Identifying brain hierarchical structures associated with Alzheimer’s disease using a regularized regression method with tree predictors

Yi Zhao | Bingkai Wang | Chin-Fu Liu | Andreia V. Faria | Michael I. Miller | Brian S. Caffo | Xi Luo

1Department of Biostatistics and Health Data Science, Indiana University School of Medicine, Indianapolis, Indiana, USA
2Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
3Center for Imaging Science, Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, USA
4Department of Radiology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
5Department of Biostatistics and Data Science, The University of Texas Health Science Center at Houston, Houston, Texas, USA

Correspondence
Yi Zhao, Department of Biostatistics and Health Data Science, Indiana University School of Medicine, Indianapolis, IN, USA.
Email: yz125@iu.edu

Abstract
Brain segmentation at different levels is generally represented as hierarchical trees. Brain regional atrophy at specific levels was found to be marginally associated with Alzheimer’s disease outcomes. In this study, we propose an \( \ell_1 \)-type regularization for predictors that follow a hierarchical tree structure. Considering a tree as a directed acyclic graph, we interpret the model parameters from a path analysis perspective. Under this concept, the proposed penalty regulates the total effect of each predictor on the outcome. With regularity conditions, it is shown that under the proposed regularization, the estimator of the model coefficient is consistent in \( \ell_2 \)-norm and the model selection is also consistent. When applied to a brain sMRI dataset acquired from the Alzheimer’s Disease Neuroimaging Initiative (ADNI), the proposed approach identifies brain regions where atrophy in these regions demonstrates the declination in memory. With regularization on the total effects, the findings suggest that the impact of atrophy on memory deficits is localized from small brain regions, but at various levels of brain segmentation. Data used in preparation of this paper were obtained from the ADNI database.

KEYWORDS
hierarchical predictors, path analysis, penalized linear models, structural neuroimaging, tree-based regularization

1 | INTRODUCTION

In the problem of linear regression with high-dimensional data, \( \ell_1 \)-regularization and its variations are ubiquitously studied. Well-known examples include the lasso (Tibshirani, 1996), the elastic net (Zou & Hastie, 2005), the adaptive lasso (Zou, 2006), the fused lasso (Tibshirani et al., 2005), the group lasso, GL (Yuan & Lin, 2007), the generalized lasso (Tibshirani & Taylor, 2011), and many others. Among these, the fused lasso, the GL, and the generalized lasso accommodate certain structural information in the design matrix to achieve desired profiles of the model coefficients. Motivated by genomic data, considering the network information of the predictors as a priori, Li and Li (2008) and Pan et al. (2010) introduced network-based regularization approaches to conduct variable selection. In the studies, the priori networks are undirected graphs and no hierarchy between variables is assumed. Embedding the compositional nature of microbiome data, Wang and Zhao (2017) and Wang et al. (2017)
introduced tree-guided regularizations to linear regression. The connections between the variables (bacterial taxa) were defined by a phylogenetic tree, where only the leaf nodes are observable and the internal nodes are the conceptual taxonomic levels. Recently, Yan and Bien (2021) considered a tree-guided regularization as well to aggregate rare features to improve the performance of prediction, where the trees are predefined from prior knowledge and/or external data sources. Again, only leaf nodes are observed and aggregated to construct denser features at higher levels of the tree. For neuroimaging studies, Liu et al. (2014) introduced an approach that first employs the hierarchical agglomerative clustering on voxel-level brain structural data to construct a binary tree and then applied overlapping GL regularization for feature selection and ultimately sample classification. In this study, we consider a scenario where the predictors possess a hierarchical tree structure and data at all levels of the tree are observable. Treating the hierarchical tree as a directed acyclic graph, we interpret the model parameters from a path analysis perspective and propose a lasso-type penalty that regulates the total effect of the predictor on the outcome.

