

Research Article

Evaluating the Impact of Different Factors on Voxel-Based Classification Methods of ADNI Structural MRI Brain Images*

Ramon Casanova,¹ Joseph A. Maldjian,² and Mark A. Espeland¹ for the Alzheimer's Disease Neuroimaging Initiative

¹Department of Biostatistical Sciences, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA

²Department of Radiology, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA

Address correspondence to Ramon Casanova, casanova@wfubmc.edu

Received 21 January 2011; Revised 31 July 2011; Accepted 2 August 2011

Abstract In this work we introduce the use of penalized logistic regression (PLR) to the problem of classification of MRI images and automatic detection of Alzheimer's disease. Classification of sMRI is approached as a large scale regularization problem which uses voxels as input features. We evaluate how differences in sMRI pre-processing steps such as smoothing, normalization, and template selection affect the performance of high-dimensional classification methods. In addition, we compared the relative performance of PLR to a different approach based on support vector machines. To study these questions we used data from the Alzheimer Disease Neuroimaging Initiative (ADNI). The ADNI project follows a protocol consisting of acquisition of two images in each session, image correction steps and further evaluation by experts to obtain the optimized images. We evaluated here the impact of this optimization process on the performance of high-dimensional machine learning techniques.

Keywords machine learning; Alzheimer disease; regularization; GLMNET; ADNI; SVM; logistic regression; elastic net

1 Introduction

The use of machine learning techniques is becoming very popular in the neuroimaging community (Mur et al. [37]; Pereira et al. [40]). These methods probe data in a multivariate fashion, overcoming shortcomings of other more conventional approaches (Davatzikos [16]). An especially challenging area of application is the classification and regression of structural MRI brain images. Many researchers

use dimension reduction measures such as the use of input features generated from predefined regions of interest (ROI) in the brain (Magnin et al. [35]) or principal component analysis (PCA) (Teipel et al. [45]) to reduce the size of the classification problem. Others resort to complicated procedures consisting of several steps based on downsampling of the images (Vemuri et al. [50]) or image processing methods (Fan et al. [19]). Recently, Ashburner, Kloppel, Cuingnet and colleagues have shown that voxel-based classification using kernel methods is not only feasible but produces very good results when classifying sMRI brain images (Ashburner [1]; Chu [11]; Cuingnet et al. [12]; Cuingnet et al. [14]; Kloppel et al. [31]). In a recent comparison of several of the most successful methods, a linear SVM method (Kloppel et al. [31]) was one of the best performers (Cuingnet et al. [13]). This comparison was based on the use of SPM5 normalization tools and a more recent high-dimensional normalization method based on diffeomorphic transformations called DARTEL (Ashburner [1]). Here, we evaluate the influence of smoothing, normalization, and template choice on the performance of two classification methods: penalized logistic regression (PLR) and linear SVM (hard and soft margin) when combined with a high-dimensional image warping technique called symmetric normalization (SyN) (Avants et al. [3]) as implemented in the Advance Normalization Tools (ANTS) software package. Penalized logistic regression (PLR) has been used before in genetics to analyze microarray and sequence data (Liu et al. [34]; Park and Hastie [39]; Shevade and Keerthi [44]; Zhu and Hastie [53]) and in the context of neuroimaging applications PLR has been used before to analyze fMRI (Ryali et al. [43]; Yamashita et al. [52]) and regional volumes of sMRI (Casanova et al. [6,7]) data. Here we use PLR to solve classification problems of very large size that result when voxels from sMRI images are used as input features. To our best knowledge we have been the first to use PLR for AD automatic detection (Casanova et al. [8, 9]) via large-scale regularization. Since we use voxels as

* Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.ucla.edu/>). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. ADNI investigators include (complete listing available at <http://adni.loni.ucla.edu/>).

input features, the sizes of the classification problems we solve here ($\approx 10^6$) are much larger than those reported in the fMRI literature (usually $\approx 10^4$) what makes this problem very challenging. Results related to SyN performance in the context of high-dimensional classification have not been reported before and neither SVM has been compared to penalized logistic regression. The issue of smoothing is an interesting one. In the investigation by Kloppel and colleagues, smoothing was not employed, suggesting this resulted in the best performance. However, data smoothing is a common practice for other high-dimensional classification methods (Fan et al. [19]; Vemuri et al. [50]) and for univariate analyses (Good et al. [24]).

To study these questions, we used data obtained from the Alzheimer Disease Neuroimaging Initiative (ADNI) (Mueller et al. [36]). The ADNI project follows a protocol consisting of acquisition of two images in each session, with selection of the best image set by expert reviewers (optimized images) for further pre-processing and analysis (Jack et al. [28]). We made an effort here to evaluate and quantify the impact of this optimization process on the performance of high-dimensional machine learning techniques.

2 Material and methods

2.1 ADNI database

Data used in the preparation of this work were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and nonprofit organizations as a \$60 million, 5-year public private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. 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.

2.2 MRI scans

We used baseline 1.5T T1-weighted MRI data, as described in the ADNI acquisition protocol (Jack et al. [28]), from 49 subjects with AD and 49 cognitively normal controls (CN). The selected controls did not convert to MCI across the follow-up period of 36 months. The average age and baseline MMSE score was 76 and 29.9 for the control group, and 75 and 23.6 for the AD group, respectively. The two groups were matched approximately by sex (AD—24 m, 25 f and CN—25 f, 24 m). The ADNI protocol acquires 2 sets of structural data at each visit. These are rated for image quality and artifacts by ADNI investigators (Jack et al. [28]). To enhance standardization across sites and platforms, the best quality set undergoes additional pre-processing, including correction for gradient nonlinearity (Jovicich et al. [29]) and correction for intensity nonuniformity (Narayana et al. [38]). These optimally pre-processed images were downloaded from the ADNI database and used for subsequent analysis in this study.

