Biomarker discovery for sparse classification of brain images in Alzheimer's disease

Eva Janousova[1,3], Maria Vounou[1], Robin Wolz[2], Katherine R. Gray[2], Daniel Rueckert[2], Giovanni Montana[1], for the Alzheimer's Disease Neuroimaging Initiative[*]

[1] Department of Mathematics, Imperial College London, UK
[2] Department of Computing, Imperial College London, UK
[3] Institute of Biostatistics and Analyses, Masaryk University, Brno, CZ
⟨g.montana@imperial.ac.uk⟩

Abstract

We describe a computationally efficient biomarker discovery approach, based on a combination of penalised regression and a resampling method, for the identification of localised brain regions that are highly discriminative between two groups of brain images. The proposed procedure has been applied for classification of brain images in subjects with Alzheimer's disease and mild cognitive impairment using baseline FDG-PET data and both baseline and longitudinal MRI data. Results of nine independent classification experiments show that the selected biomarkers are consistent with well-known patterns of atrophy, hypometabolism and progression of Alzheimer's disease that have been reported in previous studies. The overall classification performance, which has been assessed for statistical significance, is comparable to related state-of-the-art findings.

1 Introduction

Neuroimaging techniques such as magnetic resonance imaging (MRI) and positron emission tomography (PET) can be used to identify diagnostic and prognostic biomarkers of neurodegenerative diseases. Such biomarkers play a crucial role in the detection of presymptomatic stages of the disease and prediction of its later development. It is well known that Alzheimer's disease (AD) is associated with tissue loss over time [Fox et al., 1996], especially in the hippocampus, hence hippocampal volume [Chupin et al., 2009], shape [Csernansky et al., 2005] or atrophy are often used as biomarkers [Leung et al., 2010]. Extraction of such biomarkers is based on prior knowledge about the disease and its underlying

---

processes. Data-driven biomarkers, however, do not require any a priori hypotheses about which structures are involved in the disease process [Hua et al., 2008], and are extracted using data-driven approaches with the use of whole-brain image data. A recent comparison of ten MRI-based high-dimensional pattern classification methods [Cuingnet et al., 2011] reported good accuracies in distinguishing AD patients from healthy controls using whole-brain approaches. One of the key challenges for data-driven biomarker extraction is the large number of voxels, which leads to problems for many classical feature selection techniques.

In this paper we propose a simple and computationally efficient approach for data-driven biomarker extraction using penalised regression combined with a data resampling scheme. The algorithm selects a small number of spatially localised voxels that are highly discriminative of the disease. This approach was used in [Janousova et al., 2011] for the classification of baseline MR and PET images. Here we further apply the method to longitudinal MRI data, to demonstrate that the approach can be used to monitor disease progression.

2 Methods

2.1 Voxel selection using penalised regression

We assume to have a collection of brain images observed on a random sample of n independent subjects. Each image $x_{i\cdot}$, with $i = 1, \ldots, n$, is represented by a p-dimensional vector of intensities, one intensity measure per voxel, and the data is arranged in a $(n \times p)$ matrix $X = (x_{1\cdot}, \ldots, x_{n\cdot})$. We also assume that the n individuals have been labelled as one of two classes, which we denote by D (diseased) and H (healthy controls), with sample sizes of n_D and n_H, respectively, and $n = n_D + n_H$. 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 assume that the vector $y = (y_1, \ldots, y_n)$ is mean centred and the columns of X, $x_{\cdot j}$, with $j = 1, \ldots, p$, corresponding to the vector of intensities for each voxel, have zero mean and unit variance.

