Fast brain-wide search of highly discriminative regions in medical images: an application to Alzheimer's disease

Eva Janousova[13]
e.janousova@imperial.ac.uk
Maria Vounou[1]
maria.vounou03@imperial.ac.uk
Robin Wolz[2]
r.wolz@imperial.ac.uk
Katherine Gray[2]
katherine.gray03@imperial.ac.uk
Daniel Rueckert[2]
D.Rueckert@imperial.ac.uk
Giovanni Montana[1]
g.montana@imperial.ac.uk

[1] Department of Mathematics
Imperial College London
London, UK

[2] Department of Computing
Imperial College London
London, UK

[3] Institute of Biostatistics and Analyses
Masaryk University
Brno, CZ

Abstract

We propose a fast algorithm for the identification of localised brain regions from medical images that discriminate between two groups of individuals. The method is based on a combination of penalised regression and a data resampling procedure. We apply this approach to both MRI and PET images for the classification of subjects with Alzheimer's disease and mild cognitive impairment. We show that the voxels selected by the algorithm form connected brain regions which are well known to be affected by Alzheimer's disease. A linear statistical classifier trained on the selected voxels achieves cross-validated classification results that are comparable to those obtained by current state-of-the-art methodologies.

1 Introduction

Early and accurate diagnosis of Alzheimer's disease (AD) is crucial for a timely and optimal treatment of patients. Neuroimaging techniques, such as MRI or PET, can help identify diagnostic biomarkers [4] and even predict later development of the disease. A comparison of supervised and unsupervised techniques for biomarker extraction has recently been published [3]. While *supervised* biomarkers encode prior knowledge about the disease and brain anatomy, such as hippocampal volume, shape or atrophy [2], *unsupervised* biomarkers do not require any a priori knowledge of structures that are involved in the disease process. In this paper we consider the latter approach and propose a statistical model for the automatic extraction of brain regions that can potentially be used as biomarkers. One of the key challenges of unsupervised biomarker extraction is given by the fact that the number of voxels

that need to be processed and tested is very large, so that many classical feature selection techniques will not scale up well or provide reliable results.

We propose a simple and computationally efficient approach for the unsupervised selection of discriminative voxels. The approach relies on a penalised regression model that encourages sparsity in the regression coefficients, thus performing voxel selection. This solution is combined with a data resampling scheme that provides a measure of discriminative importance for each individual voxel.

2 Methods

Let X be a $n \times p$ matrix composed of n brain images of independent subjects each with p voxels. We assume that the n individuals have clinical labels D (diseased) and H (healthy controls). The class label for subject i is represented by a binary variable y_i, such that $y_i = 1$ if individual i is in class D and $y_i = 0$ otherwise. We also assume that the vector y is mean centred and the columns of X, x_j, where $j = 1, \ldots, p$, have zero mean and constant variance.

We use the n images to identify localised groups of voxels whose intensities are highly discriminative between the two classes. We achieve the desired voxel selection by means of penalised regression. We treat the class indicator y as a response variable in a linear regression model whose predictors are given by the voxel intensities. Assuming a least squares loss function, we aim to estimate the linear regression coefficients $\beta = (\beta_1, \ldots, \beta_p)$ such that they satisfy two main properties: (a) only the coefficients corresponding to most discriminative voxels are non-zero, thus yielding a sparse estimate $\hat{\beta}$, which is achieved by introducing an l_1 penalty [7], and (b) the non-zero coefficients of correlated voxels should be smoothed towards a common value to allow for the selection of voxels in groups, which is obtained by imposing an l_2 penalty on the regression coefficients [8]. Accordingly, the estimates are found by minimising the following penalised least squares problem

$$\arg\min_{\beta} \left\{ \|y - X\beta\kappa\|_2^2 + \lambda \|\beta\kappa\|_1 + \mu \|\beta\kappa\|_2^2 \right\} \tag{1}$$

where $\lambda > 0$ and $\mu > 0$ are regularisation parameters introduced for the l_1 and l_2 penalties respectively, and where the scaling factor $\kappa = (1+\mu)^{-1}$ corrects for the double shrinkage caused by applying both penalties. By setting μ to infinity, we reduce the number of free parameters down to only one, λ, whilst still maintaining the grouping effect. This leads to a very computationally cheap estimation algorithm. The optimal β coefficients can be computed one element at a time by applying a simple soft-thresholding function,

$$\hat{\beta}_j = \text{sign}(x_j' y) \left(| x_j' y | - \frac{\lambda}{2} \right)_+ \qquad j = 1, \ldots, p \tag{2}$$

where $(\alpha)_+$ is defined as $\max(0, \alpha)$.

