

Classification of ADNI PET images via regularized 3D functional data analysis

Xuejing Wang, Bin Nan, Ji Zhu, Robert Koeppe & Kirk Frey

To cite this article: Xuejing Wang, Bin Nan, Ji Zhu, Robert Koeppe & Kirk Frey (2017) Classification of ADNI PET images via regularized 3D functional data analysis, *Biostatistics & Epidemiology*, 1:1, 3-19

To link to this article: <http://dx.doi.org/10.1080/24709360.2017.1280213>



Published online: 13 Mar 2017.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



Classification of ADNI PET images via regularized 3D functional data analysis

Xuejing Wang^a, Bin Nan^b, Ji Zhu^c, Robert Koeppe^d and Kirk Frey^d

^aEli Lilly and Company, Indianapolis, IN, USA; ^bDepartment of Biostatistics, University of Michigan, Ann Arbor, MI, USA; ^cDepartment of Statistics, University of Michigan, Ann Arbor, MI, USA; ^dDepartment of Radiology, University of Michigan, Ann Arbor, MI, USA

ABSTRACT

We propose a penalized Haar wavelet approach for the classification of three-dimensional (3D) brain images in the framework of functional data analysis, which treats each entire 3D brain image as a single functional input, thus automatically takes into account the spatial correlations of voxel-level imaging measures. We validate the proposed approach through extensive simulations and compare its classification performance with other commonly used machine learning methods, which show that the proposed method outperforms other methods in both classification accuracy and identification of the relevant voxels. We then apply the proposed method to the practical classification problems for Alzheimer's disease using positron emission tomography images obtained from the Alzheimer's Disease Neuroimaging Initiative database to highlight the advantages of our approach.

ARTICLE HISTORY

Received 12 September 2016
Accepted 26 December 2016

KEYWORDS

Classification; elastic net; functional logistic regression; Haar wavelets; PCA; PET

1. Introduction

The clinical diagnosis of Alzheimer's disease (AD), the most common cause of dementia, uses a variety of tests including patient's family history, physical examination, mini-mental state examination, and/or neuroimaging. Recently, functional neuroimaging technologies, such as the single-photon emission computed tomography (SPECT) and the positron emission tomography (PET), are rapidly becoming powerful tools in the diagnosis of AD since these technologies have made it possible to reveal pathophysiological changes before irreversible anatomical changes are present. For example, ¹⁸F-Fluorodeoxyglucose (FDG) is a widely used radioactive tracer in PET imaging, and FDG-PET provides useful information about the cerebral glucose metabolic rate. Studies have demonstrated reduced glucose metabolism in a small number of brain regions such as the temporal and parietal lobes in AD patients comparing to normal subjects [1,2]. As such difference becomes noticeable, researchers are increasingly interested in distinguishing AD patients from normal subjects by utilizing their brain images. As a non-negligible complementary way in the diagnosis of AD, PET imaging has high specificity and sensitivity, even a long period before the full-blown dementia is developed.

A large number of brain imaging studies have been performed in patients with AD and its prodromal stage, mild cognitive impairment (MCI), in an effort to assist in the early diagnosis of AD. Traditional methods to discriminate between patients with AD and normal control subjects are mostly based on voxel-wise analysis. However, each image contains as many as millions of voxels, which can be a major cause of practical limitation. To overcome the curse of dimensionality, dimension reduction techniques have been developed prior to classification. One common way is to group the voxels into anatomical regions and average the voxel values within each region of interest (ROI) without taking into account any heterogeneity among the voxels. Prior knowledge of what specific regions may be correlated to the disease is generally desirable; however, this knowledge is not always available in practice. In order to account for the spatial correlation between voxels as well as to reduce the dimension of imaging data, principal component analysis (PCA) has been performed in the literature, which reduces the feature space to a smaller number of principal components (PCs, called eigenimages) while still preserves the largest portion of variability [3,4]. The PC scores are then used as predictors in, for example, the logistic regression. However, each PC is usually comprised of weighted contributions of all voxels within the brain, so PCA is usually less accurate and may blur the true relationship between the progression of disease and voxels as what we will show later in numerical studies. Recent development in random matrix theory shows that the PCs obtained from the large-scale sample covariance matrix are not valid [5,6].

There has been a growing interest in developing machine learning classification techniques due to the large number of voxels. Support vector machines (SVMs) are one of these techniques used for binary classification [7]. They aim to find the hyperplane that maximizes the distance from the nearest training points while correctly separating two classes. To avoid the curse of dimensionality and improve the prediction performance, SVMs are often performed on selected features, including selected voxels [8,9], ROIs [10], or even PCs [11,12]. In an attempt to incorporate the spatial correlation, Stoeckel and Fung [13] and Hinrichs *et al.* [14] presented modified versions of SVMs that are implemented by setting similar weights to neighbouring voxels at a very local level. Although SVM-like methods have been shown to achieve high classification accuracy rates, they are not optimized for selecting sensitive and interpretable disease-related brain subregions and fail to provide estimates for the probability that a given subject has the disease or not. To address this issue, logistic regression can be implemented, often with a regularization for variable selection to prevent overfitting. Regularized logistic regression models with the potential of taking into account the highly correlated predictors in imaging have been proposed for this purpose. Shen *et al.* [15] developed an ROI-based regularized logistic regression model with the elastic net penalty [16], a linear combination of lasso and ridge penalties, to classify AD subjects from others. The elastic net penalty enables to select groups of highly correlated ROIs. Their method does not consider correlations between voxels within each ROI. Casanova *et al.* [17] discussed regularized logistic regression with the elastic net penalty in the context of a large-scale regularization problem in which voxels are used as predictors. For the typical large- p -small- n classification problem, they showed that it can be solved efficiently using the coordinate descent algorithm [18], and it is preferable to SVMs in terms of classification accuracy. However, they did not evaluate the performance in identifying relevant voxels, and in fact they only implemented the ridge regression.