Our aim is to use the n images to identify localised groups of voxels whose intensities are highly discriminative between the two classes. Ideally, a set S of discriminative voxels should have cardinality $|S| << p$. 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 discriminative voxels in S are non-zero, thus yielding a sparse estimate $\hat{\beta}$, and (b) the non-zero coefficients of correlated voxels should be smoothed towards a common value to allow selection of voxels in groups. Both effects are achieved by imposing some constraints on the regression coefficients. Specifically, we use the elastic net penalty [Zou and Hastie, 2005] which consists of two additive penalty terms: a constraint on the l_1 norm of the coefficients, which is known to induce sparse solutions [Tibshirani, 1996], and a constraint on their l_2 norm, which is known to induce a grouping effect on correlated variables [Zou and Hastie, 2005, Hoerl and Kennard, 1970]. Accordingly, the elastic net 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 the scaling factor $\kappa = (1 + \mu)^{-1}$ corrects for the double shrinkage caused by applying both the l_1 and l_2 penalties. The free parameters that need to be tuned are λ, which controls the amount of sparsity (i.e. the number of voxels with non-zero regression coefficients) and μ, which controls the amount of smoothing imposed on the regression coefficients associated with correlated voxels.

To gain further insights into the effects of the penalty terms and regularisation parameters, we re-write the terms to be minimised in Equation (1) as $-2y'X\beta + \beta'\kappa \left(X'X + \mu I\right)\beta + \lambda \left\|\beta\right\|_1$. Minimising this expression with respect to β provides the solution to an equivalent penalised regression problem having only a constraint on the l_1 norm of the coefficients, whereby the covariance matrix $X'X$ has been replaced by $\kappa \left(X'X + \mu I\right)$. Using this parametrisation, it can be noted that the l_2 regularisation parameter μ acts directly on the correlations of the predictors and shrinks those by a factor κ. As noted in [Zou and Hastie, 2005], by setting μ to infinity, the number of free parameters is reduced down to only one, λ, whilst still maintaining the grouping effect. With this choice, the expression to be minimised is $-2y'X\beta + \beta'\beta + \lambda \left\|\beta\right\|_1$, which is equivalent to assuming an orthogonal design, whereby the covariance matrix $X'X$ has been replaced by the identity matrix, with only an l_1 penalty for voxel selection. This parametrisation is convenient in that it 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,

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

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

2.2 Estimation of voxel importance through selection probabilities

The regularisation parameter λ in Equation (2) controls the amount of sparsity, and therefore determines the 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, sparser solutions are obtained, and 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 minimising a prediction error does not necessarily lead to the true, underlying sparsity pattern to be discovered. Moreover, 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 truly important for prediction and are more robust against sampling errors, we adopt a data resampling scheme that has been specifically proposed for sparse predictive modelling [Meinshausen and Bühlmann, 2010]. 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, and fit the penalised regression model on each random sub-sample. For each sub-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 coeffi-

cients by using an indicator variable $v_j^{(b)}(\lambda)$ which is equal to 1 if the coefficient corresponding to voxel $x_{\cdot j}$ is non-zero, or 0 otherwise. Using all B sub-samples, a measure of voxel importance is finally computed by estimating the selection probabilities

$$P_j(\lambda) = \frac{1}{B} \sum_{b=1}^{B} v_j^{(b)}(\lambda) \quad j = 1, \dots, p \tag{3}$$

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 the zero vector estimate $\hat{\beta} = \mathbf{0}$, with no voxels selected in the set S. The final set of voxels to be included in S is obtained by choosing a lower bound λ_{min} and 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   Experimental Results

3.1   Subjects

The penalised regression approach was applied to MRI and [^{18}F]-fluorodeoxyglucose (FDG)-PET data obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database[1]. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research, approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be followed for 2 years.

We used T1-weighted 1.5 T baseline MRI scans from all available 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 images of all subjects were aligned with the MNI152 brain template [Mazziotta et al., 1995] using a coarse non-rigid registration regularised by a 10*mm* B-spline control-point grid [Rueckert et al., 1999]. Brain extraction was performed based on automated tissue classification using SPM5[2]. Image intensities were normalised to the template using linear regression prior to performing Gaussian smoothing with a 4*mm* full width at half-maximum Gaussian kernel. The small Gaussian kernel was used to enable detection of discriminative regions in smaller brain structures [Honea et al., 2005]. Follow-up MRI images acquired 24 months from baseline were available for 510 subjects: 105 AD, 117 pMCI, 123 sMCI, and 165 CN. All follow-up scans were aligned with their baseline scans using a fine 2.5*mm* B-spline control-point spacing non-rigid registration [Rueckert et al., 1999]. The Jacobian determinants extracted from the resulting deformation fields represent intra-subject development (expansion/contraction) on a voxel basis. The Jacobian maps were then transformed to the reference space using the transformations computed for the baseline scans. PET images were available for 287 subjects: 71 AD, 62 pMCI, 85 sMCI, and 69 CN. Each

