: Prediction accuracy comparison for different lambda

Now, with the accuracy analyzed it is important to have a perception of how many voxels are needed for each method to achieve a classification score, so in Figure 11 the number of features selected by each method and their correspondent accuracy obtained are represented. Using this plot it is possible once again to conclude that both methods that used the manually selected regions obtained far better results using much less voxels when compared to the others, which helps to prove that the manual selection performed by Dr Durval is accurate. Additionally, the two Lasso tests were the only that, using less

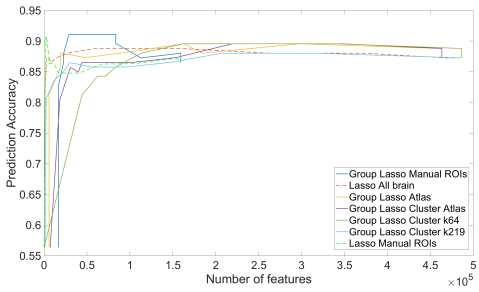
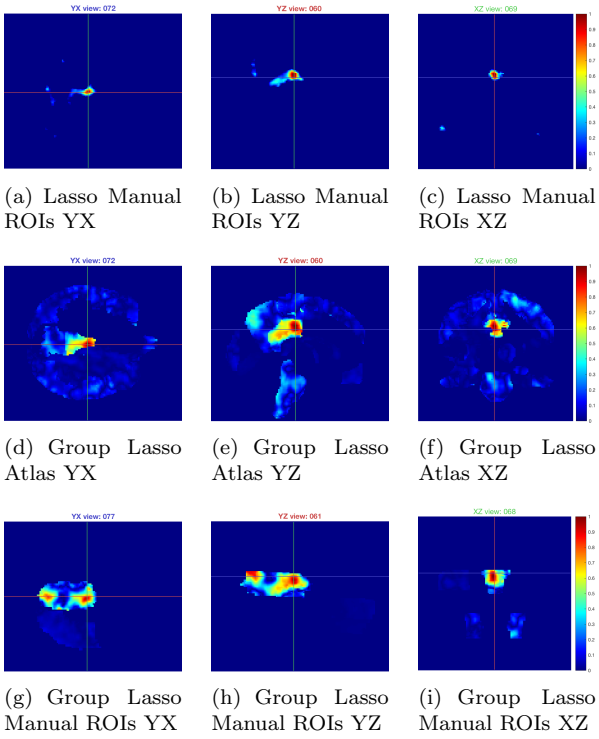


Figure 11: Prediction accuracy comparison for the different number of features

Table 2: Brain regions that have the highest 25% weights of each method, corresponding to the red areas

Method	Total voxels [-]	Highest voxels [%]	Area name
Lasso All brain	60	46	Cingulate Gyrus
		21	Postcentral Gyrus
		19	Precentral Gyrus
Lasso Manual ROIs	35	91	Cingulate Gyrus
Group Lasso Atlas	97	100	Cingulate Gyrus
Group Lasso Cluster Atlas	89	60	Cingulate Gyrus
		27	Postcentral Gyrus
Group Lasso Manual ROIs	120	75	Cingulate Gyrus
		25	Precuneous Cortex
Group Lasso Cluster k64	50	52	Cingulate Gyrus
		22	Right Cerebral WM
		18	Postcentral Gyrus
Group Lasso Cluster k219	111	71	Cingulate Gyrus
		21	Left Cerebral WM



(a) Lasso Manual ROIs YX (b) Lasso Manual ROIs YZ (c) Lasso Manual ROIs XZ

(d) Group Lasso Atlas YX (e) Group Lasso Atlas YZ (f) Group Lasso Atlas XZ

(g) Group Lasso Manual ROIs YX (h) Group Lasso Manual ROIs YZ (i) Group Lasso Manual ROIs XZ

Figure 12: Weights of each method; red indicates a greater absolute value and the blue indicate a value near to zero.

than one thousand features, were able to perform a reliable feature selection and creating models with an accuracy above 85%, with the one that used only voxels from the manually selected regions being the best and achieving a value of accuracy above 90%.