The raw scans had not undergone any corrections for nonuniformity or nonlinearity, and were not selected on the basis of image quality. The raw scans were simply chosen as the first of the 2 repeat T1 scans available in the ADNI database per subject. Both the raw (nonoptimized T1 data) and optimized T1 data were used in our analyses.

2.3 Normalization algorithm

We evaluated the performance of one normalization algorithm called symmetric diffeomorphic registration. Symmetric diffeomorphic registration (SyN) uses diffeomorphisms (differentiable and invertible maps with differentiable inverse) to capture both large deformations and small shape changes (Avants et al. [3]). In the largest evaluation of nonlinear brain registration algorithms to date, SyN was found to be a top-ranking performer, providing among the best results according to overlap and distance measures and delivering the most consistently high accuracy across subjects and label sets (Klein et al. [30]). The SyN normalization procedures have been implemented in the freely available ANTS software toolbox. A separate pipedream toolbox is also available, which scripts the procedures for implementation on grid computing systems. We created a series of in-house MATLAB wrappers for accessing the ANTS and pipedream programs allowing user-tunable parameter modifications and work-flow definition with run-time batch script generation for implementation on the Sun Grid Engine.

2.4 Templates

For ANTS, a variety of templates were used as targets for the normalization procedure including the high-resolution 1 mm

ICBM individual subject template (<http://www.loni.ucla.edu/ICBM/>) and three study-specific custom templates. The custom templates included a template consisting of 12 randomly selected normal subjects using the raw T1 data (ADNI12-Raw); a template constructed from the same 12 subjects using the optimized T1 data (ADNI12-Opt); and a template constructed from 49 AD, 49 CN, and 42 MCI subjects using the raw T1 data (ADNI140-Raw). The AD and CN subjects in the ADNI140 template were also part of the subject sample for the analyses.

2.4.1 Custom template construction

The 3 custom templates (ADNI12-Raw, ADNI12-Opt, and ADNI140-Raw) were built using a diffeomorphic shape and intensity averaging technique (Avants and Gee [2]; Avants et al. [3]). Parameters for the template construction procedure included a four-level Gaussian pyramid as the multi-resolution strategy and the cross-correlation similarity metric, with 200 maximum iterations. This yielded an unbiased average shape and appearance template, as well as the set of diffeomorphisms and inverse diffeomorphisms that map from template to each individual. The resulting study template was normalized to the ICBM T1 label atlas (<http://www.loni.ucla.edu/ICBM/>) using SYN (Avants and Gee [2]; Avants et al. [3]) in order to allow automated label information to be obtained from each subject. The ICBM atlas distribution includes a high-resolution T1 image, segmented labels, and a brain mask. A 2-step normalization procedure was used to generate a highly accurate skull-stripped version of the custom templates normalized to the ICBM atlas. Following the first normalization step to the ICBM template including the skull, a skull stripped version of each custom template was generated using the inverse transformation and the ICBM brain mask applied to the custom template. A second SYN normalization was then performed using the skull-stripped custom template and the skull-stripped ICBM atlas. For each custom template, these procedures produced a full custom template (with scalp), a scalp-stripped version of the custom template, a custom template brain mask, and fully invertible parameters defining a transform to custom template space to ICBM atlas space.

ANTS spatial normalization. The native space raw and optimal T1-weighted images obtained from the ADNI database for each subject were used for image analysis. The optimized images had already undergone gradient field inhomogeneity correction, and N3 bias correction. For the raw T1 data, we added an N3 bias correction step prior to normalization. Each subject T1 image was normalized to each of the custom study templates using SYN (Avants and Gee [2]; Avants et al. [3]) in a 2-step process. Each subject image was normalized to the full custom template. The inverse transformation and the custom template brain mask

were then used to generate a skull-stripped version of the subject T1 image in native space. A second SYN normalization was then performed between the skull-stripped native T1 image and the skull-stripped custom template. The previously computed custom template to ICBM normalization parameters were combined with the native space to custom template parameters, generating a set of transformation parameters to bring native space T1 images into ICBM space with a single resampling step. The nonlinear transformations from the SYN procedure provide deformation tensor fields describing the voxel-based shape changes from the template to each subject's brain. The Jacobian determinants of these deformation fields indicate the fractional volume expansion and contraction at each voxel required to match the template. These maps can be used directly as in tensor-based morphometry to determine population differences, or they can be combined with the segmentation maps to generate maps of tissue-specific volume change (modulated maps). The native space grey matter segmentation maps generated from the SPM8 new segment procedure were brought into template space using the combined SYN transform. The Jacobian maps were then multiplied by the respective GM segmentation maps to limit analysis to grey matter volume changes. For simplicity, all the machine learning analyses presented later are restricted to GM tissue only.

2.5 Normalization strategies

Eight normalization strategies were compared utilizing either the raw or optimized T1 data as input. These included the ANTS normalization procedure using the 4 templates (ADNI12-Raw, ADNI12-Opt, ADNI140-Raw, and ICBM) and the raw or optimized T1 data as input. The modulated grey matter maps produced from each of these processing methods were used as the input features to the machine learning algorithms. In order to limit the machine learning analysis to the same voxels across the subjects, grey matter segmentation masks were generated for all of the templates using the SPM new segment tool. The grey matter template segmentations were thresholded at 0.5 and binarized to generate the final mask used across subjects in each of the machine learning analyses.

2.6 Degree of smoothing

We studied the influence of isotropic Gaussian smoothing on the performance of voxel-based classification of sMRI across different normalization strategies. We analyzed three different degrees of smoothness: (1) No-smoothing; (2) GM images smoothed with a Gaussian filter with full width half maximum (FWHM) of 4 mm; and (3) GM images smoothed with a Gaussian filter with FWHM = 8 mm.