In this paper, we propose a highly effective and computationally efficient regularized functional logistic regression approach using Haar wavelets, which automatically preserves the spatial correlation information of voxels by viewing each subject's image as a realization of 3D functional predictor. Functional logistic regression has been used in the classification of functional data. For example, Reiss and Ogden [19] considered the problem by applying functional PCA to images and demonstrated their method in two-dimensional (2D) settings. Reiss *et al.* [20] and Ciarleglio and Ogden [21] considered the wavelet-based approach for the image predictors. Zhao *et al.* [22] proposed a general wavelet-based lasso approach in functional linear regression, but only in the 1D case. The general goal of functional data analysis (FDA) is to estimate the coefficient function that describes the association between an outcome and a functional predictor. In this study, we are interested in finding out which voxels are most responsive in determining the disease status. In particular, we assume that only few brain subregions are predictive to the disease status. Properly regularized FDA with Haar wavelet expansion is able to yield a sparse coefficient function estimate (taking value zero at most places in the brain) and also enjoys the advantage of preserving the spatial correlation among voxels. Note that this desirable property was discussed in detail in Wang *et al.* [23] for linear models. We extend their method to logistic regression models with image predictors. To demonstrate advantages of the proposed approach, we compare it with other classification methods including regularized voxel-level logistic regression with the elastic net penalty and PCA-based logistic regression.

The data used in this paper are baseline FDG-PET images of 403 subjects from the Alzheimer's Disease Neurological Initiative (ADNI) database, including 95 AD patients, 206 MCI patients, and 102 normal controls (NC). The rest of this paper is organized as follows. We present the proposed approach and also describe two other classification methods in Section 2. Numerical results for analysing the simulated and real data-sets are presented in Section 3, demonstrating the superior voxel selection and classification performance of the proposed approach. Final conclusions are provided in Section 4.

2. Materials and methods

In this section, we present three logistic regression-based methods for the classification of brain images and briefly introduce the data-set we use in preparation of this paper. In particular, we describe how the proposed approach is applied in the functional regression framework for analysing brain images, and also explain in detail why we choose to use Haar wavelets.

2.1. Haar-wavelet-based regularized functional logistic regression (HW-RFLR)

Logistic regression is commonly used for a binary response variable Y . Functional logistic regression is developed to relate the response variable Y to a functional predictor. Here we treat each subject's 3D brain image as a functional predictor X_i , where $X_i(u, v, w)$ is the covariate value, e.g. the FDG-PET image measure, of the voxel located at (u, v, w) for subject i . Suppose Y_i takes values either 0 or 1, indicating the disease status of subject i . We fit

the following 3D functional logistic regression model

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \iiint X_i(u, v, w)\beta(u, v, w) dudvdw, \quad i = 1, \dots, n, \quad (1)$$

where $\pi_i = P(Y_i = 1|X_i)$ for subject i and $\beta(u, v, w)$ is the 3D regression coefficient function. In this study, we are particularly interested in the assumption that $\beta(u, v, w) = 0$ over large regions, and potential discontinuities of β are allowed.

Choosing proper basis functions to represent β in the above regression model is a critical step. Among a variety of basis functions, we choose 3D Haar wavelets to decompose β owing to the following desirable properties. First, the use of Haar wavelets provides a way of overcoming the issue of multicollinearity caused by large spatial correlation among neighbouring voxels. Haar wavelets consist of piecewise constant functions. Our estimation procedure tends to estimate β to be zero or non-zero altogether for a cluster of neighbouring voxels instead of a single voxel. Second, as mentioned earlier, we assume that only few brain subregions are predictive, implying sparsity of the coefficient function. Exact zero regions can be yielded by the sparsity of wavelet coefficients (see Wang et al. [23], Appendix B). Third, Haar wavelets can be applied as a signal compression technique. They provide a good approximation of the original function with only a subset of non-zero wavelet coefficients, which can be achieved by zeroing out the wavelet coefficients that are smaller than a prespecified threshold value. The dimensionality can thus be reduced if we only consider the non-zero subset.

3D Haar wavelets can be obtained by tensor products of 1D Haar wavelets. For simplicity, we assume that $0 \leq u, v, w \leq 1$. 1D Haar wavelets can be constructed from a mother wavelet function and a scaling function. The mother wavelet function $\psi(t)$ is given by

$$\psi(t) = \begin{cases} 1 & \text{if } 0 \leq t < 1/2, \\ -1 & \text{if } 1/2 \leq t < 1, \\ 0 & \text{otherwise;} \end{cases}$$

and the scaling function $\phi(t)$ is given by

$$\phi(t) = \begin{cases} 1 & \text{if } 0 \leq t < 1, \\ 0 & \text{otherwise.} \end{cases}$$

All 1D Haar wavelets are obtained as translated and dilated versions of the above functions:

$$\begin{aligned} \psi_{j,k}(t) &= \sqrt{2^j} \psi(2^j t - k), \\ \phi_{j,k}(t) &= \sqrt{2^j} \phi(2^j t - k), \end{aligned}$$

where $j = 0, 1, \dots$ and $k = 0, 1, \dots, 2^j - 1$. The index j refers to dilations and k refers to translations and $\sqrt{2}$ is the normalizing factor. It can be seen that these basis functions are orthogonal to each other, and the support becomes smaller as j increases. The functions $\phi_{j,k}(t)$ and $\psi_{j,k}(t)$ are usually referred to as averaging and differencing operations,

respectively. Let us now consider tensor products of three elements with each of them being $\phi(t)$ or $\psi(t)$. The total number of different combinations is $2^3 = 8$. The 3D scaling function is the tensor product of three 1D scaling functions $\phi(t)$. 3D mother wavelet functions are the remaining seven tensor products considering all cross-spatial horizontal, vertical, and diagonal directions. 3D Haar wavelets are generated as adapted translations and dilations of these functions by using a tensor product of three 1D Haar wavelets. For example,

$$\phi_{j,\{k,l,m\}}(u, v, w) = \phi_{j,k}(u)\phi_{j,l}(v)\phi_{j,m}(w).$$

For more details about the construction of 3D Haar wavelets, see [24].

Let $B(u, v, w)$ denote the collection of 3D Haar wavelet basis functions constructed above using tensor products. We can now decompose $X_i(u, v, w)$ and $\beta(u, v, w)$ as follows:

$$X_i(u, v, w) = C_i^T B(u, v, w), \quad \beta(u, v, w) = B(u, v, w)^T \eta, \quad (2)$$

where C_i is the *known* wavelet coefficient vector of X_i and η is the *unknown* coefficient vector of β . Then by the orthogonality of wavelet basis functions, the 3D functional logistic regression reduces to the following multiple logistic regression by plugging (2) into (9):

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + C_i^T \eta. \quad (3)$$

Once an estimator of η is obtained from (3), an estimator of β can be obtained from (2).