This is motivated by structural magnetic resonance imaging (sMRI) studies. With the availability of high-quality 3D images, it is now possible to measure brain morphometry and investigate associations with mental disorders. For example, in the study of Alzheimer’s disease (AD), sMRI is considered as a direct reflection of the density of neurofibrillar tangles, an established pathological hallmark of AD. It captures atrophy in gray matter due to the loss of neurons, synapses, and dendritic de-\emph{}arborization (atrophy in white matter due to the loss of structural integrity of white matter fiber tracts presumably a consequence of demyelination and dying back of axonal processes) and an \emph{ex vacuo} (increases in the volume of the cerebral spinal fluid (CSF) caused by the loss of encephalic volume). Thus, measurements acquired from sMRI have been widely used to identify (regional) markers of AD (Vemuri & Jack, 2010). In order to extract regional structural data, anatomical brain segmentation is generally applied. Multi-atlas segmentation (MAS) is a popular approach, which has the advantage of coordinating representations from multiple segmented atlases and correcting errors through a label fusion process. Several MAS approaches offer hierarchical segmentations at various granularity levels (Djamanakova et al., 2014; Wu et al., 2015; Doshi et al., 2016). Djamanakova et al. (2014) introduced a segmentation that has a five-level hierarchical structure, starting from a coarse segmentation (Level 1) into major areas (telencephalon, diencephalon, metencephalon, mesencephalon, and CSF) to a fine segmentation (Level 5) defining hundreds of structures, as small as gyri and deep nucleae. The hierarchical tree structure of this multi-level segmentation is presented in Figure B.1. When studying the association with AD symptoms, such as memory decline, the granularity level of the region of interests (ROIs) that play a role may diverge across the brain. For example, the hippocampus, which is a Level-4 region in the segmentation, is a consistently identified brain region related to memory deficits in all stages of AD (Pini et al., 2016). Another well-known marker area is the entorhinal cortex (Pini et al., 2016), a Level-5 region. Both the hippocampus and entorhinal cortex are part of the limbic system, which is a Level-3 region. Thus, it is beneficial to include data extracted from all levels of segmentation into feature selection and, in the meantime, taking the hierarchical structure and data dependence into consideration.

Considering the tree structure as a directed acyclic graph, we introduce an $\ell_1$-type regularization, which incorporates the structural information by taking the influence matrix of the predictors as the penalty matrix. The influence matrix can be either obtained from the weighted adjacency matrix or estimated from the data following the hierarchical structural specification. Using a path diagram, we demonstrate that this is equivalent to regularizing the total effect of each predictor on the outcome. Different from imposing variable fusion, the proposed approach selects variables with significant total effect. In sMRI studies, the assumption that brain regions on the same tree branch have a similar impact on the outcome may not hold. For example, considering the Level-3 regions on the branch of the cerebral cortex, structural atrophy occurs following the trajectory of medial temporal (the limbic system)–temporal–parietal–frontal as AD progresses, while less evidence suggests the association between occipital atrophy and AD. Thus, constancy among these regions cannot be assumed. The total effect can be interpreted in the sense of an aggregation of the effects from the child nodes (as well as the marginal effect of itself). Regularizing the total effect of each node enables the variable selection across various hierarchies on the tree and at the same time identifies the aggregated effect.

The rest of the paper is organized as follows. In Section 2, we introduce regularization functions for data following a hierarchical tree structure and show that (under regularity conditions) the regularized estimator of the model parameter and the estimated active set are both consistent. Section 3 presents the simulation results demonstrating the performance of the proposed regularizations under various scenarios. In Section 4, we apply the proposed approach to the data collected in the Alzheimer’s Disease Neuroimaging Initiative (ADNI). The goal is to investigate the association between atrophy in the brain and memory deficits, where the brain volumetric data are extracted from a hierarchical segmentation. Using the proposed
regularization function, the identified relevant brain regions are in line with existing literature. In addition, the findings infer local impacts of atrophy on memory decline. Section 5 summarizes the paper with a discussion.