After discussing the accuracy results and how it relates to the number of features selected, a complete analysis of the feature selection localization in the brain is needed. But in order to do this, is necessary to understand how the methods perform the feature selection. To conduct this analyses from each pair of method and area, the best correspondent model, in respect to accuracy obtained, is selected and their weight structure analyzed, through visualizations present in Figure 12. To fully describe the information present in this figures values for which regions have the higher weights in Table 2, are presented.

Comparing the visualizations from Figure 12 it is possible to observe that they have similar spatial feature selection and similar weights intensity distribution across all the methods. As expected with a high regularization value of $\lambda = 0.251$, the Lasso Manual ROIs is the method that has less number of voxels selected. More precisely, a total of 22606 voxels were selected when using this model and the majority of the features selected have weights with a higher value (the red areas in the Figure 12 (a), (b) and (c)). The most relevant weights are all located in the cingulate gyrus. Employing the same segmentation, the manual ROIs, but resorting to the Group Lasso algorithm the model achieved the best accuracy result overall. The feature selection for the $\lambda = 0.1$, in the Figure 12 (g), (h) and (i), selected almost the entire frontal medial cortex and Amygdala areas that are shown in other studies [11] to be one of the important areas for the AD diagnosis. The highest weights in this model are located in the cingulate gyrus and precuneous cortex.

6.1. Discussion

It is summarized in the Table 3 a comparison of the best results from each method and area obtained in this work. The combination of voxels spatial information with the Group Lasso improved the classifier results when compared to the Lasso method. In the clustered atlas and the region with 64 clusters the results were very similar to the fixed Atlas Oxford-Harvard with all segmentations obtaining a score of accuracy of 89.56%. Comparing the feature selection made by the manual ROIs with the other segmentations used, it is possible to conclude that this selection is accurate and improved the classification in the both methods used. In fact, in both cases was the best score obtained. In the best created models some of the most selected areas and with the most important weights, in Table 2, were the parahippocampal gyrus, the amygdalas and the

Table 3: The best models from each method and region with all the scores from the classifier performance and percentage of the brain voxels selected.

Method	Regions	Lambda [%]	Accuracy [%]	Sensivity [%]	Specificity [%]	Running time [s]	Features used [%]
Lasso	All brain	0.0251	88.79	89.67	88.39	15.32	12.22
	Manual ROIs	0.2512	90.88	86.00	94.64	4.08	4.65
Group Lasso	Manual ROIs	0.1000	91.10	90.00	92.32	1.21	17.38
	Atlas	0.0631	89.56	88.00	91.07	6.06	35.14
	Cluster Atlas	0.0398	89.56	89.67	89.83	6.25	47.31
	Cluster k=64	0.0631	89.56	89.67	89.82	6.06	33.39
	Cluster k=219	0.0251	88.02	87.67	88.39	8.34	42.61

cyngulate gyrus. These areas have been mentioned in previous studies [5, 13, 11, 3] as being some of the most relevant areas to the AD diagnosis.

Finally, it is important to note that besides using these type of manual segmentations in the tests, they are not the main target for the studies in this area because it needs a previous assignment from a specialist in the field. However, they are useful and can be used as a benchmark to other tested areas.

7. Conclusions

The work developed in this thesis addressed the complex problem in the neuroimaging field applied to the AD diagnosis using machine learning algorithms. The sparse learning methods in order to reduce the feature space of the problem and improve the accuracy in the AD diagnosis were exposed and explored. Alternative brain image segmentations were also created to test how different regions adjusted to the CN subjects from the database were able to improve AD detection.

The solutions presented with Lasso and Group Lasso were capable of obtaining accurate models, between 85% and 91% accuracy, with the PET high dimensional data through feature selection therefore resulting in a computational cost reduction. This results achieved, showed how powerful these sparse methods are, based in the $L1-norm$ to perform feature selection, using less than 12.6% of the features in some cases, which at the same time were able to build the classifier efficiently.

The Group Lasso algorithm, that uses group of voxels as input and takes into account the spatial information between them, improved the results when compared to the Lasso using the same voxels. The segmentations obtained in this work with the clustered regions had their best result at exactly the same accuracy as the best result of the pre-labeled Atlas. In general, this regions were not capable to bring any improvements regarding the classification performance to the already segmentated regions. The only improvement showed were in the case tested with $k = 64$ clusters that obtained the same best result as the Atlas but with less features selected to create the model, meaning that with this clustered segmentation it was performed a more accurate feature selection.