2.7 Penalized logistic regression

Logistic regression is a common choice when the response variable Y is binary. It models the class-conditional

probabilities through a linear function of predictors:

$$\log \left(\frac{\Pr(Y = 1/x)}{\Pr(Y = 2/x)} \right) = \beta_0 + x^T \beta. \quad (1)$$

In problems where the number of predictors (voxels) is much greater than the number of examples (ADNI participants' GM images in our case), it is necessary to apply regularization (Tikhonov and Arsenin [46]). The regularization is performed by adding a penalty term to the estimation criterion having the effect of regulating the complexity of the model by introducing constraints on the estimated solution. In this work, we evaluate the performance of PLR, as implemented in the GLMNET library (Friedman et al. [21,22]), when applied to classification of sMRI images. To estimate PLR, the GLMNET library solves the problem defined by

$$\begin{aligned} \min_{\beta_0, \beta \in R^{p+1}} & -C(\beta_0, \beta, x_i, y_i) + \lambda P(\beta), \\ C(\beta_0, \beta) &= \frac{1}{N} \sum_{i=1}^N y_i (\beta_0 + x_i^T \beta) \\ & \quad - \log(1 + e^{(\beta_0 + x_i^T \beta)}), \\ P(\beta) &= \sum_{j=1}^p \left[\frac{(1-\alpha)}{2} \beta_j^2 + \alpha |\beta_j| \right], \end{aligned} \quad (2)$$

where N is the number of examples (GM images corresponding to 98 subjects in our case), $x_i \in R^p$ is the i th example (GM image corresponding to the i th subject) or feature vector containing the GM voxels entering the analysis, $p \approx 7.4 \times 10^5 - 9.1 \times 10^5$ is the number of voxels entering the analysis depending on the normalization template of choice, $y_i \in \{1, 2\}$ is the i th label (1 for CN and 2 for AD participants), β_0, β are the parameters of the model, and λ is the regularization parameter.

The general form defined by equation (2) contains the so-called elastic net penalty (Zou and Hastie [54]) which is a linear combination of the L_1 and L_2 penalties that produces a trade-off between the properties of both penalties. We set the value of α to zero to enforce the L_2 penalty, which similarly to ridge regression (Hoerl [27]) has the effect of shrinking the coefficients of the equation (2) final solution. The GLMNET library solves the problem described above by using a very efficient optimization technique called coordinate descent (Friedman et al. [20]). The basic idea of the method is to solve a sequence of one-dimensional optimization problems by fixing all variables except one. The process is iterated until convergence. This class of methods has been independently developed by different groups (Daubechies et al. [15]; Fu [23]; Krishnapuram and Hartemink [32]) and its convergence has been proven for classes of convex optimization problems (Tseng [47]). The optimal value of the regularization parameter is estimated using cross-validation as explained later.

2.8 Support vector machine

This is one of the more common techniques used in different ways to analyze sMRI data (Ashburner [1]; Davatzikos et al. [17]; Kloppel et al. [31]; Lao et al. [33]; Vemuri et al. [50]) which is the one originally proposed by the developers of this method (Boser et al. [5]; Vapnik [48]). There are many sources describing in detail the principles behind SVM (Bishop [4]; Hastie et al. [26]) and we refer the reader to those while here we briefly describe our implementation. Kloppel and colleagues used a hard-margin linear SVM (HM-SVM) (Chu [11]; Kloppel et al. [31]), while more recently Cuingnet and colleagues implemented a soft-margin linear SVM (SM-SVM) (Cuingnet et al. [13]). We implemented and evaluated versions of both algorithms. Linear SVMs search for the hyperplane that separates two classes with maximum margin. When the two classes are separable, the hard-margin SVM (HM-SVM) formulation can be applied. The soft-margin SVM allows dealing with overlapping between classes by permitting some of the training examples to be misclassified. Its classical formulation (primal) is

$$\min_{\beta, b} \frac{1}{2} \|\beta\|_2^2 + C \sum_{i=1}^N \xi_i \quad (3)$$

subject to

$$y_i (\beta^T x_i + b) \geq 1 - \xi_i, \quad (4)$$

$$\xi_i \geq 0, \quad i = 1, \dots, N. \quad (5)$$

In terms of the regularization framework, the formulation of the soft-margin linear SVM can be recast as

$$\min_{\beta, b} \sum_{i=1}^N \max(0, 1 - y_i (\beta^T x_i + b)) + \lambda \sum_{j=1}^p \beta_j^2, \quad (6)$$

where $\lambda = (2C)^{-1}$ plays the role of regularization parameter (Bishop [4]). As evident in equation (6), the soft-margin linear SVM, as originally developed, is an example of L_2 penalized regularization technique. We will use the same method here, but note that other variations of SVMs using different types of penalties have been recently proposed (Fan et al. [18]; Wang et al. [51]).

To solve the problem posed by equations (3), (4), (5), or (6), we follow a similar methodology as described by (Chu [11]) using the kernel approach based on the LIBSVM library (Chih-Chung et al. [10]). Images after normalization are vectorized and treated as examples. A linear kernel matrix is generated by computing the inner products across all examples. This is provided to the LIBSVM library as a pre-computed kernel. For the SM-SVM, the C parameter was tuned using cross-validation as explained below. To produce the HM-SVM, we fixed a high value of C in this case 10^4 . The map of weights in the voxel space can be

obtained as $\beta = \sum_{i=1}^{nSV} y_i a_i x_i$, where a_i and nSV are the coefficients and the number of support vectors, respectively, but we did not generate maps in this work.