---

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

PET image was converted to a 30-minute static and affinely aligned with the corresponding native-space MRI. An affine transformation was preferred over a rigid one because it can account for any scaling or voxel size errors which remain after phantom correction of the MRI [Clarkson et al., 2009]. The non-linear transformation parameters estimated to map the baseline 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 using a cluster of relatively preserved regions derived from an independent dataset [Yakushev et al., 2009], and re-sampled to the higher resolution of the MRI. $1,650,857$ voxel intensities in the PET images and baseline MRI scans and the same number of Jacobian determinants representing longitudinal changes in MRI images were used to perform voxel selection with the proposed method after correcting for age and gender using a linear regression model.

3.2 Classification results

We report on nine independent classification experiments, whereby we compare two groups in each experiment: AD vs CN, pMCI vs CN, and pMCI vs sMCI, using PET data and baseline and longitudinal MRI 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. 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 tuning, therefore there are only two parameters λ_{min} and π to be optimised which we collect in a parameter vector $\boldsymbol{\theta} = \{\lambda_{min}, \pi\}$.
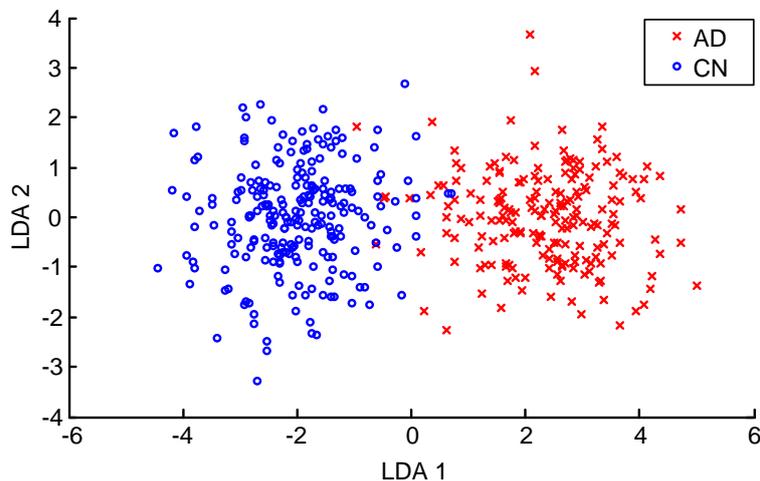
Figure 1: Two-dimensional projections of training data points observed on the selected voxels (baseline MRI data) show that AD and CN subjects are linearly separable. LDA1 and LDA2 are the first and the second latent factor extracted using LDA, respectively.

The optimal parameter vector $\boldsymbol{\theta}^*$ was obtained by 10-fold cross-validation of three performance measures: accuracy, sensitivity and specificity. The cross-validated performance measures are reported in Table 1. Using baseline MRI data, the accuracy index in LDA clas-

Table 1: Number of selected voxels (`vox`) and 10-fold cross-validated classification performance measures in percentage - accuracy (`acc`), sensitivity (`sen`), and specificity (`spe`) - using LDA and SVM with Gaussian kernel.