On the other hand, the tests executed with the regions that were manually selected by an expert in the field obtained clearly the best results in the both Lasso methods used. This is partly explained because, before running the sparse methods there were already a feature selection made by the physician, where the regions selected only use 35% of the total brain voxels. This pre-selection combined with the power of the feature selection from the sparse methods produced the best results from all the tests performed in this thesis.

Finally, the main objectives of this work to model and detect Alzheimer's Disease using PET images were successfully attained. The framework presented were capable to report quite accurate results to assess the AD state and capable to perform adequate feature selection that brought improvements to the model. It was also recognized that the spatial information provided to the algorithm by the regions can improve the final results, specially the one where specialized knowledge was used.

In order to improve the work developed in this thesis and the application of sparse learning methods to the AD diagnosis some recommendations for future work are suggested. One of the conclusions achieved with this work was that the regions that were already made a pre-selection of the brain voxels obtained the best results. Due to that observable behavior, one course in which the future work can be redirect is in order to introduce an additional step to create this pre-selection without medical knowledge and before modeling the data using the sparse methods. As a suggestion to surpass this limitation is to test a different type of unsupervised method combined with prior knowledge, as an example the creation of regions of interest using the watershed transform segmentation combined with the atlas registration. Other type of method based in the knowledge is to create new clustered regions based on the previously manually identified regions from several experts, as an example adopting the manual regions used in this thesis to create clusters of regions based only in the voxels selected.

References

[1] ML Study k-fold cross-validation.

http://karlrosaen.com/ml/ learning-log/2016-06-20/. Accessed: 2018-04-18.

[2] Orientation and Voxel-Order Terminology: ras, las, lpi, rpi, xyz and all that. http://www.grahamwideman.com/gw/brain /orientation/orientterms.htm. Accessed: 2018-04-15.

[3] E. Franko, O. Joly, and for the Alzheimer's Disease Neuroimaging Initiative. Evaluating alzheimer's disease progression using rate of regional hippocampal atrophy. *PLOS ONE*, 8(8):1–11, 08 2013.

[4] D. Janet Cochrane Miller. Radiology Rounds. *Radiology Rounds*, 4, 2006.

[5] B. F. Jones, J. Barnes, H. B. M. Uylings, N. C. Fox, C. Frost, M. P. Witter, and P. Scheltens. Differential regional atrophy of the cingulate gyrus in alzheimer disease: a volumetric mri study. *Cerebral cortex*, 16 12:1701–8, 2006.

[6] D. Kennedy, C. Haselgrove, J. Breeze, A. Neuropsychiatric, and L. Seidman. Templates and Atlases included with FSL. 2016.

[7] A. Klein and J. Tourville. 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in Neuroscience*, 6:171, 2012.

[8] J. Liu, J. Chen, and J. Ye. Large-scale sparse logistic regression. pages 547–555, 2009.

[9] J. Liu, S. Ji, and J. Ye. Slep: Sparse learning with efficient projections. 05 2013.

[10] Y. Nardi and A. Rinaldo. On the asymptotic properties of the group lasso estimator for linear models. 2008.

[11] S. Poulin, R. Dautoff, J. C Morris, L. Barrett, and B. Dickerson. Amygdala atrophy is prominent in alzheimer's disease and relates to symptom severity. 194:7–13, 09 2011.

[12] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital Signal Processing*, 10, 2000.

[13] M. Saroshi, G. Bruno, B. Stanley, F. K. A., F. N. L., and K. D. E. Metabolic reduction in the posterior cingulate cortex in very early alzheimer's disease. *Annals of Neurology*, 42(1):85–94.

[14] M. W. Weiner and D. P. Veitch. The alzheimer's disease neuroimaging initiative : A review of papers published since its inception. 2012.

[15] I. Yakushev, A. Hammers, A. Fellgiebel, and Schmidtmann. SPM-based count normalization provides excellent discrimination of mild Alzheimer's disease and amnestic mild cognitive impairment from healthy aging. *NeuroImage*, 44, 2009.