It should be noted that the wavelet expansion of a given function is determined by the coarsest and the finest levels of decomposition. In practice, we only observe $X(u, v, w)$ discretely, e.g. at a finite number of uniformly distributed voxels in a cube. Representing observed $X(u, v, w)$ by a set of wavelet coefficients is called discrete wavelet transform. In this case, the finest level is always given as the operations on adjacent voxels, and thus only the coarsest level needs to be determined, which is referred to as the level of decomposition in this paper. The number of basis functions in the set $B(U, V, W)$ or the length of C_i depends on the level of decomposition.

The estimation of η in (3) is accomplished by fitting the model via a penalized maximum log-likelihood:

$$\hat{\eta} = \arg \max_{\eta} \frac{1}{n} \sum_{i=1}^n \{Y_i \log \pi_i + (1 - Y_i) \log (1 - \pi_i)\} - \lambda \|\eta\|_1, \quad (4)$$

where $\|\cdot\|_1$ denote the L_1 norm. Such a penalty is called the lasso penalty [25] that makes many estimated coefficients to be exactly zero. The constant $\lambda \geq 0$ is a tuning parameter that determines how much shrinkage is applied to the vector η . This regularized logistic regression problem can be efficiently solved by the coordinate descent algorithm [18].

The estimation of $\beta(u, v, w)$ is then obtained by

$$\widehat{\beta}(u, v, w) = B(u, v, w)^T \widehat{\eta}.$$

Note that when the wavelet transform is performed under different levels of decomposition, the obtained $\widehat{\eta}$ is different, corresponding to different collections of $B(u, v, w)$. As a result, the estimator $\widehat{\beta}(u, v, w)$ would be different. We set the level of decomposition as another tuning parameter in addition to λ in (4), and their optimal values will be determined by certain criterion using a data-driven approach. Hence an identified brain region consists of a cluster of neighbouring cubes, where the sizes of the cubes are controlled by the level of decomposition.

2.2. Elastic net regularized logistic regression (EN-RLR)

The elastic net is considered as a generalized version of lasso which encourages a grouping effect by allowing strongly correlated predictors to be in or out of the model together. It also enjoys the computational advantage of lasso. It should be noted that Haar wavelets are constructed to account for the grouping effect of the neighbouring voxels, so we only implement the L1 penalty in Section 2.1. To compare with the approach proposed in Section 2.1, here we evaluate the performance of EN-RLR at the voxel level rather than the ROI level [15], which is given by

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \sum_{u,v,w} X_i(u, v, w) \beta(u, v, w), \quad i = 1, \dots, n, \quad (5)$$

where (u, v, w) are integers indicating the location of the corresponding voxel. Denote the total number of voxels by p . Since $p \gg n$, regularization is needed to prevent overfitting. The elastic net method maximizes the following regularized log-likelihood function:

$$\frac{1}{n} \sum_{i=1}^n \{Y_i \log \pi_i + (1 - Y_i) \log (1 - \pi_i)\} - \lambda P_\alpha(\beta), \quad (6)$$

where $P_\alpha(\beta) = \sum_{u, v, w} \{\alpha |\beta(u, v, w)| + (1 - \alpha) \beta(u, v, w)^2\}$. It can be seen that the penalty $\lambda P_\alpha(\beta)$ is a mixture of L_1 and L_2 penalties, and when $\alpha = 1$, (6) is simplified to the lasso problem. We set both λ and α as tuning parameters, whereas in [17], α is set to be zero to enforce the L_2 penalty, resulting in a ridge regression. Note that (4) is for the wavelet-transformed images whereas (6) is for the original images.

2.3. Principal-component-based logistic regression (PC-LR)

PCA is a widely used tool for dimension reduction. It projects the original images into the eigenspace such that the variance of the projection along each component, the so-called principal component (PC), is maximized [3]. Each PC is referred to as an eigenimage. As most of the variability of images are captured by a small number of PCs, we retain the first few PCs with greater variances. The associated PC scores are treated as predictors in the

logistic regression model. The original coefficient function β can be obtained by the inverse transform of the coefficients of PC scores. The model can be written as follows:

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + S_i^T \zeta, \quad i = 1, \dots, n. \quad (7)$$

where S is the matrix composed of PCs of the matrix X . The original β will be estimated by $\widehat{\beta} = V\widehat{\zeta}$, where the columns of V are corresponding loadings of the PCs. The number of PCs used in the regression may affect the classification performance. Thus, in this work, we treat the number of PCs as a tuning parameter.

2.4. PET imaging data

PET imaging data analysed in this paper were obtained from the ADNI database (adni.loni.ucla.edu). The ADNI project was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and non-profit organizations, as a \$60 million, five-year public-private partnership. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers for disease progression in very early AD 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 co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the USA and Canada.

Detailed information about how FDG-PET images were acquired is available on the ADNI website. The processing steps can be summarized as follows. First, six five-minute frame scans were acquired 30–60 min after injecting FDG to the participants. These frame scans were co-registered to the first frame and then averaged to create a single image. After this step, the co-registered, averaged PET images were reoriented into a standard $160 \times 160 \times 96$ voxel image grid with 1.5-mm cubic voxels and the anterior-posterior axis of the subject is parallel to the anterior commissure (AC)–posterior commissure (PC) line. Finally, smoothing is performed to produce a uniform resolution. The data-set used in the present work consists of 403 participants' baseline scans, including 102 NC participants, 206 MCI participants, and 95 AD participants. After a careful review of each AD case by co-author Dr Frey, 15 out of the original 95 AD cases are found to be actually non-AD dementia cases, thus removed from the analysis. By the 48th month of the follow-up, 86 out of 206 MCI patients have converted to AD. To reduce the dimensionality, we set the values of voxels outside the brain and in the ventricles to zero and exclude the columns whose elements are all zero in the wavelet coefficient matrix obtained after applying 3D Haar wavelet transform to the images. Typically, the dimensionality can be reduced to about 700,000 from more than two million, which is a significant decrease.