			LDA			SVM		
Data	Comparisons	vox	acc	sen	spe	acc	sen	spe
Baseline MRI	(B1) AD vs CN	221	86.2	82.8	89.2	87.9	84.8	90.5
	(B2) pMCI vs CN	386	81.7	76.8	85.3	83.2	76.2	88.3
	(B3) pMCI vs sMCI	288	69.7	68.5	70.5	70.4	63.7	75.1
Longit MRI	(L1) AD vs CN	11 394	87.0	81.0	90.9	90.3	87.5	92.1
	(L2) pMCI vs CN	12 664	83.3	79.5	86.1	86.9	81.2	90.9
	(L3) pMCI vs sMCI	10 593	71.3	68.4	74.0	82.1	81.5	82.9
PET	(P1) AD vs CN	2 020	87.1	87.3	87.0	87.9	88.7	87.0
	(P2) pMCI vs CN	1 178	84.0	80.6	87.0	84.0	82.3	85.5
	(P3) pMCI vs sMCI	1 463	70.1	72.6	68.2	72.1	61.3	80.0

sification experiments is between 69.7% (for the pMCI vs sMCI group) and 86.2% (for the AD vs CN group). Remarkably, only between 221 and 386 high-probability voxels were able to achieve those classification accuracies. As expected, superior classification performance is achieved when using longitudinal data, and requires less than 13k voxels in all cases. The accuracy of classification of PET data using LDA classifier is from 70.1% to 87.1% using less than 3k selected voxels. In the same table we also report on the classification results obtained using a Support Vector Machine (SVM) classifier with Gaussian kernel [Smola and Schölkopf, 2004]. The cross-validated performance measures obtained using the non-linear SVM classifier are slightly higher than those obtained using LDA.

In order to assess the statistical significance of the accuracy index reported in Table 1, we carried out non-parametric inference using permutation testing. Holding the optimal θ^* constant, we randomly permuted the elements of the response vector y, and repeated this procedure K times. Each permutation gave a new response vector $y^{(k)}$, with $k = 1, \ldots, K$. For each k, we applied the LDA and SVM classifier to the data $\{X, y^{(k)}\}$, where X is the matrix containing the selected voxels, and produced the corresponding 10-fold cross-validated accuracy measures. This procedure approximates the sampling distribution of the accuracy index under the null hypothesis of no association between the voxel intensities in $S(\lambda_{min}^*, \pi^*)$ and the response, and an empirical p-value can be easily computed. Using $K = 1000$ permuted data sets, all the accuracy results in Table 1 were all found to be highly statistically significant (p-values < 0.001).

