Journal of Neuroscience Methods 357 (2021) 109157

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Folded concave penalized learning of high-dimensional MRI data in Parkinson's disease

Changcheng Li [a], Xue Wang [b], Guangwei Du [c,e,*], Hairong Chen [c], Gregory Brown [c], Mechelle M. Lewis [c,d], Tao Yao [b], Runze Li [a,**], Xuemei Huang [c,d,e,f,g]

[a] Department of Statistics, Penn State University, University Park, PA, United States
[b] Alibaba DAMO Academy, Seattle, WA, United States
[c] Department of Neurology, Penn State Hershey Medical Center, Hershey, PA, United States
[d] Department of Pharmacology, Penn State Hershey Medical Center, Hershey, PA, United States
[e] Department of Radiology, Penn State Hershey Medical Center, Hershey, PA, United States
[f] Department of Neurosurgery, Penn State Hershey Medical Center, Hershey, PA, United States
[g] Department of Kinesiology, Penn State Hershey Medical Center, Hershey, PA, United States

ARTICLE INFO

Keywords:
Parkinson's disease
Magnetic resonance imaging
Diffusion tensor imaging
Folded concave penalized learning
Support vector machine
Logistic regression

ABSTRACT

Background: Brain MRI is a promising technique for Parkinson's disease (PD) biomarker development. Its analysis, however, is hindered by the high-dimensional nature of the data, particularly when the sample size is relatively small.

New Method: This study introduces a folded concave penalized machine learning scheme with spatial coupling fused penalty (fused FCP) to build biomarkers for PD directly from whole-brain voxel-wise MRI data. The penalized maximum likelihood estimation problem of the model is solved by local linear approximation.

Results: The proposed approach is evaluated on synthetic and Parkinson's Progression Marker Initiative (PPMI) data. It achieves good AUC scores, accuracy in classification, and biomarker identification with a relatively small sample size, and the results are robust for different tuning parameter choices. On the PPMI data, the proposed method discovers over 80 % of large regions of interest (ROIs) identified by the voxel-wise method, as well as potential new ROIs.

Comparison with Existing Methods: The fused FCP approach is compared with L1, fused-L1, and FCP method using three popular machine learning algorithms, logistic regression, support vector machine, and linear discriminant analysis, as well as the voxel-wise method, on both synthetic and PPMI datasets. The fused FCP method demonstrated better accuracy in separating PD from controls than L1 and fused-L1 methods, and similar performance when compared with FCP method. In addition, the fused FCP method showed better ROI identification.

Conclusions: The fused FCP method can be an effective approach for MRI biomarker discovery in PD and other studies using high dimensionality data/low sample sizes.

1. Introduction

Brain magnetic resonance imaging (MRI) is a promising technique for identifying biomarkers for diseases such as Parkinson's disease (PD). Performing analyses on MRI images without pre-defined regions of interest (ROIs) typically involves voxel numbers $\geq 10^6$ that are input variables on the voxel level. The number of human subjects, however, is small relative to the data dimension of MRI images ($n<<p$) Casanova et al. (2011); Fan et al. (2008); and Yasui et al. (2003). Reconstruction and selection of useful biomarkers from such high-dimensional data using limited sample sizes is challenging. Thus, there is a need in the biomedical research field for statistical learning methods that efficiently analyze such data.

Analysis of high dimensional data poses a non-trivial challenge to traditional approaches in statistical learning. It is common to assume data sparsity, meaning that only a small number of variables are relevant to the scientific problem under investigation even though there are a large number of variables available at the initial stage of modeling.

* Corresponding author at: Penn State Hershey Medical Center, 500 University Dr., H-037, Hershey, PA, 17033-0850, United States.
** Corresponding author at: Department of Statistics, Penn State University, University Park, PA 16802, United States.
E-mail addresses: guangweidu@pennstatehealth.psu.edu (G. Du), rzli@psu.edu (R. Li).

https://doi.org/10.1016/j.jneumeth.2021.109157
Received 1 July 2020; Received in revised form 17 March 2021; Accepted 22 March 2021
Available online 26 March 2021

Under the sparsity assumption, Lasso-type regularization has been proposed and successfully applied to many biomarker identification problems [e.g., Gu et al. (2013); and Wu et al. (2012)], and these methods have become the gold standard for high-dimensional learning Meinshausen and Bühlmann (2006); Meinshausen and Yu (2009); Van de Geer (2008); and Zhang and Huang (2008). A rich literature is devoted to exposing conditions under which the Lasso entails a theoretical guarantee of performance Bickel et al. (2009); Bunea et al. (2007); Cai et al. (2010); Candes and Tao (2007); Van De Geer and Bühlmann (2009); and Wainwright (2009). One of the main advantages of Lasso-type regularizations is that they allow efficient computation schemes to solve for the solution with global optimality. Lasso-type regularizations, however, can lead to extra estimation bias due to its statistical properties.

To address the drawback of Lasso-type approaches, a regularized method with folded concave penalty was proposed [FCP, Fan and Lv (2011)]. This later method enjoys desirable theoretical properties such as the unbiasedness and strong oracle property for high-dimensional sparse estimation. It has been shown to require weaker conditions and entail better statistical properties than Lasso Fan et al., 2014; Meinshausen and Bühlmann (2006); and Zou (2006). Two mainstream FCP functions are developed, i.e. the smoothly clipped absolute deviation [SCAD, Fan and Li (2001)] and minimax concave penalty [MCP, Zhang (2010)]. The FCP regularization method also has been applied successfully to many biomarker identification problems [e.g., Liu et al. (2016a, b)].

Another major challenge in extracting biomarkers from MRI data is the spatial correlation among adjacent voxels. The most popular approach to this problem has been univariate analysis (voxel-wise analysis) [e.g., Haynes and Rees (2006)]. Over the last several decades, however, multivariate statistical learning approaches have been applied in an increasing number of neuroimaging studies to model effects across multiple voxels Friston et al. (1995); Hanke et al. (2009). Compared to classic univariate analysis, these multivariate statistical learning approaches have shown better predictive performance. Spatial correlation also can be useful in guiding the construction of regions of interest (ROIs) [e.g., Friston et al. (1996)]. The classic methods usually separate ROI construction and model estimation, which may introduce extra noise in the model Tohka et al. (2016). As such, creating a unified model to simultaneously construct ROIs and estimate coefficients may improve analysis methods. Tibshirani et al. (2005) proposed using fused Lasso to explore a model with features that can be ordered in some meaningful way. The authors applied this method to analyses of protein mass spectroscopy and gene expression data with favorable results. And Lee et al. (2014) proposed a fused Lasso logistic regression approach to analyze corpus callosum thickness in early Alzheimer's disease.

We propose a fused MCP penalized approach to address the two major challenges in MRI data analysis: high-dimensionality and spatial correlation amongst voxels. In this paper, we aim to develop the novel approach as a reliable and improved method for identifying markers from high dimensional MRI data that also utilizes spatial correlations between adjacent voxels. In comparison with previous MRI biomarker studies using FCP penalized approaches such as Liu et al. (2016b), the newly proposed method does not rely on pre-defined ROIs but automatically constructs MRI metrics from the high-dimensional voxel-wise data by accommodating the fused penalty. This new penalty method also is compatible with many popular statistical/machine models [e.g., logistic regression, support vector machine (SVM), and local discriminate analysis (LDA)]. We first tested the model using synthetic data to verify the proposed approach successfully can recover MRI image signal using a sample size typical for most MRI research. We then applied the model to a real dataset (Parkinson's Progression Marker Initiative, PPMI) to determine its performance compared to other approaches.

2. Methods

We first propose a general classification framework for 3D MRI data using the fused MCP penalty (Section 2.1). We then introduce three popular classification methods, logistic regression, SVM, and LDA (Section 2.2), that can be used with the proposed fused MCP penalized method.