3. Results

3.1. Simulations

We conduct simulation studies to evaluate the performance of the three classification methods. The images, covariates in the logistic regression, are obtained from the ADNI data-set. For illustrative purposes, we extract the same 160×160 axial slice from each subject and use it as the 2D functional covariate $X_i(u, v)$. For a given coefficient function β , we randomly generate the response variable Y_i from a Bernoulli distribution with success probability π_i determined by the following 2D model:

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \iint X_i(u, v)\beta(u, v) dudv, \quad i = 1, \dots, n. \quad (8)$$

The regression coefficient function $\beta(u, v)$ is chosen to be non-zero at two small round regions, see [Figure 2\(a\)](#). The scale of $\beta(u, v)$, together with the intercept β_0 , is adjusted to achieve a Bayes error rate [26] around 0.15 which is the lowest possible error rate for any classifier. We consider two case-control ratios, the ratio of the number of occurrence ($Y_i = 1$) to the number of non-occurrence ($Y_i = 0$), at $r = 1:1$ and $r = 1:2$, by randomly choosing 200 and 300 subjects from the ADNI data-set, respectively, for their 2D images. Such designed simulation study keeps the original spatial correlation structure of the ADNI FDG-PET images, whereas the disease status is randomly generated with the original disease status completely ignored.

The selection of tuning parameters is involved in all three methods. We consider a variety of criteria for determining optimal tuning parameters, including cross-validated deviance (CV-DEV), cross-validated misclassification error rates (CV-MER), cross-validated area under the ROC curve (CV-AUC), AIC, and BIC, where the use of CV-DEV and CV-MER is discussed in [18], and CV-AUC criterion, specially designed for optimizing the classification performance for binary outcomes, is discussed in [27]. AIC and BIC are criteria that penalize the number of free parameters, which are common for variable selection in high-dimensional models. Note that in the calculation of AIC and BIC, the degrees of freedom (df) need to be determined. An unbiased estimate of df when only L_1 penalty is used is the number of non-zero coefficients in the model [28], while an unbiased estimate of df is derived as the trace of the modified hat matrix, when a mixture of penalties is presented, see [29] for details. In PC-LR, df is estimated as the number of PCs used in the model.

We apply a 10-fold cross-validation to evaluate prediction accuracy. Specifically, each simulated data-set is randomly partitioned into 10 folds. Among them, nine folds are used as the training set to fit model (8) by each of the optimal tuning parameter selection criteria; the remaining fold is used as a test set to calculate the predicted probability $\hat{\pi}_i$ for each test observation. The procedure is repeated 10 times with each of the 10 folds used exactly once as the test set. For predictions based on each cut-off value of $\hat{\pi}_i$, we compute sensitivity and specificity and then construct the empirical ROC curves by changing the cut-off point of $\hat{\pi}_i$. Area under the ROC curve (AUC) is calculated to provide an overall measure of the discriminative ability of each of the three classification models. The procedure is repeated 100 times by generating 100 independent sets of binary response variables. The average ROC curves with average AUCs are presented in [Figure 1](#). It shows that