Figure 2 shows coronal slices with voxels in $S(\lambda_{min}^*, \pi^*)$, in yellow, for all comparisons in Table 1. As an illustration, the insets show the whole range of selection probabilities $P_j(\lambda_{min}^*)$ for all voxels, without any thresholding. It can be noted how the l_2 penalty enforces group selection. In the AD vs CN, pMCI vs CN and also pMCI vs sMCI comparisons, using baseline MRI images, the selected voxels form connected regions in both hippocampus and amygdala. In the case of longitudinal MRI data, the most discriminative voxels are clustered in the hippocampus and lateral ventricles. In experiments using PET data, the selected voxels form clusters in posterior cingulate gyrus and superior parietal gyrus.
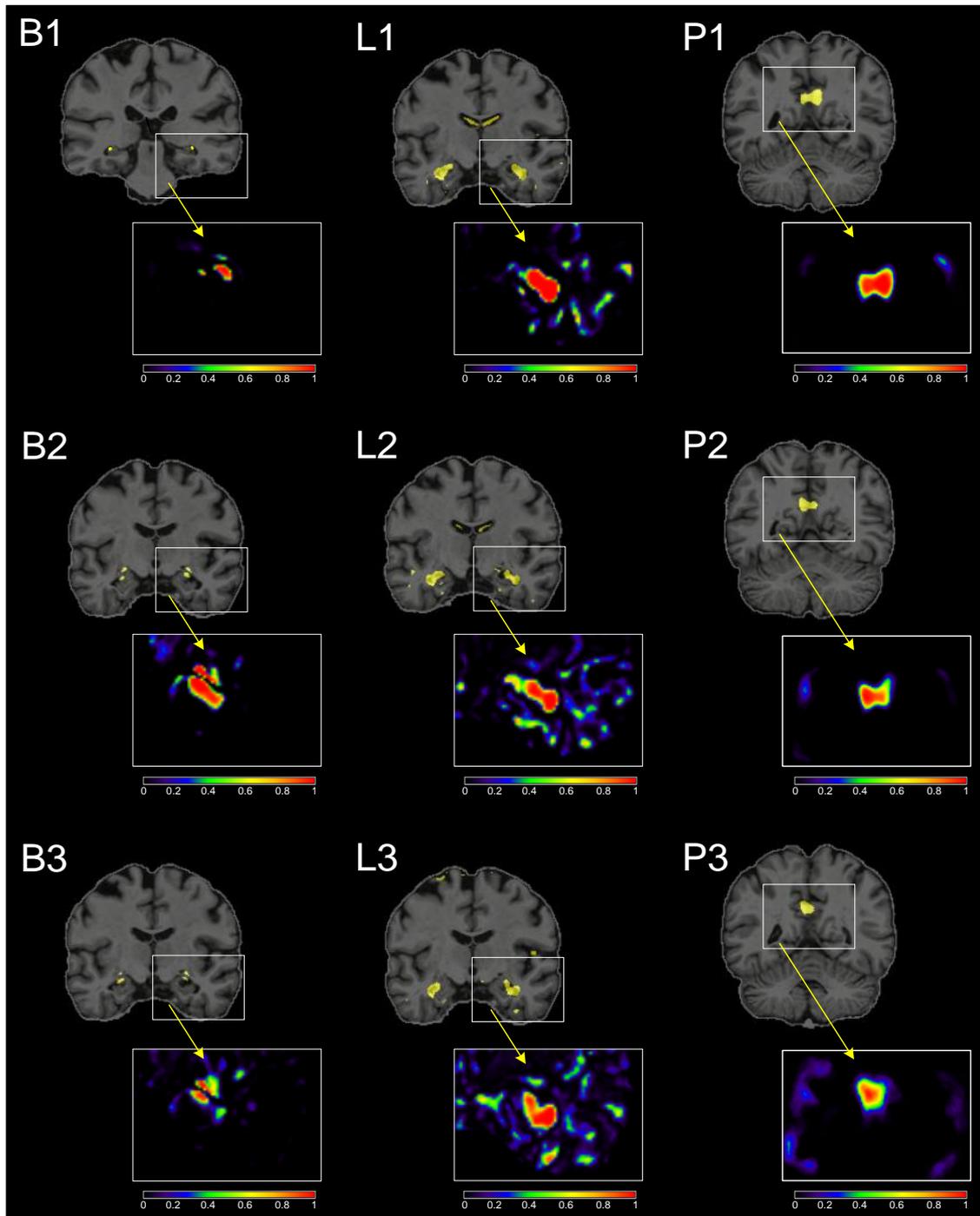
Figure 2: Coronal slices showing the voxels selected by penalised regression (in yellow) in the whole brain; the selection probabilities $P_j(\lambda)$ of all voxels are shown in the insets.

4   Discussion and Conclusion

We have presented a penalised regression approach using the elastic net penalty, for the selection of voxel intensities that are highly discriminative between two groups of MR images. A measure of voxel importance is obtained by employing a data resampling procedure. For a two-class classification problem, it is known that treating the binary response as a continuous variable and minimising the least squares error is equivalent (up to a scaling factor) to the solution obtained by linear discriminant analysis (LDA) [Duda et al., 2001]. In this respect, our method provides a sparse LDA solution. The resulting estimation algorithm has very low computational complexity which makes it suitable for voxel-wise whole-brain studies.