2.1. Fused MCP penalized method for MRI data classification

Suppose we have n MRI images $m_i, i = 1, 2, \ldots, n$. Each image contains three dimensional voxel data, where *dx*, *dy*, and *dz* are the digits in each dimension. Each image then can be represented by $x_{l,dx,dy,dz}$. We assume a classification model $f(x_{l,dx,dy,dz}, \beta_{dx,dy,dz})$ for the factor (disease) of interest, where f is the classification function, $\beta_{dx,dy,dz}$ is the parameter of the classification model associated with voxels of the image, and l is the subject index. Note that the parameter $\beta_{dx,dy,dz}$ of the classification model can provide information regarding the biomarker(s) related to the factor (disease) of interest. Assume function q measures the difference between prediction of the classification model $f(x_{l,dx,dy,dz}, \beta_{dx,dy,dz})$ and the true label y_l. The loss function L associated with the classification problem then is defined as follows:

$$L(\beta_{dx,dy,dz}) = \frac{1}{n}\sum_{l=1}^{n} q\left(f\left(x_{l,dx,dy,dz}, \beta_{dx,dy,dz}\right), y_l\right), \tag{1}$$

where y_l is the indicator of the factor (disease) of interest for subject l, so $y_l = 1$ for the patient, and $y_l = 0$ for the control, and we want to find $\beta_{dx,dy,dz}$ that minimizes the loss function L. Many statistical/machine learning models (e.g., logistic regression, SVM, LDA, etc.) can be formulated under the framework of (1), and some popular classification models and the corresponding loss functions are provided in Section 2.2.

To handle high-dimensional MR images, we need to add the regularization term. L1 and/or L2 regularization terms commonly are used in the literature [e.g., Tohka et al. (2016)]. However, L1 and/or L2 regularization terms will introduce extra estimation bias. For example, the classical L1 penalty penalizes all coefficients with the same amount of penalty. To achieve the variable selection consistency, the tuning parameter needs to surpass the order of root-*n* asymptotically, which can lead to an asymptotically noneligible bias in the estimation of the non-zero coefficients Zou, 2006. As a result, we proposed using MCP Zhang (2010) to address the bias issue. MCP is defined as follows:

$$p_\lambda(t) = \lambda \int_0^{|t|} \left(1 - \frac{x}{a\lambda}\right)_+ dx = \begin{cases} \frac{1}{2}a\lambda^2, & |t| > a\lambda, \\ \lambda t - \frac{1}{2a}t^2, & \text{otherwise}, \end{cases} \tag{2}$$

where a and λ are tuning parameters to control the shape of the penalty function. When the absolute value of t is large enough, $p_\lambda(t)$ becomes a constant. This means that the MCP approach does not penalize coefficients with absolute values over a certain threshold, and we can attain asymptotical unbiasness via MCP.

Besides the penalty for sparsity, we needed one more term to induce spatial constraints:

$$f_\rho(\beta) = \rho \sum_{dx,dy,dz} \Big(\left(\beta_{dx,dy,dz} - \beta_{dx\pm1,dy,dz}\right)^2 + \left(\beta_{dx,dy,dz} - \beta_{dx,dy\pm1,dz}\right)^2 + \left(\beta_{dx,dy,dz} - \beta_{dx,dy,dz\pm1}\right)^2 \Big) \tag{3}$$

where ρ is a positive constant. (3) promotes the adjacent voxels to have similar weights in the classification parameter $\beta_{dx,dy,dz}$. In our setting, this term works like a data-driven clustering function and automatically generates ROIs from the voxel-wise data. Our proposed method is to minimize:

$$\min_{\beta_{dx,dy,dz}} \left\{ \left(L(\beta_{dx,dy,dz})\right) + \sum_{i\in dx, j\in dy, k\in dz} p_\lambda(\beta_{i,j,k}) + f_\rho(\beta_{dx,dy,dz}) \right\} \quad (4)$$

The first term in (4) is a measure on how well the model distinguishes patients from controls. The second and third terms control the model complexity in sparsity and spatial similarity, respectively. Since brain has intrinsic substructures, it is common sense to assume that only parts of the whole brain associate with the target response and that adjacent voxels may have similar parameters in the classification model. In (4) we use the second and third terms to consider explicitly these assumptions when estimating the model in a data-driven manner. We use the tuning parameters a, λ, ρ to balance the trade-off between the goodness of fit and model complexity.

2.1.1. Solution scheme

Problem (4) is non-convex. Solving it globally can be very difficult Liu et al. (2016a). To deal with this issue, we first obtain an initial solution using the Lasso method, and then adopt the Local Linear Approximation (LLA) method from Zou and Li (2008); Fan et al. (2014) to obtain a one-step sparse estimate based on the initial Lasso solution. In sum, we have a two-step approximation scheme (2sAS) for the problem (4). More specifically, we first solve the following problem:

$$\min_{\beta_{dx,dy,dz}} \left\{ L(\beta_{dx,dy,dz}) + \lambda \sum_{i\in dx, j\in dy, k\in dz} |\beta_{i,j,k}| + f_\rho(\beta_{dx,dy,dz}) \right\} \quad (5)$$

Denote the first step solution as $\beta^{(1)}_{dx,dy,dz}$, and we then go on to solve the second step:

$$\min_{\beta_{dx,dy,dz}} \left\{ L(\beta_{dx,dy,dz}) + \sum_{i\in dx, j\in dy, k\in dz} p'_\lambda\left(\beta^{(1)}_{i,j,k}\right)|\beta_{i,j,k}| + f_\rho(\beta_{dx,dy,dz}) \right\}, \quad (6)$$

where $p'_\lambda(.)$ is the first order derivative of $p_\lambda(.)$. One may view the 2sAS as a refining procedure. In the first step, we approximate the FCP penalty function by the L1 penalty and solve only the L1 regularized problem. In the second step, we approximate the FCP penalty by the weighted L1 penalty and solve the weighted L1 regularized problem. The objective functions in both steps are convex and can be solved efficiently. The tuning parameter of the L1 regularization is changed adaptively according to the solution resulted from the first step. Recall that the absolute value of the first order derivative is correlated negatively with the magnitude of the input in the MCP case. If we use the value of the first order derivative of the previous step's solution, we will set a smaller penalty for more significant parts in the previous step's solution and vice versa. Hence, the 2sAS procedure will correct the bias in the initial Lasso-type solution. Fan et al. (2014) established the oracle optimality of the 2sAS procedure for the FCP penalized estimation.

2.1.2. Tuning parameter selection

In the proposed method, we introduce tuning parameters to control sparsity and spatial similarity. We used cross-validation for tuning parameter selection. Specifically, the dataset is separated randomly into two parts, a training set to train the model and a test dataset to test it. We repeat this procedure multiple times, and the average test accuracy is used to rank the tuning parameter choices. The one with the highest rank will be selected. Besides the cross-validation, the bootstrap and Bayesian error estimation approaches also can be applied in our method. Readers interested in those ideas can refer to Meinshausen and Bühlmann (2010) and Huttunen and Tohka (2015) for more details.

2.2. Classification methods

In this section, we introduce three popular classification methods: logistic regression, SVM, and LDA. The three methods need loss functions L, and the proposed fused MCP penalized method can be used as the loss function in all three methods for high dimensional data classification.

2.2.1. Logistic regression

Logistic regression has been popular for modeling the relationship between a binary response variable and a set of input (explanatory) variables. It can be used further for the prediction of binary responses based on newly observed input variables. Logistic regression has been applied to the study of disease diagnosis by Alkan et al. (2005) and Kennedy et al. (1996), among others. The current approach is distinguished from these prior studies, however, in that we deal with the issue of high dimensionality, given a large number of input variables but a small number of samples. Liao and Chin (2007) considered an under-sampled logistic regression and proposed a parametric bootstrap to reduce the model prediction error. In contrast, this study presents a substantially different statistical learning approach to handle the issue of high dimensionality. We assume the logistic model:

$$p_l = \frac{1}{1+\exp\left(-\sum_{i\in dx, j\in dy, k\in dz} x_{l,i,j,k}\beta_{i,j,k}\right)}, \quad (7)$$

where p_l is the probability that l-th MRI subject is associated with PD and the log-likelihood function is

$$L(\beta_{dx,dy,dz}) = \sum_l y_l \log(p_l) + \sum_l (1-y_l)\log(1-p_l), \quad (8)$$

where p_l is from (7). We then plug (8) back into (6) to fit the model.

2.2.2. Support vector machine (SVM)

SVM is another popular machine learning classification approach that has been applied successfully to many real-world neuroimaging problems [e.g., Othman et al. (2011) and Singh and Kaur (2012)]. Original SVM suffers from the high-dimensionality issue. In this section, we verify our approach also is capable of being used in the SVM method. We assume the loss function as follow:

$$L(\beta_{dx,dy,dz}, b) = \frac{1}{n}\sum_{l=1}^{n} \max\left(0, 1 - y'_l\left[\sum_{i\in dx, j\in dy, k\in dz} x_{l,i,j,k}\beta_{i,j,k} - b\right]\right), \quad (9)$$

where $y'_l = -1$ if $y_l = 0$, otherwise $y'_l = 1$. Once the model is learned, we may use the sign of $\sum_{i\in dx, j\in dy, k\in dz} x_{l,i,j,k}\beta_{i,j,k} - b$ to predict y'_l.