2.9 Cross-validation procedure

For PLR and SM-SVM methods, to achieve the two goals of evaluating their generalization capabilities and estimating, the optimal values of the regularization parameters, we combined a scheme of two nested cross-validations (CV) with grid search. This is done to avoid upward bias in the estimates of accuracies produced by the CV procedure (Chu [11]; Guyon and Elisseeff [25]; Ryalı et al. [42]; Varma and Simon [49]). We implemented an external K_1 -fold CV where at each step we leave one fold for testing and the $K_1 - 1$ remaining for training and validation. These last two procedures are implemented by using a nested K_2 -fold CV. We divide the K_1-1 folds into K_2 folds and we leave one fold for validation and $K_2 - 1$ for training combined with a grid search to determine the optimal parameters. The grid we used in our analyses was $\lambda = \{0.5, 1, 5, 10, 11, 12.98, 99, 100, 200, 500, 1000\}$ for PLR and $C = \{2^{-6}, 2^{-4}, 0, 2^{12}, 2^{14}\}$ for the SM-SVM. At each grid point the classifier was trained using the training data and its performance was assessed using the fold left for validation by estimating the classification accuracy. We selected the regularization parameters that produce maximum average accuracy across the K_2 folds of the internal CV procedure. The classifier was then retrained using the whole data in the $K_1 - 1$ folds left for training and validation and the selected optimal regularization parameters. The classifier's generalization capability was then evaluated by computing the classification accuracy using the fold originally left for testing in the external CV. This was repeated K_1 times and the average classification accuracy is computed. Finally, the classifier weights were computed using the whole data set and the average values of the selected regularization parameters across the K_1 folds. In our analyses we used $K_1 = 10$ and $K_2 = 10$. In the case of the HM-SVM because there is no tuning of regularization parameters we used a conventional 10-fold CV to evaluate the classifier's generalization capability.

3 Evaluation

We computed overall classification accuracy, sensitivity and specificity to evaluate classifier performance:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}}, \quad (7)$$

$$\text{SEN} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (8)$$

$$\text{SPE} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (9)$$

where TP are AD patients correctly identified as AD, TN are controls correctly classified as controls, FN are AD patients incorrectly identified as controls and FP are controls incorrectly identified as AD. To study the variation due to different CV partitions we repeated the computation 10 times and reported the mean and standard deviation of the three metrics described above. For simplicity all our analyses were restricted to grey matter only (GM). Differences in mean levels of accuracy, sensitivity, and specificity across these repetitions were assessed using analyses of variance.

4 Results

In Table 1 we present results when the voxel-based classification analyses (CN vs. AD) were performed using non-smoothed data and the three methods described above. Additionally, in Table 2 the results of analyses of variance of summary measures across 10 trials are presented. The first observation is that when the study customized template generated using optimized data (ADNI12-Opt) was used, all classification methods produced lower values of the metrics (when compared to other templates results) both when using raw or optimized sMRI data. Something similar but in a much lesser degree occurred when the standard ICBM template was employed with the optimized data or SVMs.

The use of raw or optimized data did not originate large differences on classification performance when the study customized templates based on raw data (ADNI12-Raw and ADNI140-Raw) were employed for the analyses. Finally, across all the four templates explored in this work, PLR performance was significantly better than the two SVMs in terms of the three metrics while the SVMs performed better in terms of sensitivity when the study customized templates ADNI12-Raw and ADNI140-Raw were used. In general, no large differences in performance between the two SVM based methods were observed.

In Tables 3 and 4 the results of the evaluation of Gaussian smoothing effects on voxel-based classification performance is presented for two methods (PLR and HM-SVM) and three different templates (ADNI12-Raw, ADNI140-Raw and ICBM). Smoothing the images led to worse results according to all metrics when voxel-based classification was performed based on the ICBM template. This occurred for both classification methods and both raw and optimized data. When the study customized template ADNI12-Raw was used, the same trend of decreased performance due to smoothing was observed for most situations, with the exception of slight improvements in overall accuracy and sensitivity when the PLR and optimized images were combined.

Finally, when the template ADNI140-Raw was used, the best results in this study according to all metrics were achieved for both methods. This occurred when slight smoothing (FWHM = 4mm) was applied to the images.

Template	Data	Method	ACC	SEN	SPE
ADNI112-Raw	Raw	PLR	85.4 (2.0)	82.5 (3.2)	88.8 (2.3)
		SVM-HM	83.2 (1.6)	82.0 (2.7)	84.1 (4.5)
		SVM-SM	84.7 (1.7)	82.8 (2.4)	87.1 (3.6)
	Opt	PLR	85.3 (1.9)	80.9 (3.7)	91.0 (0.8)
		SVM-HM	83.4 (1.2)	81.3 (2.6)	87.0 (2.2)
		SVM-SM	84.3 (1.7)	81.6 (2.5)	87.4 (3.0)
ADNI112-Opt	Raw	PLR	79.3 (1.7)	76.7 (3.1)	83.1 (3.3)
		SVM-HM	69.2 (3.0)	67.1 (6.9)	71.7 (4.0)
		SVM-SM	68.7 (2.9)	69.0 (2.9)	73.1 (3.4)
	Opt	PLR	79.3 (1.7)	77.2 (3.5)	81.4 (2.4)
		SVM-HM	74.5 (2.3)	75.7 (3.7)	74.7 (3.3)
		SVM-SM	75.8 (3.3)	77.1 (2.1)	76.3 (4.4)
ADNI140-Raw	Raw	PLR	83.8 (1.2)	84.0 (3.1)	83.7 (2.2)
		SVM-HM	85.6 (1.9)	83.8 (2.5)	87.8 (2.1)
		SVM-SM	85.2 (1.2)	83.9 (2.3)	87.5 (1.8)
	Opt	PLR	84.7 (2.8)	84.3 (3.5)	86.0 (4.6)
		SVM-HM	82.0 (1.8)	82.5 (1.8)	82.9 (3.5)
		SVM-SM	84.4 (1.6)	84.9 (2.7)	84.3 (1.8)
ICBM	Raw	PLR	85.4 (1.4)	82.1 (1.5)	89.9 (2.6)
		SVM-HM	81.6 (1.3)	77.1 (3.1)	87.0 (2.4)
		SVM-SM	80.9 (1.7)	77.4 (2.4)	85.6 (2.8)
	Opt	PLR	80.7 (1.4)	81.4 (2.9)	81.2 (2.7)
		SVM-HM	80.3 (2.0)	79.2 (2.2)	82.9 (2.5)
		SVM-SM	78.9 (2.0)	78.3 (2.7)	80.2 (3.5)

Table 1: The performance of PLR, HM-SVM, and SM-SVM in terms of overall accuracy, sensitivity, and specificity is shown for different normalization methods and types of data.