The regularisation parameter λ in Equation (2) controls the amount of sparsity, and therefore determines a set S containing the selected voxels. When λ is exactly zero, no penalty is imposed and all p voxels enter the set S. As λ increases away from zero, less voxels are retained. At its maximum value λ_{max}, no voxel is selected and S becomes the empty set. A common approach to model selection in sparse regression involves tuning λ, for instance by cross-validating the prediction error obtained for all values of $\lambda \in [0, \lambda_{max}]$, and then choosing the value of λ that provides the smallest cross-validated error. However, the selection

of the optimal λ might suffer from sampling errors, in the sense that a different λ, hence a different sparsity pattern, might arise from an independent data set.

To select highly discriminative voxels that are more robust against sampling errors, we adopt a data resampling scheme that has been proposed for sparse predictive modelling [6]. This procedure aims to obtain a measure of voxel importance by repeatedly fitting the sparse regression model on random subsets of the data set and keeping track of voxels that are consistently associated to non-zero regression coefficients. Specifically, for $\lambda \in [\lambda_{min}, \lambda_{max}]$, we draw B random sub-samples with replacement. For each sample we obtain a sparse estimate $\hat{\beta}^{(b)}(\lambda)$, where $b = 1, \ldots, B$. For each estimate, we determine which voxels have non-zero regression coefficients by using an indicator variable $v_j^{(b)}(\lambda)$ which is equal to 1 if the coefficient corresponding to voxel x_j is non-zero, or 0 otherwise. Using all B sub-samples, a measure of voxel importance is finally computed by estimating the selection probabilities

$$\mathrm{P}_j(\lambda) = \frac{1}{B} \sum_{b=1}^{B} v_j^{(b)}(\lambda) \quad j = 1, \ldots, p \tag{3}$$

Rather than tuning the regularisation parameter λ, we search for a set of voxels with high probability over a range $[\lambda_{min}, \lambda_{max}]$. The upper bound λ_{max} is determined to be the lowest value of λ that results in empty set S. The lower bound λ_{min} is determined using a search procedure. The final set of voxels to be included in S is obtained by choosing a threshold π on the selection probabilities, hence we denote by $S(\lambda_{min}, \pi)$. The optimal λ_{min} and π are chosen to minimise a measure of cross-validated classification error.

3 Experiment and Results

3.1 Data Sets

The proposed method was applied to MRI and FDG-PET data obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database[1]. T1-weighted 1.5 T baseline MRI scans were available for 838 subjects: 198 AD patients, 409 subjects with mild cognitive impairment (MCI) and 231 cognitively normal elderly subjects (CN). Within the MCI group, 168 subjects have so far been diagnosed with AD and are denoted by pMCI (progressive MCI), whereas the remaining subjects are denoted by sMCI (stable MCI). The baseline MR images were aligned with the MNI152 brain template using a coarse non-rigid registration regularised by a 10*mm* B-spline control-point grid. Brain extraction was performed based on automated tissue classification using SMP5[2]. Image intensities were normalised to the template using linear regression prior to performing Gaussian smoothing with a 4*mm* FWHM Gaussian kernel. PET images were available for 287 subjects: 71 AD, 62 pMCI, 85 sMCI, and 69 CN. Each PET image was converted to a 30-minute static and affinely aligned with the corresponding native-space MRI. The non-linear transformation parameters estimated to map the MR images to the MNI template were then applied to the MR-space PET images using a trilinear interpolation. These images were smoothed to a common isotropic spatial resolution, normalised and resampled to the higher resolution of the MRI. $1,650,857$ voxel intensities in MRI and PET images were used to perform voxel selection with the proposed method after correcting for age and gender using a linear regression model.

[1] www.loni.ucla.edu/ADNI
[2] www.fil.ion.ucl.ac.uk/spm

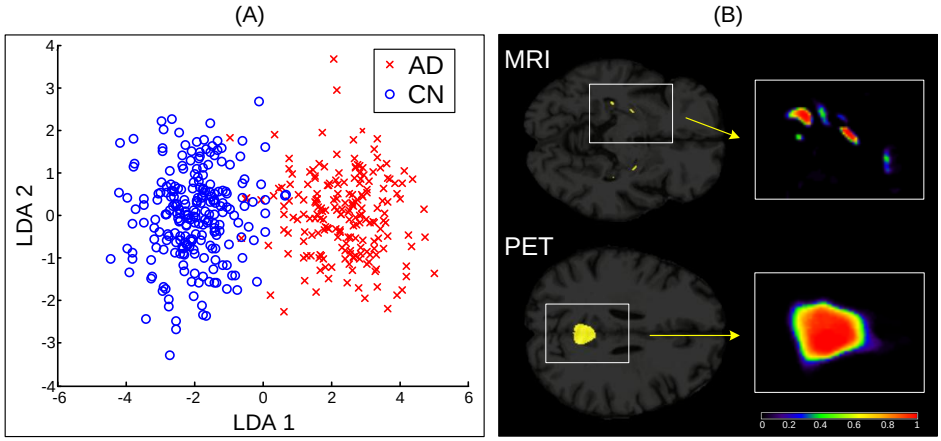

Figure 1: (A) Two-dimensional projections of training data points observed on the selected voxels (MRI data) show that AD and CN subjects are linearly separable. (B) Transversal slices showing the selected voxels (in yellow) and insets displaying selection probabilities for AD vs CN in MRI and PET data.