2.2.3. Linear discriminant analysis (LDA)

In addition to logistic regression and SVM, LDA is another popular classification method in statistics learning. LDA solves for the linear decision boundary in the underlying space that has the best discriminant ability among classes. In a low dimension classification problem, LDA performs well [e.g., Adeli et al. (2016)]. In the high dimensional setting, however, vanilla LDA leads to poor results Fan et al., 2012. We will explore the performance of a penalized LDA with our proposed penalties. The loss function of LDA is:

$$L(\beta_{dx,dy,dz}) = \frac{1}{n}\sum_{l=1}^{n}\left(\sum_{i\in dx, j\in dy, k\in dz} x_{l,i,j,k}\beta_{i,j,k}\right)^2, \quad (10)$$

subject to $\sum_{i\in dx, j\in dy, k\in dz}\mu_{i,j,k}\beta_{i,j,k} = 1$, where $\mu_{i,j,k} = \frac{1}{n}\sum_{l=1}^{n}\left[\left(x_{l,i,j,k}y_l = 1\right) - \left(x_{l,i,j,k}y_l = 0\right)\right]$. Problem (10) requires solving for a constrained learning problem but one also may consider the penalized version suggested in Fan et al. (2012).

3. Experiments

We first validated the proposed approach using synthetic datasets. The proposed procedure then was applied to real-world PPMI data that contains different types of MRI maps to separate PD patients from

controls.

3.1. Synthetic data

We generated a 2D plane (60 × 60) and set each point in the plane with a parameter $\beta_{I,j}$. The ROIs are characterized via $\beta_{I,j} \neq 0$. The subject is represented by a 60 × 60 matrix X. We considered two mechanisms for the generation of X.

- The first row and column are generated via a standard normal distribution and the remaining parts follow a linear relationship $x_{i+1,j+1} = 0.3x_{i+1,j} + 0.3x_{i,j+1} - 0.2x_{i,j} + \varepsilon$, where ε is the random error from a standard normal distribution, and $x_{i,j}$ is the element in matrix X induced by i-th row and j-th column.
- The first row and column are generated via the Rician distribution Gudbjartsson and Patz (1995), and the remaining parts follow a nonlinear relationship $x_{i+1,j+1} = 0.1x_{i+1,j}^2 + 0.1x_{i,j+1}^2 - 0.05x_{i,j}^2 + \varepsilon$, where ε is also from the Rician distribution. The probability density function of the Rician distribution is

$$f\left(x|v,\sigma\right) = \frac{x}{\sigma^2}\exp\left(\frac{-(x^2+v^2)}{2\sigma^2}\right)I_0\left(\frac{xv}{\sigma^2}\right),$$

where $I_0(z)$ is the modified Bessel function of the first kind with order zero, and v, σ are two non-negative parameters of the Rician distribution. In this paper, we use the Rician distribution with $v = 2$ and $\sigma = 1$.

We used logistic regression to construct the label for every subject. After generating the synthetic data, we then compared the prediction accuracy, ROI recovery, and ROI stability with respect to the tuning parameters of the proposed method (the fused MCP logistic model) against the L1\fused-L1\MCP models and voxel-wise analysis method.

3.2. Distinguishing PD patients from control subjects

After examining the performance of the proposed method on synthetic data, we compared the current approach with several popular methods in identifying PD patients in the PPMI dataset (Trojanowski, 2013). Details of the PPMI dataset are provided in subsections 3.2.1, 3.2.2, and 3.2.3. We selected eight different MRI feature maps including tensor-based morphometry (TBM), voxel-based morphometry (VBM), regional analysis of volumes examined in normalized space (RAVENS), and 5 diffusion tensor imaging (DTI) scalars: fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), radial diffusivity (RD), and free-water diffusion (FD). We assessed the performance of the proposed method by using three classification models: logistic regression, SVM, and LDA. For each model, we compared the Receiver Operating Characteristic (ROC) area under the curve (AUC) and accuracy scores of our method with L1/fused-L1/MCP penalties. In the experimental stage, we first randomly selected 155 of the 225 subjects as the training dataset. On each training dataset, we used 5-fold cross-validation for tuning parameter selection and fitted models with the selected tuning parameters on the training dataset. We propose a grid search for tuning parameters λ and ρ while fixing the value of a to reduce computation complexity. Here we fixed $a = 0.01$ and obtained relatively good numerical results, and we refer readers to Zhang (2010) and Kim et al. (2018) for guidance and heuristics for the selection of tuning parameters. Then we calculated AUC and accuracy scores using the test dataset that included the remaining 70 subjects. This procedure was replicated 1000 times. Mean AUC and classification accuracy scores are reported. We also compared the stability with respect to the tuning parameters of the proposed method against the L1\fused-L1\MCP methods.

3.3. Subjects

One hundred fifty-two PD patients and 73 controls (225 subjects in total) with high resolution T1-weighted and diffusion MRI data obtained using Siemens 3 T (TrioTim) scanners were downloaded from the PPMI website, along with demographic information including age, gender, and education (details in Table 1).

Table 1
Demographic data for subjects included from the PPMI dataset.

	Control(n = 73)	PD(n = 152)	P-values
Age, years (SD)	60.6 (10.8)	60.8 (9.7)	0.864
Sex, F/M	27/46	55/97	0.907
Education, years (SD)	15.6 (3.3)	15.4 (3.0)	0.587

3.4. MRI acquisition and image processing

T1-weighted (T1w) images were obtained with isotropic 1 mm spatial resolution. The DTI images were obtained with isotropic 2 mm spatial resolution (see https://www.ppmi-info.org/ for detailed MRI acquisition methods).

3.5. Generating whole brain voxel-wise MRI features

T1w images were processed for morphometric features. All images first were inspected visually to ensure that the image quality was sufficient for further analysis. The field inhomogeneity of T1w images was removed using the N4 algorithm (Tustison et al., 2010). A cohort-specific unbiased template was built on T1w images from all subjects using the buildtemplate script in ANTs v2.1. Three different morphometric features (VBM, TBM, and RAVENS) then were extracted using SPM 12, ANTs v2.1, and DRAMMS 1.4.1, respectively, following published methods Ashburner and Friston (2000); Ou et al. (2011); and Tustison et al. (2014). DTIPrep (Neuro Image Research and Analysis Laboratory, University of North Carolina, Chapel Hill, NC) was used for DTI image processing and feature extraction (Oguz et al., 2014). Fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), and radial diffusivity (RD) maps then were estimated. A two-pool diffusion model was built to estimate a free-water diffusion component (FD map) to represent free water of the interstitial fluid and non-directional diffusion components of cellular structure (Pasternak et al., 2009). DTI maps were co-registered to the T1w template using the FA map as a moving image through a non-rigid warping with ANTs v2.1 (Avants et al., 2008).

4. Results and discussion

In the current work, we proposed the fused MCP penalized method to identify biomarkers from MRI images that effectively address the challenge of high-dimensionality and spatial correlation amongst voxels. To assess whether the proposed method provides an advantage, we compared its performance to that of other methods using both synthetic data and real data from the PPMI study. We first tested the accuracy, biomarker identification, and stability with respect to tuning parameters of the proposed method using synthetic data. These analyses demonstrated the proposed method had good accuracy in group classification and biomarker identification with a relatively small sample size. We then applied our method to PPMI imaging data to test its classification accuracy and stability with respect to tuning parameters using different classification models and MRI maps. The results indicated that the proposed method had higher/comparable mean AUCs compared to the L1, fused-L1, and MCP classification models regardless of whether logistical regression (Table 2), SVM (Table 4), or LDA (Table 6) was used. In addition, mean accuracy also was higher/comparable using the proposed method (Tables 3, 5, and 7), and it performed better/comparably across the different MRI maps (Tables 2–7, Fig. 7). The classification performance of the proposed method also was robust with respect to the tuning parameter λ (Fig. 8), ρ (Fig. 9), and a (Fig. 10) in the PPMI imaging dataset.

Table 2
Mean AUC scores for the classification methods using logistic regression (LR) on the PPMI dataset.