	Attribute	Accuracy		Sensitivity		Specificity	
		Mean (SE)	<i>p</i> -value	Mean (SE)	<i>p</i> -value	Mean (SE)	<i>p</i> -value
Template	ADNI12-Raw	84.4 (0.4)	< 0.001	81.9 (0.5)	< 0.001	87.6 (0.5)	< 0.001
	ADNI12-Opt	74.5 (0.4)		73.8 (0.5)		76.7 (0.5)	
	ADNI140-Raw	84.3 (0.4)		83.9 (0.5)		85.4 (0.5)	
	ICBM	81.3 (0.4)		79.2 (0.5)		84.5 (0.5)	
Data	Raw	81.1 (0.3)	0.89	79.0 (0.3)	0.003	84.1 (0.4)	0.02
	Opt	81.1 (0.3)		80.4 (0.3)		82.9 (0.4)	
Method	PLR	83.0 (0.3)	< 0.001	81.1 (0.4)	< 0.001	85.6 (0.4)	< 0.001
	SVM-HM	80.0 (0.3)		78.6 (0.4)		82.3 (0.4)	
	SVM-SM	80.3 (0.3)		79.4 (0.4)		82.7 (0.4)	

Table 2: Results from analyses of variance of summary measures across 10 trials.

The application of more smoothing (FWHM = 8 mm) to the sMRI images led, in most of the cases, to worse classification performances. The PLR tended to perform better than the HM-SVM across all situations and produced the best performances.

5 Discussion

The first important contribution of this work is the introduction of PLR to solve classification problems of large size that result when sMRI images are used for automatic detection of AD. Contrary, to other high-dimensional methods applied before that perform feature selection

(Davatzikos et al. [17]; Vemuri et al. [50]), dimension reduction steps (Magnin et al. [35]; Teipel et al. [45]) or use the kernel approach (Ashburner [1]; Cuingnet et al. [12]; Cuingnet et al. [14]; Klöppel et al. [31]), we approached the sMRI classification problem as a large-scale regularization problem using voxels as input features. We have shown here that classification problems with a number of features approaching 1 million can be solved with levels of accuracy that are very competitive with other methods in the field (Cuingnet et al. [13]). This is possible through the use of coordinate-wise descent techniques implemented in GLMNET. We compared PLR to the linear SVM approach proposed by (Klöppel et al. [31]) and we found in this

Template	Data	Method	FWHM	Accuracy	Sensitivity	Specificity
ADNI12-Raw	Raw	HM-SVM	NS	83.2 (1.6)	82.0 (2.7)	84.1 (4.5)
			4	81.8 (1.7)	79.7 (2.4)	85.6 (2.5)
			8	80.2 (1.6)	77.5 (3.5)	84.2 (2.6)
		PLR	NS	85.4 (2.0)	82.5 (3.2)	88.8 (2.3)
			4	83.5 (1.6)	81.4 (2.9)	86.6 (2.8)
			8	81.3 (2.4)	79.4 (3.2)	84.0 (3.5)
	Opt	HM-SVM	NS	83.4 (1.2)	81.3 (2.6)	87.0 (2.2)
			4	82.8 (1.9)	80.5 (2.4)	86.6 (2.5)
			8	79.8 (1.4)	77.7 (3.5)	84.0 (2.8)
		PLR	NS	85.3 (1.9)	80.9 (3.7)	91.0 (0.8)
			4	85.6 (1.8)	82.6 (1.4)	89.0 (3.9)
			8	82.9 (2.5)	81.7 (3.7)	84.8 (2.6)
ADNI140-Raw	Raw	HM-SVM	NS	85.6 (1.9)	83.8 (2.5)	87.8 (2.1)
			4	86.5 (2.2)	84.6 (1.5)	89.0 (2.3)
			8	85.0 (1.7)	85.0 (2.4)	85.3 (3.6)
		PLR	NS	83.8 (1.2)	84.0 (3.1)	83.7 (2.2)
			4	88.6 (1.3)	84.8 (1.7)	92.8 (0.8)
			8	86.0 (1.5)	83.2 (2.4)	88.5 (2.4)
	Opt	HM-SVM	NS	82.0 (1.8)	82.5 (1.8)	82.9 (3.5)
			4	85.3 (1.5)	84.9 (2.2)	85.1 (2.9)
			8	85.3 (1.8)	84.1 (2.6)	86.2 (2.9)
		PLR	NS	84.7 (2.8)	84.3 (3.5)	86.0 (4.6)
			4	89.8 (1.3)	85.8 (3.6)	93.2 (2.9)
			8	86.0 (1.5)	83.2 (2.4)	88.5 (2.4)
ICBM	Raw	HM-SVM	NS	81.6 (1.3)	77.1 (3.1)	87.0 (2.4)
			4	77.2 (2.1)	74.9 (2.5)	80.8 (3.0)
			8	73.1 (2.5)	72.7 (3.1)	74.6 (3.0)
		PLR	NS	85.4 (1.4)	82.1 (1.5)	89.9 (2.6)
			4	83.6 (1.6)	78.3 (3.6)	88.7 (3.0)
			8	81.1 (2.1)	77.0 (3.6)	86.4 (2.7)
	Opt	HM-SVM	NS	80.3 (2.0)	79.2 (2.2)	82.8 (2.5)
			4	76.0 (2.2)	74.8 (3.0)	79.0 (3.8)
			8	71.0 (2.5)	70.0 (4.1)	74.4 (4.1)
		PLR	NS	80.7 (1.4)	81.4 (2.9)	81.2 (2.7)
			4	77.9 (1.6)	75.8 (3.3)	81.0 (2.0)
			8	73.3 (2.2)	69.9 (3.4)	78.4 (4.8)