2 | METHODS

In our application, to study the association between brain volume and memory among an elder population with mild cognitive impairment (MCI) or AD, a measurement of memory is considered as the outcome and the multi-level brain volumetric data with a hierarchical tree structure are considered as the predictors. Let \( \mathbf{Y} = (y_1, ..., y_p)^\top \in \mathbb{R}^p \) denote the outcome vector of \( n \) subjects and \( \mathbf{X} = (x_1, ..., x_p)^\top \in \mathbb{R}^{n \times p} \) the \( p \)-dimensional design matrix. The following linear regression model is considered:

\[
\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon, \tag{1}
\]

where \( \boldsymbol{\beta} \in \mathbb{R}^p \) is a \( p \)-dimensional vector of model coefficients, and elements in \( \epsilon \in \mathbb{R}^n \) are the model errors, assumed to be independent and identically distributed following a normal distribution with mean zero and variance \( \sigma^2 \). Different from a regular design matrix, the columns of \( \mathbf{X} \) possess a directed hierarchical tree structure. For the brain volumetric data, the tree structure is determined based on the segmentation at various granularity levels, where a brain region at a coarser segmentation level is divided into multiple regions at a finer level. We propose to model data with such type of property using the following formulation:

\[
\mathbf{X} = \mathbf{X}\mathbf{A} + \epsilon, \tag{2}
\]

where \( \mathbf{A} = (a_{ij}) \in \mathbb{R}^{p \times p} \) is the weighted adjacency matrix of the predictors and \( \epsilon \in \mathbb{R}^{n \times p} \) is an error matrix. Without loss of generality, we assume that the predictors are centered with mean zero. Each row of \( \epsilon \) follows a \( p \)-dimensional normal distribution with covariance matrix, \( \Omega \), where \( \Omega \) is a diagonal matrix. The dependency between the predictors is fully captured by the adjacency matrix. Since a directed hierarchical tree is directed and acyclic, \( \mathbf{A} \) is an upper triangular matrix with diagonal elements zero when the \( \mathbf{X} \)'s are properly ordered. For the brain volumetric data, the element of the upper triangular part of the adjacency matrix, \( \mathbf{A} \), is determined by the segmentation ratio, which can be estimated by the ratio of the average volume of the two corresponding regions from the data. For example, the temporal lobe is a Level-3 region (indexed by \( j \)) and is divided into four Level-4 regions, including the fusiform gyrus (indexed by \( j_k \)), the inferior (\( j_k \)), middle (\( j_k \)), and superior (\( j_k \)) temporal gyrus. The \( (j, j_k) \) element in \( \mathbf{A} \) is then estimated by the ratio of the average volume of the fusiform gyrus and the average volume of the temporal lobe. Model (2) has been used to study interactions among genes and/or proteins (Shojaie & Michailidis, 2009, 2010). Here, we adopt it to portrait the hierarchical structure of the brain volumetric data.

**Definition 1.** Consider a tree, \( \mathcal{T} = \{ \mathbf{X}, \mathbf{A}\} \), defined by (2), where \( \mathbf{X} = (X_1, ..., X_p)^\top \in \mathbb{R}^p \) are the nodes in the tree and \( \mathbf{A} = (a_{ij}) \in \mathbb{R}^{p \times p} \) is the adjacency matrix of \( \mathbf{X} \). We say that \( X_i \) is a parent of \( X_j \) if \( a_{ij} \neq 0 \), and, naturally, if \( X_i \) is a parent of \( X_j \), \( X_j \) is a child of \( X_i \), for \( i, j = 1, ..., p \) and \( i \neq j \). Let \( \mathcal{P}_j = \{ i : a_{ij} \neq 0, i = 1, ..., p \} \) denote the set of parent nodes of \( X_j \) and \( \mathcal{C}_j = \{ k : a_{jk} \neq 0, k = 1, ..., p \} \) denote the set of child nodes of \( X_j \), for \( j = 1, ..., p \). If \( \mathcal{P}_j = \emptyset \), \( X_j \) is called a root node. If \( \mathcal{C}_j = \emptyset \), \( X_j \) is called a leaf node (or terminal node). If a node has both parent and child nodes, that is, \( \mathcal{P}_j \neq \emptyset \) and \( \mathcal{C}_j \neq \emptyset \), then it is called an internal node.