	VBM	TBM	REVANS	DTI-FA	DTI-MD	DTI-AD	DTI-RD	DTI-FD
L1-LR	0.608(0.0024)	0.898(0.0005)	0.782(0.0029)	0.910(0.0006)	0.606(0.0018)	0.609(0.0018)	0.607(0.0018)	0.859(0.0015)
Fused L1-LR	0.761(0.0038)	0.900(0.0005)	0.826(0.0007)	0.913(0.0005)	0.624(0.0019)	0.626(0.0019)	0.623(0.0019)	0.854(0.0008)
MCP-LR	0.882(0.0007)	**0.942** (0.0003)	**0.923**(0.0004)	**0.953**(0.0003)	**0.812**(0.001)	**0.814**(0.001)	**0.817**(0.001)	**0.933**(0.0004)
Fused MCP-LR	**0.884**(0.0007)	**0.942**(0.0003)	**0.923**(0.0004)	**0.953**(0.0003)	0.811(0.001)	**0.814**(0.001)	0.815(0.001)	**0.933**(0.0004)

Mean AUC scores from 1000 replicates, with standard deviation listed in parenthesis. *Bold text indicates the best achieved results.*

Table 3
Mean classification accuracy for the classification methods using logistic regression (LR) on the PPMI dataset.

	VBM	TBM	REVANS	DTI-FA	DTI-MD	DTI-AD	DTI-RD	DTI-FD
L1-LR	0.567(0.0018)	0.790(0.0006)	0.681(0.0026)	0.809(0.0007)	0.586(0.0016)	0.590(0.0015)	0.585(0.0016)	0.735(0.0013)
Fused L1-LR	0.674(0.0032)	0.786(0.0006)	0.719(0.0008)	0.809(0.0006)	0.595(0.0015)	0.599(0.0014)	0.596(0.0015)	0.719(0.0009)
MCP-LR	0.824(0.0005)	**0.811**(0.0004)	0.817(0.0004)	**0.813**(0.0004)	**0.684**(0.0002)	**0.685**(0.0003)	**0.684**(0.0002)	**0.813**(0.0004)
Fused MCP-LR	**0.826**(0.0005)	**0.811**(0.0004)	**0.818**(0.0004)	**0.813**(0.0004)	**0.684**(0.0002)	**0.685**(0.0003)	0.683(0.0002)	**0.813**(0.0004)

Mean classification accuracy from 1000 replicates, with standard deviation listed in parenthesis. *Bold indicates the best achieved results.*

Table 4
Mean AUC scores for the classification methods using SVM on the PPMI dataset.

	VBM	TBM	REVANS	DTI-FA	DTI-MD	DTI-AD	DTI-RD	DTI-FD
L1-SVM	0.571(0.0016)	0.647(0.0012)	0.626(0.0016)	0.697(0.0009)	0.717(0.0008)	0.718(0.0008)	0.721(0.0008)	0.592(0.0013)
Fused L1-SVM	0.553(0.0015)	0.607(0.0011)	0.604(0.0016)	0.650(0.0009)	0.698(0.0008)	0.697(0.0007)	0.700(0.0007)	0.564(0.001)
MCP-SVM	**0.879**(0.0008)	**0.909**(0.0005)	0.891(0.0006)	0.930(0.0004)	0.846(0.0006)	**0.848**(0.0006)	0.850(0.0006)	0.909(0.0005)
Fused MCP-SVM	**0.879**(0.001)	0.908(0.0006)	**0.895**(0.001)	**0.934**(0.0006)	**0.847**(0.0007)	**0.848**(0.0007)	**0.852**(0.0007)	**0.910**(0.0008)

Mean AUC scores from 1000 replicates, with standard deviation listed in parenthesis. *Bold indicates the best achieved results*.

Table 5
Mean classification accuracy for the classification methods using SVM on the PPMI dataset.

	VBM	TBM	REVANS	DTI-FA	DTI-MD	DTI-AD	DTI-RD	DTI-FD
L1-SVM	0.555(0.0014)	0.589(0.001)	0.581(0.0014)	0.642(0.0011)	0.690(0.0004)	0.691(0.0005)	0.691(0.0005)	0.547(0.0011)
Fused L1-SVM	0.537(0.0013)	0.550(0.0009)	0.558(0.0013)	0.596(0.001)	0.630(0.0013)	0.634(0.0013)	0.634(0.0013)	0.523(0.0009)
MCP-SVM	**0.823**(0.0007)	0.830(0.0005)	**0.822**(0.0006)	0.844(0.0005)	**0.724**(0.001)	**0.727**(0.001)	**0.727**(0.001)	0.836(0.0005)
Fused MCP-SVM	0.820(0.001)	**0.834**(0.0005)	0.816(0.0009)	**0.845**(0.0007)	0.680(0.0000)	0.680(0.0000)	0.680(0.0000)	**0.837**(0.0008)

Clarification accuracy from 1000 replicates, with standard deviation listed in parenthesis. *Bold indicates the best achieved results.*

Table 6
Mean AUC scores for the classification methods using LDA on the PPMI dataset.

	VBM	TBM	REVANS	DTI-FA	DTI-MD	DTI-AD	DTI-RD	DTI-FD
L1-LDA	0.670(0.0011)	0.869(0.0009)	0.598(0.0018)	0.690(0.0039)	0.589(0.0036)	0.583(0.0037)	0.590(0.0036)	0.557(0.0037)
Fused L1-LDA	0.671(0.001)	0.865(0.0009)	0.601(0.0019)	0.695(0.0038)	0.591(0.0037)	**0.589**(0.0038)	**0.597**(0.0037)	0.563(0.0038)
MCP-LDA	**0.693**(0.0009)	0.501(0.0007)	0.499(0.0008)	0.499(0.0009)	0.569(0.0036)	0.569(0.0035)	0.570(0.0036)	0.500(0.0009)
Fused MCP-LDA	0.681(0.0012)	**0.890**(0.0008)	**0.740**(0.0024)	**0.803**(0.0027)	**0.592**(0.0037)	0.587(0.0038)	**0.597**(0.0037)	**0.687**(0.0035)

Mean AUC scores from 1000 replicates, with standard deviation listed in parenthesis. *Bold indicates the best achieved results*.

Table 7
Mean classification accuracy for the classification methods using LDA on the PPMI dataset.

	VBM	TBM	REVANS	DTI-FA	DTI-MD	DTI-AD	DTI-RD	DTI-FD
L1-LDA	0.685(0.0007)	0.715(0.0004)	0.580(0.0015)	0.658(0.0014)	**0.693**(0.0006)	0.693(0.0007)	0.692(0.0006)	0.549(0.0026)
Fused L1-LDA	0.684(0.0007)	**0.716**(0.0004)	0.582(0.0016)	0.660(0.0014)	**0.693**(0.0006)	**0.694**(0.0007)	**0.693**(0.0006)	0.552(0.0027)
MCP-LDA	0.676(0.0000)	0.501(0.0008)	0.499(0.0009)	0.499(0.0009)	0.550(0.0025)	0.550(0.0025)	0.550(0.0025)	0.500(0.0009)
Fused MCP-LDA	**0.692**(0.0007)	0.692(0.0003)	**0.683**(0.0024)	**0.672**(0.0017)	**0.693**(0.0006)	**0.694**(0.0007)	**0.693**(0.0006)	**0.644**(0.003)

Mean classification accuracy from 1000 replicates, with standard deviation listed in parenthesis. *Bold indicates the best achieved results.*

Table 8
Average ROI size identified by each penalized method in number of voxels.

	VBM	TBM	REVANS	DTI-FA	DTI-MD	DTI-AD	DTI-RD	DTI-FD
L1	20.5	1.3	46.1	3.0	1.0	1.0	1.0	2.7
Fused L1	122.2	4.1	75.4	3.3	1.0	1.0	1.0	9.1
MCP	52.1	17.1	39.2	15.8	886.7	910.8	730.2	16.7
Fused MCP	367.5	24.8	69.7	21.9	942.7	921.7	954.9	100.2

Table 9
Proportion of ROIs detected by the voxel-wise method that is larger than 12 voxels and also was identified by each penalized method.

	VBM	TBM	REVANS	DTI-FA	DTI-MD	DTI-AD	DTI-RD	DTI-FD
L1	22/33	2/70	59/81	0/94	0/86	0/97	0/85	2/87
Fused L1	33/33	7/70	67/81	5/94	0/86	0/97	0/85	23/87
MCP	25/33	58/70	53/81	84/94	84/86	94/97	81/85	58/87
Fused MCP	33/33	62/70	66/81	86/94	84/86	94/97	81/85	71/87

Table 10
Number of ROIs detected by the penalized methods that were >12 voxels and not identified by the voxel-wise method.