3.2 Classification results

Six independent classification experiments were performed, whereby we compare two groups in each experiment: AD vs CN, pMCI vs CN, and pMCI vs sMCI, using both MRI and PET data. For each experiment, voxel selection was carried out according to the procedure described in Section 2. Selection of the most discriminative voxels leads to linear separability of the pairs of subject groups in training sets as illustrated in Figure 1(A). Since the groups are linearly separable we can use linear discriminant analysis (LDA) to assess the discriminative power of the selected set of voxels, $S(\lambda_{min}, \pi)$. LDA does not require any parameter setting, therefore there are only two parameters λ_{min} and π to be optimised.

The optimal parameters λ^*_{min} and π^* were obtained by 10-fold cross-validation of three performance measures: accuracy, sensitivity and specificity. These performance measures are reported in Table 1. Using MRI data, the accuracy index is between 69.7% (for pMCI vs sMCI) and 86.2% (for AD vs CN). Remarkably, only less than 400 voxels were required to achieve these high classification accuracies. Slightly superior classification performance is achieved when using PET data, and requires from 1178 to 2020 voxels. The accuracy results reported in Table 1 were assessed for the statistical significance using permutation testing. All accuracy measures were found to be highly significant (p-values < 0.001).

Voxels in $S(\lambda^*_{\text{min}}, \pi^*)$ from the AD vs CN comparison using both datasets are visualised in yellow in Figure 1(B). As an illustration, the insets show the whole range of selection probabilities $\text{P}_j(\lambda^*_{min})$ for all voxels, without any thresholding. The selected voxels form

Table 1: Classification performance measures in percentages - accuracy (`acc`), sensitivity (`sen`) and specificity (`spe`) - using LDA based on selected voxels (`vox`).

	MRI				PET			
Comparisons	`vox`	`acc`	`sen`	`spe`	`vox`	`acc`	`sen`	`spe`
AD vs CN	221	86.2	82.8	89.2	2020	87.1	87.3	87.0
pMCI vs CN	386	81.7	76.8	85.3	1178	84.0	80.6	87.0
pMCI vs sMCI	288	69.7	68.5	70.5	1463	70.1	72.6	68.2

connected regions in hippocampus and amygdala in the MRI data and in posterior cingulate gyrus and superior parietal gyrus in the PET data, in agreement with previous findings [1, 4].

4 Discussion and Conclusion

A method based on a combination of penalised regression and data resampling has been proposed here. The approach enables computationally efficient selection of highly discriminative voxels between two groups of medical images. The methodology has been applied to the sparse classification of ADNI images. The classification performance compares favourably to results of state-of-the-art studies (for example [5]) and a recent meta-analysis [3] of classification methods on ADNI data. While our results for AD vs CN classification are comparable to the best results reported in this study, we achieve better results for pMCI vs CN classification and for the clinically most interesting discrimination of pMCI from sMCI subjects. Our findings are fully consistent with patterns of AD atrophy and hypometabolism demonstrated in previous studies [1, 4].

Acknowledgements

Eva Janousova was supported by grants IGA MH CZ NR No. NS9893-4 and NS10347-3. Maria Vounou and Giovanni Montana acknowledge financial support from the EPSRC and GlaxoSmithKline. Imaging data was provided by the Alzheimer's Disease Neuroimaging Initiative (National Institutes of Health Grant U01 AG024904).

References

[1] H. Braak and E. Braak. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica*, 82(4):239–259, 1991.

[2] J.G. Csernansky, L. Wang, J. Swank et al. Preclinical detection of Alzheimer's disease: hippocampal shape and volume predict dementia onset in the elderly. *NeuroImage*, 25 (3):783–792, 2005.

[3] R. Cuingnet, E. Gérardin, J. Tessieras et al. Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *NeuroImage*, 56(2):766–781, 2010.

[4] S.T. DeKosky, M.C. Carrillo, C. Phelps et al. Revision of the criteria for Alzheimer's disease: A symposium. *Alzheimer's & Dementia*, 7:e1–e12, 2011.

[5] S. Klöppel, C. Stonnington, C. Chu et al. Automatic classification of MR scans in Alzheimer's disease. *Brain*, 131(Pt 3):681–689, 2008.

[6] N. Meinshausen and P. Bühlmann. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473, 2010.

[7] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58(1):267–288, 1996.

[8] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67(2):301–320, 2005.