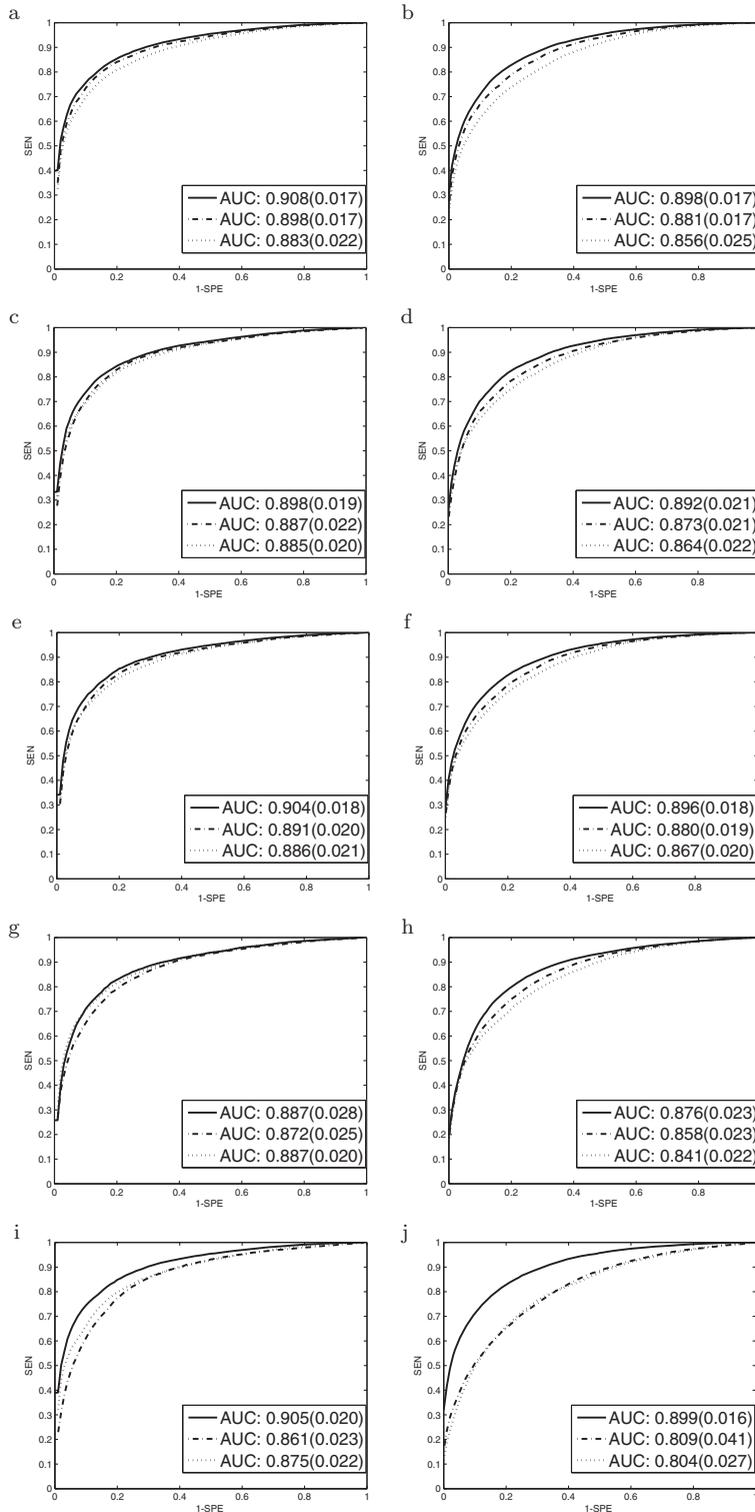


Figure 1. Simulation results. Average ROC curves and their average AUCs (standard errors) for three classification methods over 100 simulations (solid curve: HW-RFLR; dash-dot curve: EN-RLR; dotted curve: PC-LR). The left panel is for $r = 1$, the right panel is for $r = 2$. The selection criterion used in (a) and (b) is CV-DEV, (c) and (d) is CV-MCR, (e) and (f) is CV-AUC, (g) and (h) is AIC and (i) and (j) is BIC.

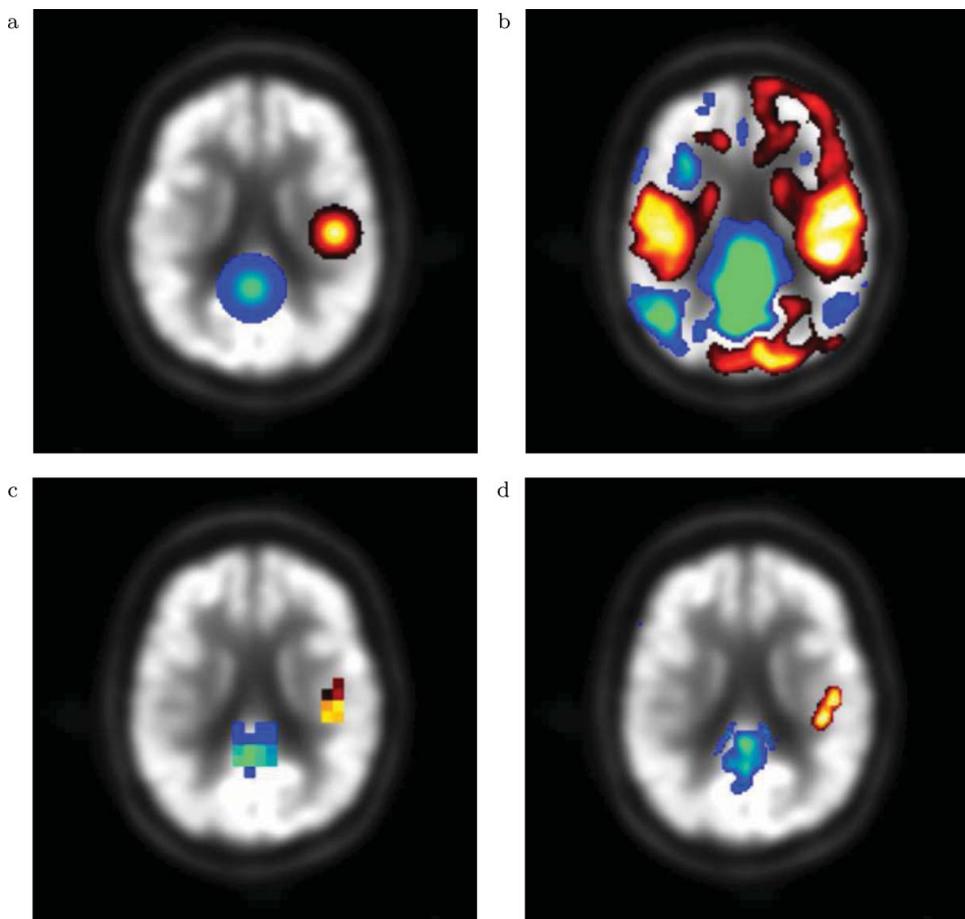


Figure 2. Simulation results. Comparison of median of $\hat{\beta}(u, v)$ using three methods over 100 replications for the simulated data sets under the case of $r = 1: 1$. (a) true $\beta(u, v)$; (b) median of $\hat{\beta}_{\text{PC-LR}}(u, v)$; (c) median of $\hat{\beta}_{\text{HW-RFLR}}(u, v)$; (d) median of $\hat{\beta}_{\text{EN-RLR}}(u, v)$.

the proposed HW-RFLR consistently dominates the other two methods. In general, our simulations indicate that the proposed HW-RFLR approach can achieve higher classification accuracy than EN-RLR and PC-LR for both balanced (i.e. $r = 1: 1$) and unbalanced (i.e. $r = 1: 2$) case-control data.

In addition to the classification performance, we also assess the performance of identifying non-zero regions of $\beta(u, v)$ on the 160×160 grid. To this end, we fit model (8) with the optimal tuning parameters selected via the selection criterion CV-DEV, the one with the best performance in above simulations. Figure 2 shows the median estimates of $\beta(u, v)$ for the 100 repetitions. The results of $r = 1: 1$ and $r = 1: 2$ are similar, so we only present the results of $r = 1: 1$. From Figure 2, we see that HW-RFLR and EN-RLR methods not only yield sparse estimates of $\beta(u, v)$, but also correctly pick up the two non-zero regions of $\beta(u, v)$, whereas PC-LR method yields non-sparse result that include a large number of falsely discovered voxels.

3.2. FDG-PET image analysis

In this subsection, we apply all three methods described in Section 2 to the ADNI FDG-PET imaging data to build classifiers for discriminating AD from NC, then predict AD conversions among MCI using baseline images. We treat each PET image as a realization of the 3D functional predictor and then fit 3D functional logistic regression model (9) for the classification of AD and NC, where $Y = 1$ indicates AD state. Similar to simulation studies, we assess the classification performance using 10-fold cross-validation. To examine the overall discriminative power, we plot the cross-validated ROC curves in Figure 3(a). Corresponding AUCs are also provided. Here we use CV-DEV as the criterion to select the tuning parameters. It can be seen that all three methods work well in discriminating AD from NC. All the methods result in classification accuracies above 90%. The same methods are also applied to the comparison of MCI converters and NC using baseline images. This is a more difficult classification problem, where we see that the proposed HW-RFLR performs the best (see Figure 3(b)).