	VBM	TBM	REVANS	DTI-FA	DTI-MD	DTI-AD	DTI-RD	DTI-FD
L1	2	0	8	0	0	0	0	0
Fused L1	2	0	5	1	0	0	0	13
MCP	128	148	66	225	15	14	17	124
Fused MCP	2	176	80	232	16	17	13	154

We then compared the ROIs identified by the proposed fused MCP penalized method and other penalized methods with ones identified by the voxel-wise method (Tables 8–10). Note that for each penalized method (L1, fused L1, MCP, and fused MCP), we combined ROIs identified in the three classification methods (logistic regression, SVM, and LDA). The results indicated that the proposed method identifies larger ROIs on average than other penalized methods (Table 8), can identify a large proportion of ROIs identified by the voxel-wise method (Table 9), and can discover new ROIs (Table 10). Collectively, these empirical studies showed the advantage of the proposed fused MCP penalized method in accuracy of classification and biomarker identification under various situations.

The classification accuracy results on the synthetic dataset for the proposed fused MCP method and other methods under finite sample sizes are summarized in Fig. 1. In many MRI imaging studies, the typical sample size is ~20 and thus it was necessary to test the finite sample size performance of our method. The proposed approach reaches its best performance among benchmarks in the sample size ranging from 10 to 500 under both settings. Especially when the sample size is small (<100), the proposed approach is markedly better than the second-best method (on the order of 0.08−0.15 higher AUC scores). An interesting observation is that in the normal distribution setting, except for our proposed approach, the voxel-wise analysis attains the best performance when the sample size is <100; above this value, the voxel-wise analysis is surpassed by the fused-L1 method, and the fused-L1 method attains similar performance as the proposed fused-MCP method under the normal distribution setting with a large sample size of 500. We believe this phenomenon is due to the biasness issue of the fused-L1 estimator: under small sample sizes, the fused-L1 estimator has a large bias that significantly impairs the classification accuracy, whereas under larger sample sizes, the bias decreases and AUC scores eventually surpass the

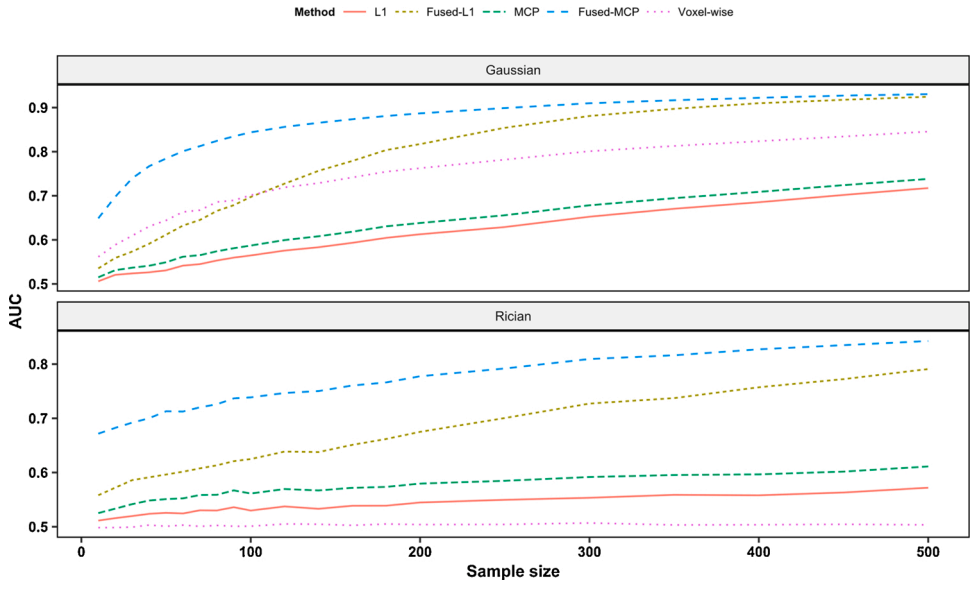

Fig. 1. AUC scores under different sample sizes using the synthetic dataset. The mean AUC scores of 1000 replicates are shown.

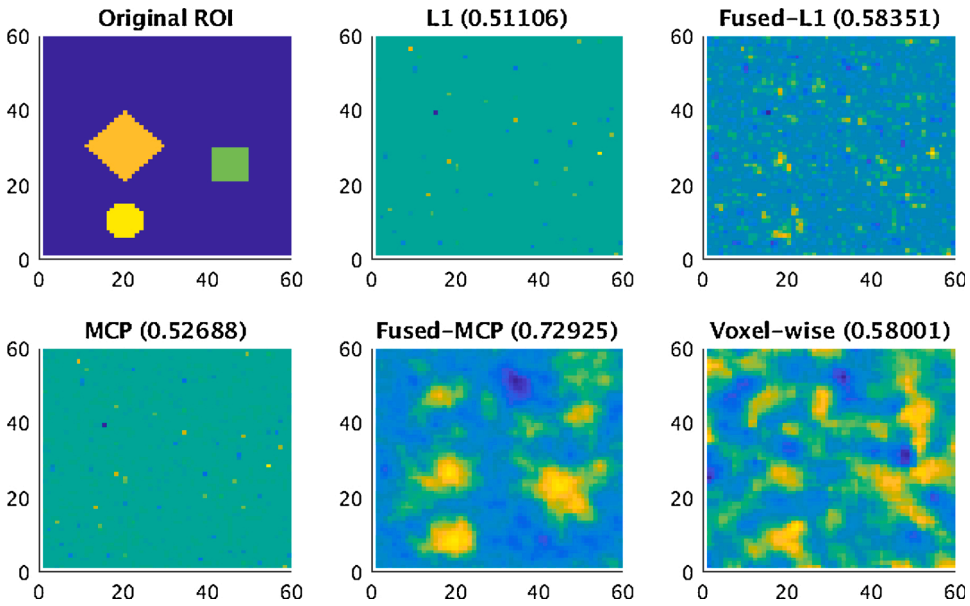

Fig. 2. Parameter recovery with 25 samples using the synthetic dataset generated using a normal distribution.

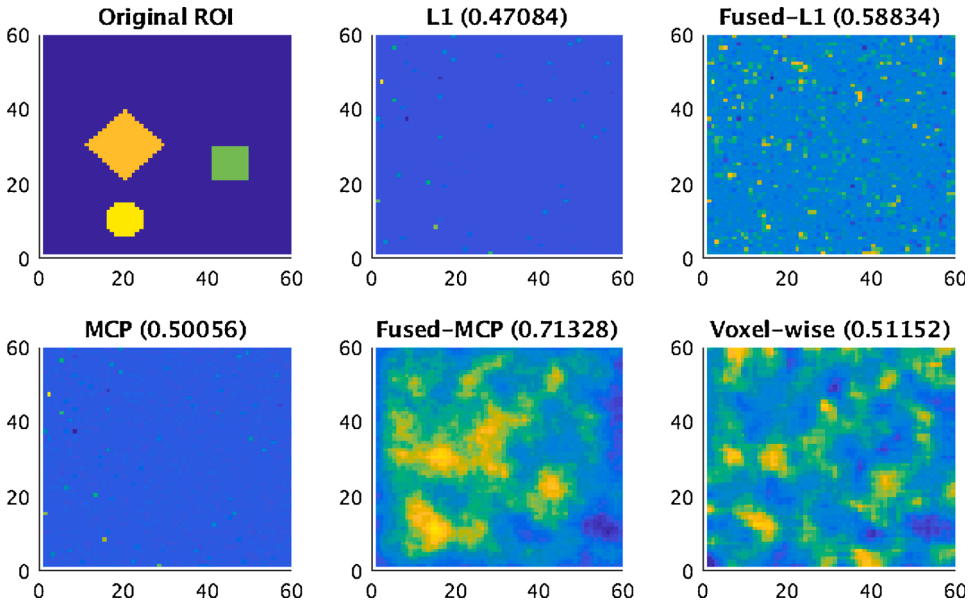

Fig. 3. Parameter recovery with 25 samples using the synthetic dataset generated using the Rician distribution.

classic voxel-wise analysis approach. Another observation is that the penalized and voxel-wise methods behave worse under the Rician distribution setting than the normal setting, especially the voxel-wise analysis. This suggests that the voxel-wise method is not as robust as the penalized method regarding the distribution of noise.

The accuracies of biomarker identification for the proposed method using the synthetic dataset under limited sample sizes are reported in Fig. 2 (for the normal distribution) and Fig. 3 (for the Rician distribution). We again chose a samples size ($n = 25$) that reasonably approximates MRI studies. Our proposed method successfully recovered all three ROIs under both distributions and was superior to the other approaches, which produced more ambiguous results. The proposed method also provided the highest biomarker identification accuracy (>70 %) compared to the other methods (47–59 %).