The methodology has been applied to the sparse classification of ADNI images. The classification performance measures obtained using LDA and SVM with Gaussian kernel are comparable to findings documented in the literature. For instance, for the AD vs control group, typical classification accuracy has been reported to vary from 85% to 95% [Fan et al., 2008, Klöppel et al., 2008, Batmanghelich et al., 2009], whereas for the pMCI vs control group comparison the accuracy varies between 70% and 81.8% [Fan et al., 2008, Batmanghelich et al., 2009] and for the pMCI vs sMCI between 70% and 81.5% [Misra et al., 2009]. Our results compare favourably to a recent meta-analysis [Cuingnet et al., 2011] 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 progressive MCI vs CN classification and for the clinically most interesting discrimination of progressive from stable MCI subjects (sMCI vs pMCI).

The estimated performance measures obtained using the non-linear SVM classifier are only slightly higher than those obtained using LDA. This observation supports the suitability of the linear classifier for the assessment of the discriminative power of the selected sets of voxels in our experiments. If SVM with Gaussian kernel was used for the selection of the optimal parameter vector θ^*, there would be two additional parameters to be optimised ,σ and C, which are the kernel width and the regularisation parameter of the SVM classifier, respectively, which would introduce much more computational complexity in our experiments.

The selected voxels form connected regions in hippocampus and amygdala in the baseline data and in posterior cingulate gyrus and superior parietal gyrus in the PET data. In the longitudinal data, the selected voxels reflect hippocampus atrophy and ventricular enlargement. These findings are fully consistent with patterns of AD atrophy and hypometabolism demonstrated in previous neuropathological and morphological studies [Braak et al., 1999, Cuingnet et al., 2011, DeKosky et al., 2011]. Both hippocampus and amygdala are among the first structures in the brain which are affected by Alzheimer's disease [Braak and Braak, 1991]. The selected voxels in the longitudinal data are also highly meaningful because hippocampal atrophy and ventricular enlargement over time has been shown to correlate with disease progression [Jack et al., 2004]. The results also show that the patterns of brain atrophy and hypometabolism in baseline time-point and brain morphological changes during follow-up are the same in both AD patients and MCI subjects.

Longitudinal FDG-PET data were not used in classification experiments as there are only 221 subjects with follow-up FDG-PET scans available in the ADNI database and the number of individuals in AD, sMCI, pMCI and CN groups is too small for the presented method. Ongoing work on the method is focused on modification of the approach for multi-class

classification and combination of image modalities.

Acknowledgements

References

N. Batmanghelich, B. Taskar, and Ch. Davatzikos. A general and unifying framework for feature construction, in image-based pattern classification. *Lecture Notes in Computer Science*, 5636:423–34, January 2009.

E. Braak, K. Griffing, K. Arai, J. Bohl, H. Bratzke, and H. Braak. Neuropathology of alzheimer's disease: what is new since a. alzheimer? *European Archives of Psychiatry and Clinical Neuroscience*, 249(9):14–22, 1999.

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

M. Chupin, E. Gérardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehéricy, H. Benali, L. Garnero, O. Colliot, and Alzheimer's Disease Neuroimaging Initiative. Fully automatic hippocampus segmentation and classification in alzheimer's disease and mild cognitive impairment applied on data from adni. *Hippocampus*, 19(6):579–587, Jun 2009.

M.J. Clarkson, S. Ourselin, C. Nielsen, K.K. Leung, J. Barnes, J.L. Whitwell, J.L. Gunter, D.L.G. Hill, M.W. Weiner, C.R. Jack, N.C. Fox, and the Alzheimer's Disease Neuroimag-

ing Initiative. Comparison of phantom and registration scaling corrections using the ADNI cohort. *NeuroImage*, 47:1506–1513, 2009.

J. G. Csernansky, L. Wang, J. Swank, J. P. Miller, M. Gado, D. McKeel, M. I. Miller, and J. C. Morris. Preclinical detection of alzheimer's disease: hippocampal shape and volume predict dementia onset in the elderly. *NeuroImage*, 25(3):783–792, Apr 2005.

R. Cuingnet, E. Gérardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot. Automatic classification of patients with alzheimer's disease from structural mri: A comparison of ten methods using the adni database. *NeuroImage*, 56:766–781, June 2011.