The estimated regression coefficient function for each classification by the proposed HW-RFLR approach are given in Figure 4, which is depicted, superimposed on the average image of all the NC images. The clusters of voxels identified by the proposed approach are sparsely located over the brain, representing the set of voxels that jointly achieve the highest cross-validated classification accuracy. It suggests that only a few brain subregions are identified to be predictive. The voxels with cold colours are negatively associated with the more severe disease state, whereas the voxels with warm colours indicate a positive relationship. Many of these identified clusters have been found discriminative in other ROI-based analyses, for example, [2,30,31]. In particular, Figure 4(a) shows the anatomic locations of voxels distinguishing probable AD subjects from cognitively normal subjects. Regions with negative coefficients correspond anatomically to the parahippocampal gyrus and anterior hippocampus (levels -48 to -30) and to the posterior cingulate and parietal association cortices (levels $+12$ to $+30$). Whereas Figure 4 shows the anatomic locations of voxels distinguishing subjects with progressive MCI from cognitively normal subjects. Regions with negative coefficients at the base of the brain correspond to the amygdala

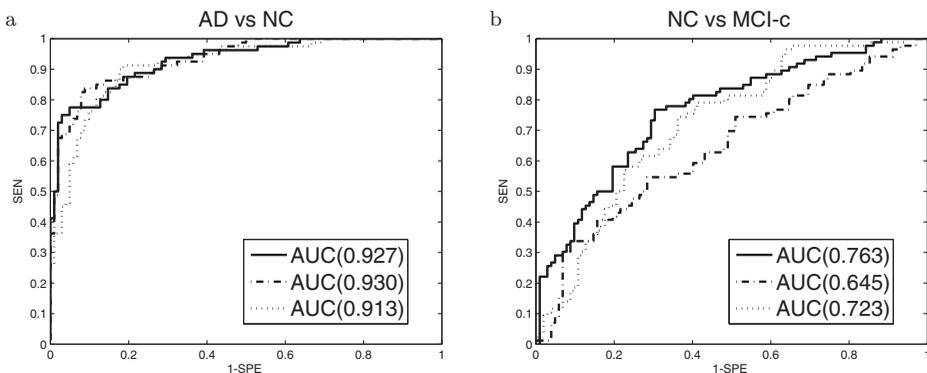


Figure 3. Cross-validated ROC curves and their AUC for three classification methods under CV-DEV criterion for pairwise classification (solid curve: HW-RFLR; dash-dot curve: EN-RLR; dotted curve: PC-LR). (a): AD vs. NC; (b) MCI converters vs. NC.

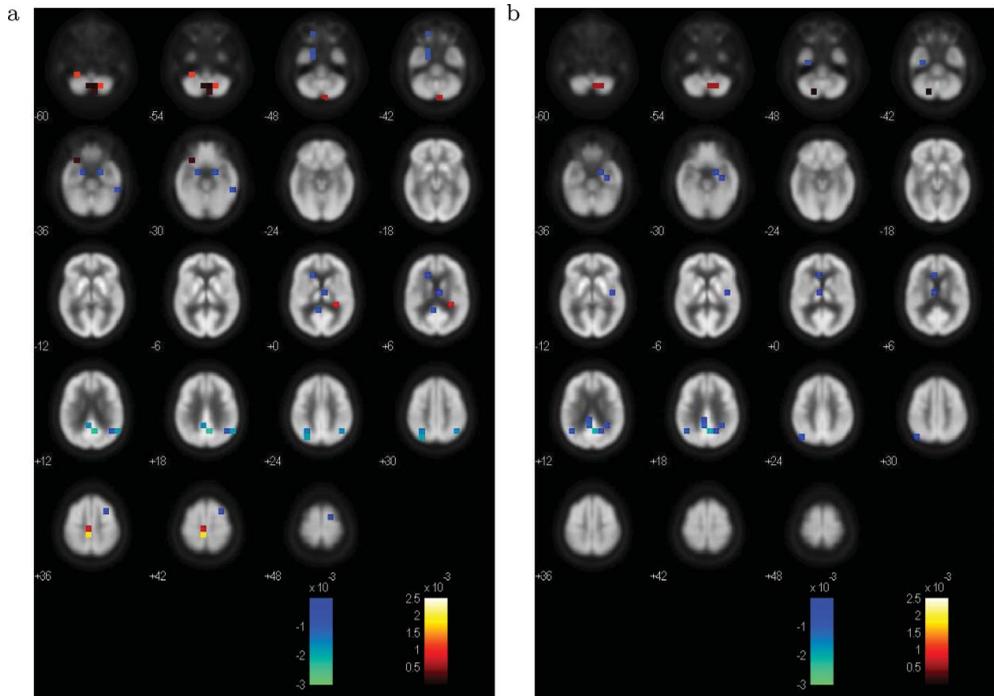


Figure 4. Voxels identified using the proposed HW-RFLR under CV-DEV criterion for pairwise classification. (a) AD vs. NC; (b) MCI converters vs. NC.

and anterior hippocampus (levels -36 and -30). Additional regions correspond to the parietal association cortex and to the posterior cingulate cortex (levels $+12$ to $+24$). Cerebellar regions (levels -60 and -54) with positive regression coefficients are most likely a result of global data normalization rather than true increased FDG metabolism in MCI subjects.

To validate the proposed method in predicting early AD, we apply the logistic regression models built on AD and NC images to the independent MCI baseline images for predicting AD conversions. The results are given in Figure 5. Note that the images used for the prediction in Figure 5 are up to four years earlier from the time of AD conversion, and the MCI non-converters are different to NC, which make it a harder problem than distinguishing AD and NC. It is interesting to see that the model built by the proposed HW-RFLR performs the best with a predictive AUC of 0.779, which is even slightly higher than the cross-validated AUC of the classification of MCI converters and NC, indicating a high prediction precision of the classifier built upon AD and NC images.