We next tested the stability of the proposed method with respect to tuning parameters. As shown in Fig. 4, the λ parameter ranged from 1/32 to 32 in a ratio of 2. The results indicated that the proposed method achieves the highest AUC values in this range of λ except for the cases of $\lambda \geq 16$ under the normal distribution. The proposed method also has a quite stable performance with $\lambda \leq 1$, and its performance only starts to decrease when $\lambda > 1$. The performance of the fused-L1 method also starts to decrease with large λ, and the decrease starts earlier than the proposed fused-MCP method. L1, MCP, and voxel-wise approaches are quite stable but attain lower AUC scores under most cases than the proposed method. To understand this, note that the L1 penalized method yields biased estimates. More specifically, every non-zero coefficient in the L1

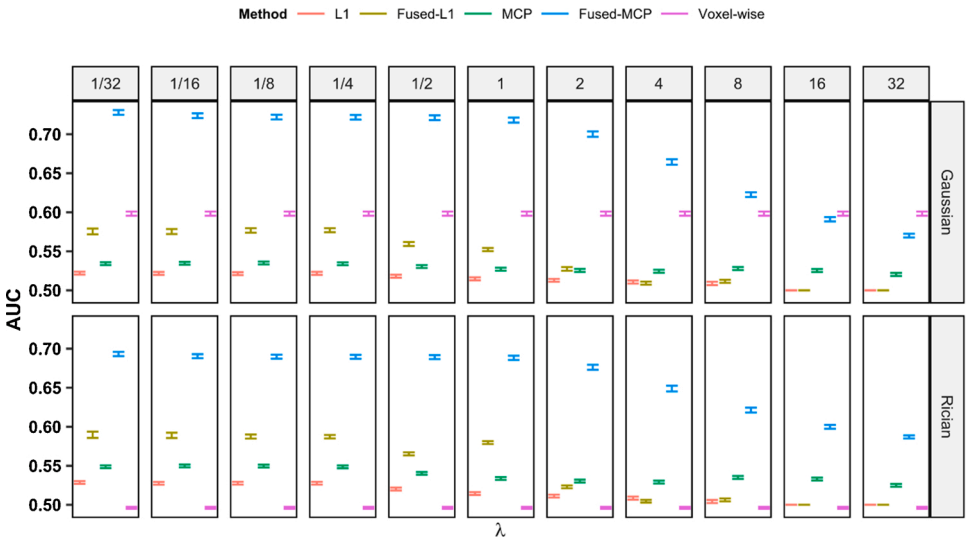

Fig. 4. AUC scores vs λ using the synthetic dataset. Methods in each block from left to right are L1, fused-L1, MCP, fused-MCP (the proposed method), and voxel-wise methods. The mean AUC scores of 1000 replicates are reported. 95 % confidence intervals are presented by error bars.

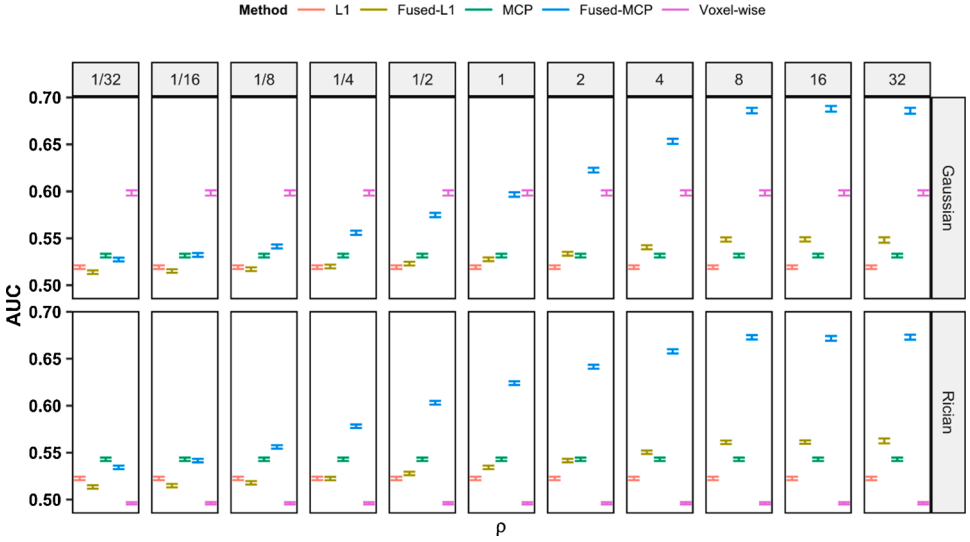

Fig. 5. AUC Scores vs ρ using the synthetic dataset. Methods in each block from left to right are L1, fused-L1, MCP, fused-MCP (the proposed method), and voxel-wise methods. The mean AUC scores of 1000 replicates are reported. 95 % confidence intervals are presented by error bars.

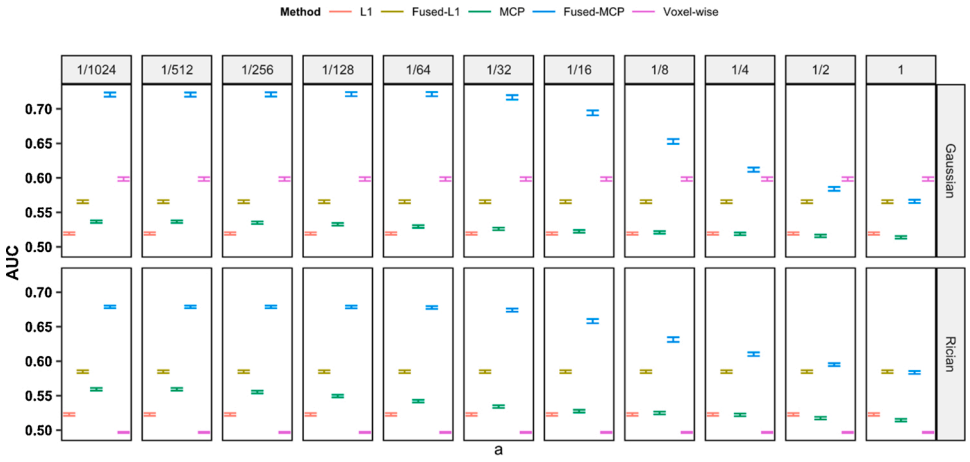

Fig. 6. AUC Scores vs *a* using the synthetic dataset. Methods in each block from left to right are L1, fused-L1, MCP, fused-MCP (the proposed method), and voxel-wise methods. The mean AUC scores of 1000 replicates are reported. 95 % confidence intervals are presented by error bars.