S.T. DeKosky, M.C. Carrillo, C. Phelps, D. Knopman, R. C. Petersen, R. Frank, D. Schenk, D. Masterman., E. R. Siemers, J. M. Cedarbaum, M. Gold, D. S. Miller, B. H. Morimoto, D. S. Khachaturian, and R. C. Mohs. Revision of the criteria for alzheimer's disease: A symposium. *Alzheimer's & Dementia*, 7:e1–e12, 2011.

R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern classification*, volume 2. John Wiley & Sons, 2001.

Y. Fan, N. Batmanghelich, Ch.M. Clark, and Ch. Davatzikos. Spatial patterns of brain atrophy in mci patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *NeuroImage*, 39(4):1731–43, February 2008.

N. C. Fox, P. A. Freeborough, and M. N. Rossor. Visualisation and quantification of rates of atrophy in alzheimer's disease. *Lancet*, 348(9020):94–97, Jul 1996.

A.E. Hoerl and R.W. Kennard. Ridge regression: Applications to nonorthogonal problems. *Technometrics*, 12(1):69–82, 1970.

R. Honea, T.J. Crow, D. Passingham, and C.E. Mackay. Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. *American Journal of Psychiatry*, 162(12):2233–2245, Dec 2005.

X. Hua, A.D. Leow, N. Parikshak, S. Lee, M.-Ch. Chiang, A.W. Toga, C.R. Jack, M.W. Weiner, P.M. Thompson, and Alzheimer's Disease Neuroimaging Initiative. Tensor-based morphometry as a neuroimaging biomarker for alzheimer's disease: an mri study of 676 ad, mci, and normal subjects. *NeuroImage*, 43(3):458–469, Nov 2008.

C. R. Jack, M. M. Shiung, J. L. Gunter, P. C. O'Brien, S. D. Weigand, D. S. Knopman, B. F. Boeve, R. J. Ivnik, G. E. Smith, R. H. Cha, E. G. Tangalos, and R. C. Petersen. Comparison of different mri brain atrophy rate measures with clinical disease progression in ad. *Neurology*, 62(4):591–600, Feb 2004.

E. Janousova, M. Vounou, R. Wolz, K. Gray, D. Rueckert, and G. Montana. Fast brain-wide search of highly discriminative regions in medical images: an application to alzheimer's disease. In *Proceedings of Medical Image Understanding and Analysis 2011*, 2011.

S. Klöppel, C.M. Stonnington, C. Chu, B. Draganski, R.I. Scahill, J.D. Rohrer, N.C. Fox, C.R. Jack, J. Ashburner, and R.S.J. Frackowiak. Automatic classification of mr scans in alzheimer's disease. *Brain*, 131:681–9, March 2008. ISSN 1460-2156.

K.K. Leung, J. Barnes, G.R. Ridgway, J.W. Bartlett, M.J. Clarkson, K. Macdonald, N. Schuff, N.C. Fox, S. Ourselin, and Alzheimer's Disease Neuroimaging Initiative. Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and alzheimer's disease. *NeuroImage*, 51(4):1345–1359, Jul 2010.

J. Mazziotta, A. Toga, A. Evans, P. Fox, and J. Lancaster. A probabilistic atlas of the human brain: Theory and rationale for its development. *NeuroImage*, 2(2):89–101, 1995.

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

Ch. Misra, Y. Fan, and Ch. Davatzikos. Baseline and longitudinal patterns of brain atrophy in mci patients, and their use in prediction of short-term conversion to ad: results from adni. *NeuroImage*, 44(4):1415–22, February 2009.

D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE Transactions on Medical Imaging*, 18(8):712–21, August 1999.

A.J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14:199–222, August 2004.

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

I. Yakushev, A. Hammers, A. Fellgiebel, I. Schmidtmann, A. Scheurich, H.G. Buchholz, J. Peters, P. Bartenstein, K. Lieb, and M. Schreckenberger. Spm-based count normalization provides excellent discrimination of mild alzheimer's disease and amnestic mild cognitive impairment from healthy aging. *NeuroImage*, 44:43–50, 2009.

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.