4. Discussion

In this paper, we have described a Haar wavelet approach for classifying brain images in the framework of 3D functional data. This approach is demonstrated to not only achieve high classification accuracy, but also be more likely to identify the most responsive clusters of voxels. The proposed regularized Haar-wavelet-based functional logistic regression

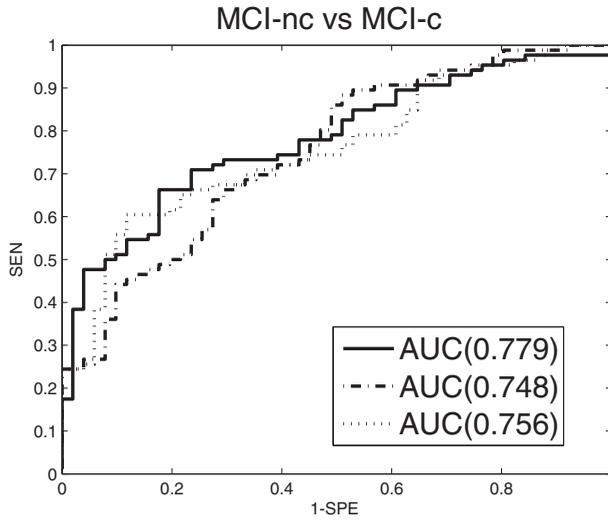


Figure 5. Predictive ROC curves and their AUC for three classification methods under CV-DEV criterion for pairwise classification (solid curve: HW-RFLR; dash-dot curve: EN-RLR; dotted curve: PC-LR).

does not impose smoothness requirement on the regression coefficient function, thus has the potential to identify the boundaries of truly predictive subregions. Our numerical results demonstrate that the proposed HW-RFLR can achieve higher classification accuracy than other methods. We also compared different tuning parameter selection criteria in this work, based on the simulations, choosing the tuning parameter based on deviance performs slightly better than others. The proposed HW-RFLR is not sensitive to the selection criteria. It should be noted that although many previous studies reported classification accuracy rates using FDG-PET image data, most of them did not perform the selection of voxels. We emphasize that the proposed HW-RFLR method integrates voxel selection into the estimation procedure, which is useful when only few brain subregions are related to the disease status. Other baseline covariates (such as age, gender, etc.), denoted by Z in the following, can be easily incorporated in the model:

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \int \int \int X_i(u, v, w) \beta(u, v, w) dudvdw + Z^T \gamma, i = 1, \dots, n. \quad (9)$$

The prediction performance for the ADNI data is primarily driven by the brain regions because the predictions based on the adjusted models are very similar to the unadjusted one. Thus the results of the adjusted models are omitted.

The proposed approach is computationally efficient partly due to the fact that Haar wavelets can further compress the data by thresholding the absolute value of wavelet coefficients without losing the ability of preserving spatial correlations among voxels. In the ADNI FDG-PET imaging data analysis, we excluded the voxels outside the brain prior to model fitting, thus dramatically reduced the number of considered voxels. The computation of the ADNI data example can be done within several hours with the implementation of coordinate descent algorithm using the MATLAB glmnet package on a 64-bit Intel Xeon 3.33 GHz server with about 35 GB of RAM. The proposed HW-RFLR method can

deal with images of any dimension, and provides highly interpretable estimates of the coefficient function due to sparsity.

Acknowledgments

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. 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.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported in part by NIH [grant number R01-AG036802]; NSF [grant number DMS-1007590], [grant number DMS-1407142], [grant number DMS-0748389].

Data collection and sharing for this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health) [grant number 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; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Amorfix Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (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 [grant number P30 AG010129], [grant number K01 AG030514].

Notes on contributors

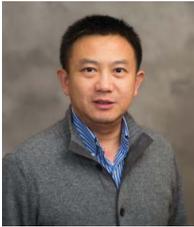


Dr Xuejing Wang obtained her BSc degree in statistics from Nankai University in 2006 and her PhD degree in biostatistics from the University of Michigan in 2013. She joined Eli Lilly and Company as a research scientist (statistics) in 2013. Her research interests include brain imaging data analysis, survival analysis and clinical trials.



Dr Bin Nan is a professor of biostatistics and statistics at the University of Michigan. He received his PhD degree in biostatistics from the University of Washington in 2001 and joined the faculty at the University of Michigan in the same year. Dr Nan's research interests are in various areas of statistics and biostatistics including semiparametric inference, failure time and survival analysis, longitudinal data, missing data and two-phase sampling designs, and high-dimensional data analysis. He is collaborating in many studies in the areas of epidemiology, bioinformatics, and brain imaging, particularly in cancer, HIV, women's health, and neurodegenerative diseases. He is a fellow of

the American Statistical Association and a fellow of the Institute of Mathematical Statistics.



Dr Ji Zhu received his BSc degree in physics from Peking University in China, and his PhD degree in statistics from Stanford University in 2003. He is now a professor in the Department of Statistics at the University of Michigan. Dr Zhu has devoted his research to developing theory and methodologies in the fields of classification, clustering, variable selection, high-dimensional data analysis and statistical network analysis. He received a CAREER award from the National Science Foundation (USA) in 2008, and he was elected as the fellow of the American Statistical Association in 2013 and the fellow of the Institute of Mathematical Statistics in 2015.



Dr Robert Koeppe received his PhD degree in medical physics from the University of Wisconsin-Madison. He is a professor of radiology at the University of Michigan. He has been the director of the University of Michigan PET Physics section for 25 years and has had major research focus on kinetic modeling and data analysis of dynamic PET data. He has extensive experience in brain radiotracer studies and has collaborated with neurologic investigators at Michigan his entire career. He has also been involved in many large and small multicenter trials, including the ADNI and DIAN trials performing quality assurance and data analysis on PET image sets from other centers.



Kirk A. Frey, MD, PhD is the David E Kuhl Collegiate Professor of Nuclear Medicine in Radiology, Professor of Neurology and Research Professor in the Molecular and Behavioral Neuroscience Institute at The University of Michigan. Dr Frey is the chief of the Division of Nuclear Medicine and Molecular Imaging, director of the Center for Positron Emission Tomography, and co-director of the Neurology Movement Disorders Clinic. He has served as the program director for Nuclear Medicine residency training at The University of Michigan since 2004. He serves currently on the ACGME Nuclear Medicine Residency Review Committee, and in the past, has served as a member of

the Board of Directors of the American Board of Nuclear Medicine and of the Board of Directors of the American Board of Medical Specialties. He holds lifetime certifications from the American Board of Psychiatry and Neurology (Neurology, 1989) and the American Board of Nuclear Medicine (1989). Dr Frey has reported more than 200 peer-reviewed scientific manuscripts. He is a recipient of the Marc Tetalman Memorial Award and the Kuhl-Lassen Award from the Society of Nuclear Medicine and Molecular Imaging, is an elected member of the American Neurological Association and is a distinguished investigator in the Academy of Radiology Research. His research interests focus on molecular imaging of neurodegenerations, including Parkinson disease, Alzheimer disease and related disorders. He has collaborated also on research studies of radiation therapy responses in lung cancer and on the development of respiratory gating in thoracic PET imaging.