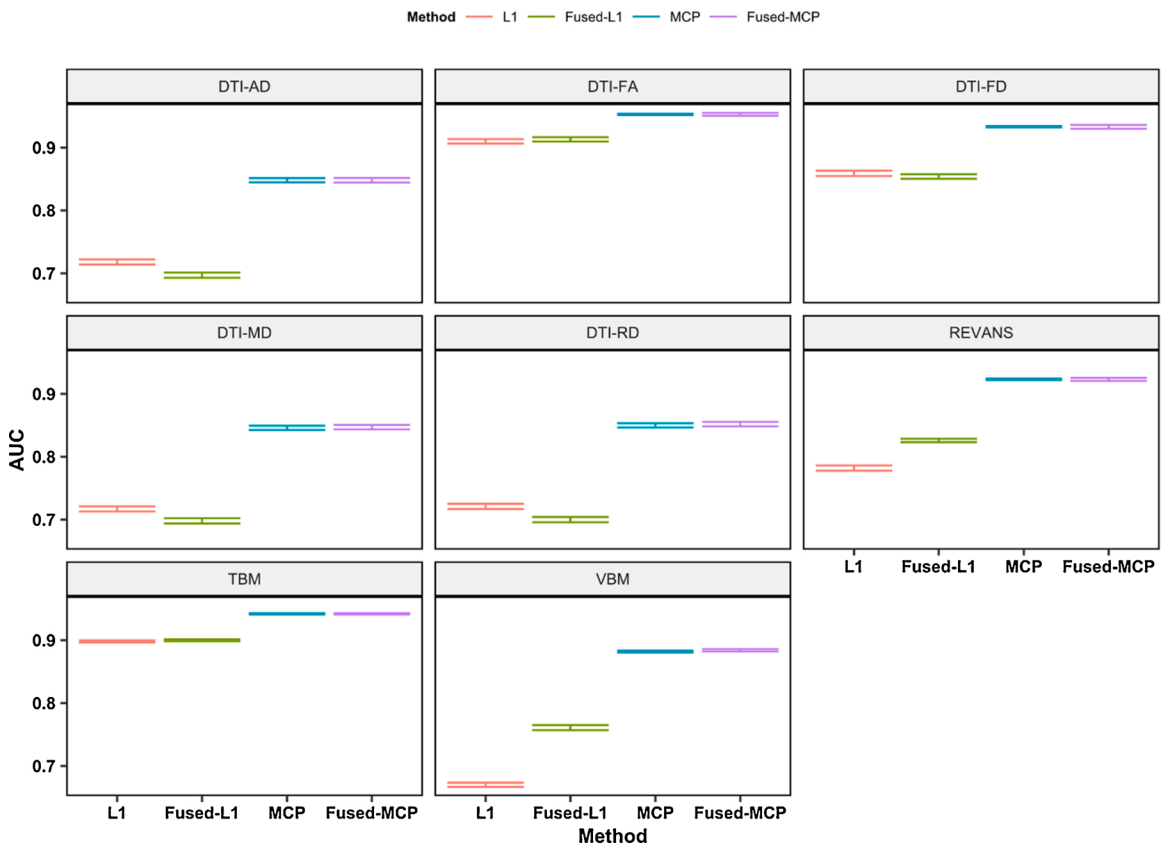

Fig. 7. Comparison of the best statistical performance among three classification models (logistic regression, SVM, and LDA) for each combination of MRI maps and penalized methods. Methods in each block from left to right are L1, fused-L1, MCP, and fused-MCP (the proposed method). The mean AUC scores of 1000 replicates are reported. 95 % confidence intervals are presented by error bars.

Fig. 8. Comparison of the statistical performance on SVM of different penalized method for each combination of MRI maps and tuning parameter λ. Methods in each block from left to right are L1, fused-L1, MCP, and fused-MCP (the proposed method). The mean AUC scores of 1000 replicates are reported. Error bars represent 95 % confidence intervals.

Fig. 9. Comparison of the statistical performance on SVM of different penalized method for each combination of MRI maps and tuning parameter ρ. Methods in each block from left to right are L1, fused-L1, MCP, and fused-MCP (the proposed method). The mean AUC scores of 1000 replicates are reported. Error bars represent 95 % confidence intervals.

Fig. 10. Comparison of the statistical performance on SVM of different penalized method for each combination of MRI maps and tuning parameter *a*. Methods in each block from left to right are L1, fused-L1, MCP, and fused-MCP (the proposed method). The mean AUC scores of 1000 replicates are reported. Error bars represent 95 % confidence intervals.

method is subject to the bias caused by λ. Since the fused-L1 method generally yields more non-zero coefficients than the non-fused L1 method, the fused-L1 method has a much larger bias in total than the non-fused L1 method. The performance of the fused-L1 also is affected by λ more than the non-fused L1 method. Since the MCP method solves the bias issue in the L1 method asymptotically, the fused-MCP method is not as sensitive as the fused-L1 method to the change in tuning parameter λ.

In Fig. 5, the ρ parameter also was varied from 1/32 to 32 in a ratio of 2. The proposed method achieved the highest AUC values when $\rho \geq 2$ under the normal distribution setting or $\rho \geq 1/8$ under the Rician distribution setting. The proposed method also attains a quite stable performance with $\rho \geq 8$. The fused methods, including fused-L1 and fused-MCP, need to use ρ large enough to utilize the spatial information, and the proposed fused-MCP approach utilized spatial information better than the fused-L1 approach as ρ increases.

In Fig. 6, the a parameter was varied from 1/1024 to 1 in a ratio of 2. The proposed method attains a stable performance with $a \leq 1/32$. This agrees with the fact that larger a may increase the finite sample bias of the MCP estimator (Zhang, 2010). A similar phenomenon also was present in the PPMI data (Fig. 10). Collectively, Figs. 4–6 show that the performance of the proposed approach is quite robust regarding the choice of tuning parameters within appropriate ranges.

When applying the proposed method to the PPMI data using logistical regression, our approach attained significant improvements over the L1/fused L1 methods and had a similar performance with the MCP logistic regression both in AUC scores and classification accuracy because of the bias of the L1/fused L1 methods (Tables 2 and 3). Note that this classification accuracy comparison of fused MCP versus non-fused MCP is different in the PPMI data from that in the synthetic data, where the fused MCP had markedly higher classification accuracy than the non-fused MCP. This may be due to different signal-to-noise ratios and different spatial structures of coefficient vectors in the synthetic and PPMI datasets. Similar results were observed when using SVM or LDA and in each case, the proposed approach achieved the best performance in most of the maps both in AUC and classification accuracy (Tables 4–7). The results of the proposed method were quite robust regarding tuning parameters λ, ρ, and small a, and Figs. 8–10 present the robustness of the proposed method on the SVM model as examples.

Regarding biomarker identification in the PPMI data, the fused methods identified larger ROIs on average than the corresponding non-fused methods because of the spatial information incorporated by the fused term. Also, with less bias than the L1 and fused L1 methods, the MCP and fused MCP methods can recover more signals and identify larger ROIs on average. In the PPMI MRI data, 12 voxels give a total volume of ~96 mm^3, which is close to a commonly used threshold for significant clusters (large ROIs) in neuroimaging studies. Compared to the voxel-wise method, the proposed fused MCP method identified >80 % of the large ROIs (>12 voxels) discovered by the voxel-wise method for all MRI maps, and the proportion was highest among the penalized methods in most of the MRI maps. This implies that both the proposed and classical voxel-wise methods discovered important ROIs in the PD group. Whereas the voxel-wise method treats each voxel individually and ignores any spatial information from the MRI maps, the proposed fused MCP method is based on classification models using all voxels and incorporates spatial information into the model through the fused term. Hence, it was expected that the proposed method would have different biomarker identification results from the classical voxel-wise method. It is also worth exploring the clinical meaning of the new ROIs discovered by the proposed fused MCP method.

MRI biomarker identification for PD, together with other biomarker development challenges, suffers from high-dimensionality issues: the subject number can be much smaller than the number of the voxels and there can be very little information on the shape of the biomarkers. Traditional statistical learning approaches become invalid facing high dimensionality. In this paper, we implemented for the first time the fused MCP penalized method for MRI biomarker discovery in PD. We demonstrated that it outperformed non-fused and fused L1 penalized methods and had comparable classification accuracy performance with the non-fused MCP penalized method for various classification models including SVM, PCA, and LDA. We also demonstrated the biomarker identification accuracy of the proposed method using a synthetic dataset. Furthermore, we compared the biomarker identification of the proposed method with the classical voxel-wise method using the PPMI dataset and found that the proposed method recovered >80 % of the large ROIs identified by the voxel-wise method, as well as some new ROIs. There may be great interest in the future for comparing the fused-MCP penalized method and other penalized methods together with deep learning regularization approaches for biomarker identification, such as deep neural networks using dropout (Srivastava et al., 2014). Further studies also are needed to validate the fused-MCP method in larger PD datasets and determine its applicability for biomarker discovery in other diseases.

CRediT authorship contribution statement

Changcheng Li: Software, Formal analysis, Writing - original draft, Writing - review & editing. **Xue Wang:** Software, Data curation, Writing - original draft, Writing - review & editing. **Guangwei Du:** Conceptualization, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Hairong Chen:** Software, Data curation. **Gregory Brown:** Writing - review & editing. **Mechelle M. Lewis:** Writing - review & editing. **Tao Yao:** Resources, Funding acquisition. **Runze Li:** Conceptualization, Methodology, Resources, Writing - review & editing, Funding acquisition. **Xuemei Huang:** Resources, Writing - review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

This work was supported in part by the National Institute of Neurological Disorders and Stroke (NS082151); the Hershey Medical Center General Clinical Research Center (National Center for Research Resources, grant UL1 RR033184, which is now at the National Center for Advancing Translational Sciences, grant UL1 TR002014); the National Science Foundation grants DMS 1820702, 1953196, and 2015539; the National Institute of Allergy and Infectious Diseases (R01-AI136664); the Pennsylvania Department of Health Tobacco CURE Funds; the Michael J. Fox Foundation for Parkinson's Research (10971, 18078); the Alzheimer's Association, Alzheimer's Research UK; and the Weston Brain Institute. Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org. We thank the investigators and teams associated with PPMI who collected the data and made them available to the public.

References

Adeli, E., Shi, F., An, L., Wee, C.-Y., Wu, G., Wang, T., Shen, D., 2016. Joint feature-sample selection and robust diagnosis of parkinson's disease from MRI data. Neuroimage 141, 206–219.