Table 3: The performance of PLR, and HM-SVM in terms of overall accuracy, sensitivity, and specificity is shown for different degrees of smoothness. NS stands for no smoothing.

study that PLR performed better across a wide range of situations. PLR has two well-known advantages over SVM: (1) it models the conditional-class probabilities that in this application (AD automatic detection) could be directly used as a metric to characterize the closeness of a new sample to the two classes CN or AD. In other words, it could be used as biomarker to characterize cognitive status based on spatial brain patterns of atrophy derived from sMRI images. The SVMs do not directly model probabilities, although some approximations have been proposed in the literature (Platt [41]). (2) It extends in a natural way to the case of multi-class classification while SVMs have to rely on different strategies (e.g., one versus all, one versus one, etc.). It is worth to notice that while PLR is relatively time-efficient, the hard-margin linear

SVM is much faster due to the use of kernels and the related dimensionality reduction during the optimization phase. In a computer with two 2.66 Ghz CPUs and using the MATLAB parallel computing toolbox solving ICBM template-related classification problems approaching one million variables ($\approx 9.1 \times 10^5$ voxels) and 98 images took less than 3 minutes using HM-SVM while it takes around 1 hour to compute PLR. The search for the PLR optimal regularization parameter in a nested CV procedure is implemented in a nonoptimized MATLAB code. A different implementation based on C or FORTRAN will speed up PLR. One interesting observation about the SVM method in this study is that in the context of very large classification problems it was not possible to achieve significant increases in performance by tuning the regularization parameter via CV.

	Attribute	Accuracy		Sensitivity		Specificity	
		Mean (SE)	<i>p</i> -value	Mean (SE)	<i>p</i> -value	Mean (SE)	<i>p</i> -value
Template	ADNI12-Raw	82.9 (0.3)		80.6 (0.3)		86.3 (0.4)	
	ADNI40-Raw	85.7 (0.3)	< 0.001	84.2 (0.3)	< 0.001	87.4 (0.4)	< 0.001
	ICBM	78.4 (0.3)		76.1 (0.3)		82.0 (0.4)	
Data	Raw	82.9 (0.2)	< 0.001	80.6 (0.2)	0.13	86.0 (0.3)	< 0.001
	Opt	81.8 (0.2)		80.0 (0.2)		84.5 (0.3)	
Method	HM-SVM	81.1 (0.2)	< 0.001	79.6 (0.2)	< 0.001	83.7 (0.3)	< 0.001
	LOG-2	83.6 (0.2)		81.0 (0.2)		86.8 (0.3)	
FWHM	NS	83.4 (0.3)		81.8 (0.3)		86.0 (0.4)	
	4	83.2 (0.3)	< 0.001	80.7 (0.3)	< 0.001	86.5 (0.4)	< 0.001
	8	80.4 (0.3)		78.4 (0.3)		83.3 (0.4)	

Table 4: Results from analyses of variance of summary measures across 10 trials.

This behavior of SVMs has been reported before by (Hastie et al. [26, page 658]). The second important contribution of this study is that it provides valuable information about how high-dimensional classification methods are affected by image pre-processing steps such as normalization, choice of template, and smoothing, and the effects on classification performance of the optimization process designed by ADNI to increase the quality of the images. Since the ADNI database is a huge asset for the neuroimaging community and ANTS is a very successful normalization software this information is important. In the situation when the data were not smoothed we observed the following.

The template generated using optimized images (ADNI12-Opt) produced clearly the worst results for all methods and for both set of images. This suggests that ANTS performance was affected by the optimization process used by ADNI and that when processing ADNI sMRI images using ANTS, it is better to use the raw data. We were not able to find a clear explanation for this effect though we believe that ANTS normalization had difficulties to deal with some of the ADNI corrections made to improve the sMRI images quality. The two study customized templates built using raw images of 12 and 140 subjects produced similar overall accuracy but the increase of the number of subjects led to increases in sensitivity and losses in specificity.

The standard ICBM template in general produced performances inferior to the customized templates with the exception of the combination of PLR with raw data. It should be noticed that generating the study customized templates could be very time-consuming especially if the number of subjects is large. The combination of ICBM, PLR and raw images reached similar classification performance without the additional computational expenses of generating the customized templates. It could be a option when there is a lack of powerful computational resources. In general, not only PLR produced better performances when compared to SVMs in terms of all metrics across all types of data and templates, but also was more robust to situations that were

adverse to voxel-based classification, such as the use of optimized images to built the custom templates (ADNI12-Opt).

When we smoothed the data it was observed that the ICBM template generated the worse results across methods and types of images with a clear trend of decreased classification performance when the amount of smoothing was increased. The ADNI12-Raw custom template showed very often the same trend though in general the results were better than the ones obtained with the standard ICBM template. However, building a custom template with increased number of subjects combined with slight smoothing (FWHM = 4 mm) yielded the best results in this work. All this suggests that the effect of smoothing depends on the type of template. If a standard or a custom template generated using a relatively small number of subjects is used, smoothing the images seemed to lead to worse classification results. The additional effort in generating a custom template with a large number of subjects seems to pay off when combined with slight smoothing of the images. Even in this case application of greater amount of smoothing (FWHM = 8 mm) leads to decreases in classification performance. These findings are in relative agreement with those reported by Kloppel et al. [31], who did not apply any smoothing to the images, suggesting that no smoothing produced the best results. This contrasts with the heavier amount of smoothing usually recommended in the past (8–12 mm) for univariate voxel-based analyses such as voxel-based morphometry (Good et al. [24]) and also for other sMRI classification methods (Davatzikos et al. [17]; Vemuri et al. [50]) (FWHM = 8 mm). This seems to indicate a decrease need of the degree of smoothing needed when images are normalized with SYN and DARTEL. A possible explanation is that both methods are based on diffeomorphic transformations that are estimated via regularization approaches which include penalties to impose smoothing to the estimated transformation. Therefore, during the normalization process some degree of smoothness could be induced to the images decreasing the need of smoothing afterward.