References

- [1] Hoffman JM, Welsh-Bohmer KA, Hanson M, et al. FDG PET imaging in patients with pathologically verified dementia. *J Nucl Med.* **2000**;41(11):1920–1928.
- [2] Langbaum JB, Chen K, Lee W, et al. Categorical and correlational analyses of baseline fluoro-deoxyglucose positron emission tomography images from the Alzheimer’s disease neuroimaging initiative (ADNI). *Neuroimage.* **2009**;45:1107–1116.
- [3] Zuendorf G, Kerrouche N, Herholz K, et al. Efficient principal component analysis for multivariate 3D voxel-based mapping of brain functional imaging data sets as applied to FDG-PET and normal aging. *Hum Brain Mapping.* **2003**;18:13–21.
- [4] Higdon R, Foster NL, Koeppe RA, et al. A comparison of classification methods for differentiating fronto-temporal dementia from Alzheimer’s disease using FDG-PET imaging. *Stat Med.* **2004**;23:315–326.
- [5] Bai ZD, Ying YQ. Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann Probab.* **1993**;21(3):1275–1294.
- [6] Johnstone IM, Lu AY. On consistency and sparsity for principal components analysis in high dimensions. *J Amer Stat Assoc.* **2009**;104(486):682–693.
- [7] Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference and prediction.* New York (NY): Springer; **2009**.
- [8] Silveira M, Marques J. Boosting Alzheimer disease diagnosis using PET images. In: 20th International Conference on Pattern Recognition; **2010**. pp. 2556–2559. IEEE.
- [9] Vemuri P, Gunter JL, Senjem ML, et al. Alzheimer’s disease diagnosis in individual subjects using structural MR images: validation studies. *NeuroImage.* **2008**;39:1186–1197.
- [10] Dehghan H, Pouyan AA, Hassanpour H. SVM-based diagnosis of the Alzheimer’s disease using 18F-FDG PET with Fisher discriminant rate. In: 18th Iranian Conference on Biomedical Engineering; **2011**. pp. 37–42. IEEE.
- [11] Bonneville M, Meunier J, Bengio Y, et al. Support vector machines for improving the classification of brain PET images. *SPIE Med Imaging.* **1998**;3338:264–273.
- [12] Illán I, Górriz J, Ramírez J, et al. 18F-FDG PET imaging analysis for computer aided Alzheimer’s diagnosis. *Inf Sci.* **2011**;181:903–916.
- [13] Stoeckel J, Fung G. SVM feature selection for classification of SPECT images of Alzheimer’s disease using spatial information. *J Knowl Inf Syst.* **2007**;23(2):243–258.
- [14] Hinrichs C, Singh V, Mukherjee L, et al. Spatially augmented LP boosting for AD classification with evaluations on the ADNI dataset. *Neuroimage.* **2009**;48:138–149.
- [15] Shen L, Kim S, Qi Y, et al. Identifying neuroimaging and proteomic biomarkers for MCI and AD via the elastic net. In: Liu T, Shen D, Ibanez L, Tao X, editors. *Multimodal brain image analysis.* Vol. 7012, Lecture Notes in Computer Science. Berlin Heidelberg: Springer; **2011**. pp. 27–34.
- [16] Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc B.* **2005**;67:301–320.
- [17] Casanova R, Maldjian JA, Espeland MA. Evaluating the impact of different factors on voxel-based classification methods of ADNI structural MRI brain images. *Int J Biomed Data Min.* **2011**;1:11–20.
- [18] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* **2010**;33(1):1–22.
- [19] Reiss PT, Ogden RT. Functional generalized linear models with images as predictors. *Biometrics.* **2010**;66:61–69.
- [20] Reiss P, Huo L, Zhao Y, et al. Wavelet-domain regression and predictive inference in psychiatric neuroimaging. *Ann Appl Stat.* **2015**;9:1076–11019.
- [21] Ciarleglio A, Ogden RT. Wavelet-based scalar-on-function finite mixture regression models. *Comput Stat Data Anal.* **2016**;93:86–96.
- [22] Zhao Y, Ogden R, Reiss P. Wavelet-based lasso in functional linear regression. *J Comput Graph Stat.* **2012**;21:600–617.
- [23] Wang X, Nan B, Zhu J, et al. Regularized 3d functional regression for brain image data via haar wavelets. *Ann Appl Stat.* **2014**;8(2):1045–1064.

- [24] Muraki S. Approximation and rendering of volume data using wavelet transforms. In: Kaufman A, Nielson G, editors. *Proceedings of Visualization*; Boston, MA; 1992. pp. 21–28. Los Alamitos (CA): IEEE Computer Society.
- [25] Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc B*. 1996;58:267–288.
- [26] Fukunaga K. *Introduction to statistical pattern recognition*. 2nd ed. San Diego (CA): Academic Press; 1990.
- [27] Jiang D, Huang J, Zhang Y. The cross-validated AUC for MCP-logistic regression with high-dimensional data. *Stat Methods Med Res*. 2013;22(5):505–518.
- [28] Zou H, Hastie T, Tibshirani R. On the “degrees of freedom” of the lasso. *Ann Stat*. 2007;35:2173–2192.
- [29] Zou H. *Some perspectives of sparse statistical modeling [dissertation]*. Department of Statistics, Stanford University; 2005.
- [30] Rabinovici GD, Furst AJ, Alkalay A, et al. Increased metabolic vulnerability in early-onset Alzheimer’s disease is not related to amyloid burden. *Brain*. 2010;133:512–528.
- [31] Minoshima S, Giordani B, Berent S, et al. Metabolic reduction in the posterior cingulate cortex in very early Alzheimer’s disease. *Ann Neurol*. 1997;42.