Alkan, A., Koklukaya, E., Subasi, A., 2005. Automatic seizure detection in EEG using logistic regression and artificial neural network. J. Neurosci. Methods 148 (2), 167–176.

Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry—the methods. Neuroimage 11 (6), 805–821.

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Issmage Anal 12 (1), 26–41.

Bickel, P.J., Ritov, Y.A., Tsybakov, A.B., 2009. Simultaneous analysis of lasso and dantzig selector. Ann. Stat. 37 (4), 1705–1732.

Bunea, F., Tsybakov, A., Wegkamp, M., 2007. Sparsity oracle inequalities for the Lasso. Electron. J. Stat. 1, 169–194.
Cai, T., Wang, L., Xu, G., 2010. Stable recovery of sparse signals and an oracle inequality. IEEE Trans. Inf. Theory 56 (7), 3516–3522.
Candes, E., Tao, T., 2007. The dantzig selector: statistical estimation when p is much larger than n. Ann. Stat. 35 (6), 2313–2351.
Casanova, R., Wagner, B., Whitlow, C.T., Williamson, J.D., Shumaker, S.A., Maldjian, J. A., Espeland, M.A., 2011. High dimensional classification of structural MRI alzheimer's disease data based on large scale regularization. Front. Neuroinf. 5, 22.
Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. J. Am. Stat. Assoc. 96 (456), 1348–1360.
Fan, J., Lv, J., 2011. Nonconcave penalized likelihood with NP-dimensionality. IEEE Trans. Inf. Theory 57 (8), 5467–5484.
Fan, Y., Resnick, S.M., Wu, X., Davatzikos, C., 2008. Structural and functional biomarkers of prodromal alzheimer's disease: a high-dimensional pattern classification study. Neuroimage 41 (2), 277–285.
Fan, J., Feng, Y., Tong, X., 2012. A road to classification in high dimensional space: the regularized optimal affine discriminant. J. R. Stat. Soc. Series B 74 (4), 745–771.
Fan, J., Xue, L., Zou, H., 2014. Strong oracle optimality of folded concave penalized estimation. Ann. Stat. 42 (3), 819–849.
Friston, K.J., Frith, C.D., Frackowiak, R.S., Turner, R., 1995. Characterizing dynamic brain responses with fMRI: a multivariate approach. Neuroimage 2 (2), 166–172.
Friston, K.J., Holmes, A., Poline, J.-B., Price, C.J., Frith, C.D., 1996. Detecting activations in PET and fMRI: levels of inference and power. Neuroimage 4 (3), 223–235.
Gu, X., Yin, G., Lee, J.J., 2013. Bayesian two-step Lasso strategy for biomarker selection in personalized medicine development for time-to-event endpoints. Contemp. Clin. Trials. 36 (2), 642–650.
Hanke, M., Halchenko, Y.O., Sederberg, P.B., Hanson, S.J., Haxby, J.V., Pollmann, S., 2009. PyMVPA: a python toolbox for multivariate pattern analysis of fMRI data. Neuroinformatics 7 (1), 37–53.
Haynes, J.-D., Rees, G., 2006. Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. 7, 523–534.
Huttunen, H., Tohka, J., 2015. Model selection for linear classifiers using Bayesian error estimation. Pattern Recognit. 48 (11), 3739–3748.
Kennedy, R.L., Fraser, H.S., McStay, L.N., Harrison, R.F., 1996. Early diagnosis of acute myocardial infarction using clinical and electrocardiographic data at presentation: derivation and evaluation of logistic regression models. Eur. Heart J. 17 (8), 1181–1191.
Kim, D., Lee, S., Kwon, S., 2018. A Unified Algorithm for the non-Convex Penalized Estimation: The Ncpen Package. *arXiv preprint arXiv:1811.05061*.
Lee, S.H., Yu, D., Bachman, A.H., Lim, J., Ardekani, B.A., 2014. Application of fused lasso logistic regression to the study of corpus callosum thickness in early alzheimer's disease. J. Neurosci. Methods 221, 78–84.
Liao, J.-G., Chin, K.-V., 2007. Logistic regression for disease classification using microarray data: model selection in a large p and small n case. Bioinformatics 23 (15), 1945–1951.
Liu, H., Yao, T., Li, R., 2016a. Global solutions to folded concave penalized nonconvex learning. Ann. Stat. 44 (2), 629–659.
Liu, H., Du, G., Zhang, L., Lewis, M.M., Wang, X., Yao, T., Li, R., Huang, X., 2016b. Folded concave penalized learning in identifying multimodal MRI marker for parkinson's disease. J. Neurosci. Methods 268, 1–6.
Meinshausen, N., Bühlmann, P., 2006. High-dimensional graphs and variable selection with the Lasso. Ann. Stat. 34 (3), 1436–1462.
Meinshausen, N., Bühlmann, P., 2010. Stability selection. J. R. Stat. Soc. Series B 72 (4), 417–473.
Meinshausen, N., Yu, B., 2009. Lasso-type recovery of sparse representations for high-dimensional data. Ann. Stat. 37 (1), 246–270.
Oguz, I., Farzinfar, M., Matsui, J., Budin, F., Liu, Z., Gerig, G., et al., 2014. DTIPrep: quality control of diffusion-weighted images. Front. Neuroinf. 8, 4.
Othman, M.F.B., Abdullah, N.B., Kamal, N.F.B., 2011.). MRI brain classification using support vector machine. In: Paper Presented at the 2011 Fourth International Conference on Modeling. Simulation and Applied Optimization.
Ou, Y., Sotiras, A., Paragios, N., Davatzikos, C., 2011. DRAMMS: deformable registration via attribute matching and mutual-saliency weighting. Med. Image Anal. 15 (4), 622–639.
Pasternak, O., Sochen, N., Gur, Y., Intrator, N., Assaf, Y., 2009. Free water elimination and mapping from diffusion MRI. Magn. Reson. Med. Sci. 62 (3), 717–730.
Singh, D., Kaur, K., 2012. Classification of abnormalities in brain MRI images using GLCM, PCA and SVM. Int. J. Eng. Adv. Technol. 1 (6), 243–248.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.
Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., Knight, K., 2005. Sparsity and smoothness via the fused lasso. J. R. Stat. Soc. Series B 67 (1), 91–108.
Tohka, J., Moradi, E., Huttunen, H., Alzheimer's Disease Neuroimaging Initiative, 2016. Comparison of feature selection techniques in machine learning for anatomical brain MRI in dementia. Neuroinformatics 14 (3), 279–296.
Trojanowski, J.Q., 2013. Parkinson's progression markers initiative (PPMI). J. Neuropathol. Exp. Neurol. 72 (6), 540–598.
Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging 29 (6), 1310–1320.
Tustison, N.J., Cook, P.A., Klein, A., Song, G., Das, S.R., Duda, J.T., 2014. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. Neuroimage 99, 166–179.
Van de Geer, S.A., 2008. High-dimensional generalized linear models and the lasso. Ann. Stat. 36 (2), 614–645.
Van dsse Geer, S.A., Bühlmann, P., 2009. On the conditions used to prove oracle results for the lasso. Electron. J. Stat. 3, 1360–1392.
Wainwright, M.J., 2009. Sharp thresholds for high-dimensional and noisy sparsity recovery using $\ell 1$-constrained quadratic programming (lasso). IEEE Trans. Inf. Theory 55 (5), 2183–2202.
Wu, M.-Y., Dai, D.-Q., Shi, Y., Yan, H., Zhang, X.-F., 2012. Biomarker identification and cancer classification based on microarray data using laplace naive bayes model with mean shrinkage. IEEE/ACM Trans. Comput. Biol. Bioinf. 9 (6), 1649–1662.
Yasui, Y., Pepe, M., Thompson, M.L., Adam, B.L., Wright Jr, G.L., Qu, Y., et al., 2003. A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection. Biostatistics 4 (3), 449–463.
Zhang, C.-H., 2010. Nearly unbiased variable selection under minimax concave penalty. Ann. Stat. 38 (2), 894–942.
Zhang, C.-H., Huang, J., 2008. The sparsity and bias of the lasso selection in high-dimensional linear regression. Ann. Stat. 36 (4), 1567–1594.
Zou, H., 2006. The adaptive lasso and its oracle properties. J. Am. Stat. Assoc. 101 (476), 1418–1429.
Zou, H., Li, R., 2008. One-step sparse estimates in nonconcave penalized likelihood models. Ann. Stat. 36 (4), 1509–1533.