We need to point out several weaknesses in our study: (1) we did not implement the exact methodology introduced by Kloppel et al. [31], since they based their normalization on DARTEL (Ashburner [1]) a different high-dimensional warping method. For the HM-SVM we fixed the parameter C to a high value while the use a different procedure; (2) due to limitations in computational resources and time, we did not study other situations that could have helped to clarify why normalization templates built using optimized images led to worse results and though we made efforts to find an explanation for that effect we were not able to do so, and (3) we used in this work the L_2 PLR by setting $\alpha = 0$ in equation (2) mostly for computational reasons since our main goal was to study the impact of different pre-processing factors on PLR prediction performance. Preliminary computations (not presented here) using the L_1 PLR showed decrease in performance with respect to the L_2 PLR and very sparse discriminative maps while the elastic net PLR produced similar performance while being much more time-consuming than L_2 PLR. This is an issue that requires more careful evaluation in the future.

6 Conclusion

In this work we have introduced a new application of PLR for high-dimensional classification of structural MRI images. We have compared its performance with another extant methodology based on SVMs using ADNI data. Our results suggest that our approach often produces superior performance in prediction. We additionally have evaluated the influence of different image pre-processing factors on voxel-based classification procedures when combined with ANTS normalization. Our work provides more evidence about the high accuracy of these voxel-based procedures when solving classification problems approaching one million variables (voxels) related to sMRI images of CN and AD subjects. Future work will address comparisons of different high-dimensional warping techniques such as DARTEL and HAMMER (Fan et al. [19]), study of the influence of the type penalties and the performance of conditional probabilities produced by the PLR as biomarkers for early AD prediction.

Acknowledgments This work is funded in part by the NIH Supplement # HHSN268200464221C added to the N01-WH-4-4221 grant. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., as well as nonprofit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation,

with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health (<http://www.fnih.org/>). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation. We also would like thank the thoughtful comments provided by the reviewers that helped us to improve this work.

References

- [1] J. Ashburner, *A fast diffeomorphic image registration algorithm*, Neuroimage, 38 (2007), 95–113.
- [2] B. Avants and J. C. Gee, *Geodesic estimation for large deformation anatomical shape averaging and interpolation*, Neuroimage, 23 (2004), S139–S150.
- [3] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, *Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain*, Med Image Anal, 12 (2008), 26–41.
- [4] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, Berlin, 2006.
- [5] B. Boser, I. Guyon, and V. Vapnik, *A training algorithm for optimal margin classifiers*, in Proc. 5th Annual ACM Workshop on Computational Learning Theory, ACM, Pittsburgh, 1992, 144–152.
- [6] R. Casanova, M. A. Espeland, J. S. Goveas, C. Davatzikos, S. A. Gaussoin, J. A. Maldjian, et al., *Application of machine learning methods to describe the effects of conjugated equine estrogens therapy on region-specific brain volumes*, Magn Reson Imaging, 29 (2011), 546–553.
- [7] R. Casanova, S. Gaussoin, W. Benjamin, J. A. Maldjian, and M. A. Espeland, *An approach to feature selection based on random forest and penalized logistic regression*, in Human Brain Mapping Meeting, Orlando, USA, 2009.
- [8] R. Casanova, J. A. Maldjian, and M. A. Espeland, *A new approach to classification of mri brain images based on penalized logistic regression*, in Third Annual International Brain Conference, Montreal, Canada, 2011.
- [9] R. Casanova, C. T. Whitlow, B. Wagner, J. Williamson, S. A. Shumaker, J. A. Maldjian, et al., *A new high-dimensional machine learning approach for identifying Alzheimer Disease from MRI structural images*, Proc. Intl. Soc. Mag. Reson. Med., 19 (2011), 2555.
- [10] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, ACM Transactions on Intelligent Systems and Technology, 2 (2011), 27:1–27:27.
- [11] C. Chu, *Pattern Recognition and Machine Learning for Magnetic Resonance Images with Kernel Methods*, University College London, London, 2009.
- [12] R. Cuingnet, H. Benali, M. Chupin, and O. Colliot, *Spatial and anatomical regularization of SVM for brain image analysis*, in 24th Annual Conference on Neural Information Processing Systems, NIPS, December 2010.
- [13] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehericy, M. O. Habert, et al., *Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database*, Neuroimage, 56 (2011), 766–781.
- [14] R. Cuingnet, C. Rosso, S. Lehericy, D. Dormont, H. Benali, Y. Samson, et al., *Spatially regularized SVM for the detection of brain areas associated with stroke outcome*, Med Image Comput Assist Interv, 13 (2010), 316–323.

- [15] I. Daubechies, M. Defrise, and D. Mol, *An iterative thresholding algorithm for linear inverse problems with sparsity constraint*, Communications on Pure and Applied Mathematics, 57 (2004), 1413–1457.
- [16] C. Davatzikos, *Why voxel-based morphometric analysis should be used with great caution when characterizing group differences*, Neuroimage, 23 (2004), 17–20.
- [17] C. Davatzikos, F. Xu, Y. An, Y. Fan, and S. Resnick, *Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index*, Brain, 132 (2009), 2026–2035.
- [18] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, *LIBLINEAR: a library for large linear classification*, J Mach Learn Res, 9 (2008), 1871–1874.
- [19] Y. Fan, D. Shen, R. C. Gur, and R. E. Gur, *COMPARE: classification of morphological patterns using adaptive regional elements*, IEEE Trans Medical Imaging, 26 (2007), 93–105.
- [20] J. Friedman, T. Hastie, H. Hofling, and R. Tibshirani, *Pathwise coordinate optimization*, The Annals of Applied Statistics, 1 (2007), 302–332.
- [21] J. Friedman, T. Hastie, and R. Tibshirani, *Package glmnet: lasso and elastic net regularized generalized linear models*. R package version 1.1-4, 2009. <http://cran.r-project.org/web/packages/glmnet/glmnet.pdf>.
- [22] J. Friedman, T. Hastie, and R. Tibshirani, *Regularization paths for generalized linear models via coordinate descent*, J Stat Softw, 33 (2010), 1–22.
- [23] W. Fu, *Penalized regressions: the bridge vs. the lasso*, Journal of Computational and Graphical Statistics, 7 (1998), 397–416.
- [24] C. D. Good, I. S. Johnsrude, J. Ashburner, R. N. Henson, K. J. Friston, and R. S. Frackowiak, *A voxel-based morphometric study of ageing in 465 normal adult human brains*, Neuroimage, 14 (2001), 21–36.
- [25] I. Guyon and A. Elisseeff, *An introduction to variable feature selection*, J Mach Learn Res, 3 (2003), 1157–1182.
- [26] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2nd ed., 2009.
- [27] A. Hoerl, *Application of ridge analysis to regression problems*, Chemical Engineering Progress, 58 (1962), 54–59.
- [28] C. R. Jack Jr., M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, et al., *The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods*, J Magn Reson Imaging, 27 (2008), 685–691.
- [29] J. Jovicich, S. Czanner, D. Greve, E. Haley, A. van der Kouwe, R. Gollub, et al., *Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data*, Neuroimage, 30 (2006), 436–443.
- [30] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M. C. Chiang, et al., *Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration*, Neuroimage, 46 (2009), 786–802.
- [31] S. Kloppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Schill, J. D. Rohrer, et al., *Automatic classification of MR scans in Alzheimer's disease*, Brain, 131 (2008), 681–689.
- [32] B. Krishnapuram, L. Carin, M. A. T. Figueiredo, and A. J. Hartemink, *Sparse multinomial logistic regression: fast algorithms and generalization bounds*, IEEE Trans Pattern Anal Mach Intell, 27 (2005), 957–968.
- [33] Z. Lao, D. Shen, Z. Xue, B. Karacali, S. M. Resnick, and C. Davatzikos, *Morphological classification of brains via high-dimensional shape transformations and machine learning methods*, Neuroimage, 21 (2004), 46–57.
- [34] Z. Liu, F. Jiang, G. Tian, S. Wang, F. Sato, S. J. Meltzer, et al., *Sparse logistic regression with L_p penalty for biomarker identification*, Stat Appl Genet Mol Biol, 6 (2007), Article 6.
- [35] B. Magnin, L. Mesrob, S. Kinkingnehun, M. Pelegrini-Issac, O. Colliot, M. Sarazin, et al., *Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI*, Neuroradiology, 51 (2009), 73–83.
- [36] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. R. Jack, W. Jagust, et al., *Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's disease neuroimaging initiative (ADNI)*, Alzheimers Dement, 1 (2005), 55–66.
- [37] M. Mur, P. A. Bandettini, and N. Kriegeskorte, *Revealing representational content with pattern-information fMRI—an introductory guide*, Soc Cogn Affect Neurosci, 4 (2009), 101–109.
- [38] P. A. Narayana, W. W. Brey, M. V. Kulkarni, and C. L. Sievenpiper, *Compensation for surface coil sensitivity variation in magnetic resonance imaging*, Magn Reson Imaging, 6 (1988), 271–274.
- [39] M. Y. Park and T. Hastie, *Penalized logistic regression for detecting gene interactions*, Biostatistics, 9 (2008), 30–50.
- [40] F. Pereira, T. Mitchell, and M. Botvinick, *Machine learning classifiers and fMRI: a tutorial overview*, Neuroimage, 45 (2008), S199–S209.
- [41] J. Platt, *Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods*, in Advances in Large Margin Classifiers, I. Smola, A. Bartlett, P. Schölkopf, and D. Schuurmans, eds., MIT Press, Cambridge, MA, 1999, 61–74.
- [42] S. Ryali, K. Supekar, D. A. Abrams, and V. Menon, *Sparse logistic regression for whole brain classification of fMRI data*, Neuroimage, 51 (2010), 752–764.
- [43] S. Ryali, K. Supekar, D. A. Abrams, and V. Menon, *Sparse logistic regression for whole-brain classification of fMRI data*, Neuroimage, 51 (2011), 752–764.
- [44] K. Shevade and S. Keerthi, *A simple and efficient algorithm for gene selection using sparse logistic regression*, Bioinformatics, 19 (2003), 2246–2253.
- [45] S. J. Teipel, C. Born, M. Ewers, A. L. Bokde, M. F. Reiser, H. J. Moller, et al., *Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment*, Neuroimage, 38 (2007), 13–24.
- [46] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems*, W. H. Winston, Washington, DC, 1977.
- [47] P. Tseng, *Coordinate Ascent for Maximizing Nondifferentiable Concave Functions*, MIT, Laboratory for Information and Decision Systems, Cambridge, MA, USA, 1988.
- [48] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Berlin, Springer-Verlag, 1998.
- [49] S. Varma and R. Simon, *Bias in error estimation when using cross-validation for model selection*, BMC Bioinformatics, 7 (2006), 91.
- [50] P. Vemuri, J. L. Gunter, M. L. Senjem, J. L. Whitwell, K. Kantarci, D. S. Knopman, et al., *Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies*, Neuroimage, 39 (2008), 1186–1197.
- [51] L. Wang, J. Zhu, and H. Zou, *Hybrid huberized support vector machines for microarray classification and gene selection*, Bioinformatics, 24 (2008), 412–419.
- [52] O. Yamashita, M. A. Sato, T. Yoshioka, F. Tong, and Y. Kamitani, *Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns*, Neuroimage, 42 (2008), 1414–1429.
- [53] J. Zhu and T. Hastie, *Classification of gene microarrays by penalized logistic regression*, Biostatistics, 5 (2004), 427–443.
- [54] H. Zou and T. Hastie, *Regularization and variable selection via the elastic net*, J R Statist Soc B, 67 (2005), 301–320.