Figure 1A shows the tree structure of a toy example with \( p = 6 \) predictors. In this example, \( X_1 \) is a root node, \( X_4, X_5 \) and \( X_6 \) are leaf nodes, and \( X_2 \) and \( X_3 \) are internal nodes. From Model (2),

\[
\mathbf{X} = \epsilon(\mathbf{I} - \mathbf{A})^{-1}. \tag{3}
\]

Let \( \mathbf{D} = (\mathbf{I} - \mathbf{A})^{-1} = (d_{ij}) \in \mathbb{R}^{p \times p} \), which is also an upper triangular matrix, but with diagonal elements equal to one. The matrix \( \mathbf{D} \) is called the influence matrix (Shojaie & Michailidis, 2009), where the value of \( d_{ij} \) quantifies the overall influence of \( X_i \) on \( X_j \) if \( X_i \) is a parent of \( X_j \). Under (3), the covariance matrix of \( \mathbf{x}_i \) is \( \mathbf{D}^\top \Omega \mathbf{D} \) (for \( i = 1, ..., n \)). As \( \Omega \) is a diagonal matrix, this demonstrates that the dependencies among the \( p \) predictors are fully characterized through \( \mathbf{D} \). In the motivated sMRI example, the hierarchical brain segmentation in Figure B.1 has one tree, where the Level-0 root is the root node. The direction is from Level 0 to Level 5 as the child regions are a segmentation of the parent region with potential measurement errors.

In Model (1), \( \hat{\beta}_j \) is interpreted as the effect of \( X_j \) on \( Y \) conditional on the rest predictors, or the direct effect \( X_j \) on \( Y \).

**Definition 2.** Under Model (1), for \( j = 1, ..., p \),

\[
\hat{\beta}_j = 0 \quad X_j \not\perp \!
\!
\!
\perp \mathbf{Y}|\{X_1, ..., X_{j-1}, X_{j+1}, ..., X_p\}. \tag{4}
\]

With a tree structure, the total effect of a node can also be of great interest, where the total effect accounts for all the possible path effects to the outcome. For example, in the ADNI application, for a root/internal brain region at...
higher levels, the existence of a total effect may suggest a global impact of atrophy on the declination of memory. Otherwise, if only the leaf nodes have a total/direct effect, the impact is localized. Therefore, we introduce the following regularization criterion to estimate the parameters:

\[ \hat{\beta} = \arg \min_{\beta \in \mathbb{R}^p} \frac{1}{2} \| Y - X\beta \|^2_2 + \lambda (\| D\beta \|_1 + \alpha \| \beta \|_1), \]  

where \( \lambda, \alpha \geq 0 \in \mathbb{R} \) are the tuning parameters. Note that the regularization term above can be written as

\[ D = \left( \begin{array}{c} D \\ \alpha I_p \end{array} \right) \in \mathbb{R}^{2p \times p}, \]

where \( I_p \) is the \( p \)-dimensional identity matrix. This can be solved by the so-called generalized lasso (Tibshirani & Taylor, 2011). In practice, \( D \) can be predefined based on domain knowledge or estimated from the data. As the tree structure is defined, by properly ordering the columns of \( X \), \( D \) can be estimated by obtaining the covariance or precision matrix of \( X \) first followed by a Cholesky decomposition (Shojaie & Michailidis, 2010).

To interpret the penalty terms, denote

\[ R_1(\beta; D) = \| D\beta \|_1, \ R_2(\beta) = \| \beta \|_1, \ \text{and} \ R(\beta; D) = R_1(\beta; D) + \alpha R_2(\beta). \]

\( R_1 \) regulates the total effect of the predictors, \( R_2 \) regulates the direct effect, and \( R \) leverages both with a tuning parameter \( \alpha \). For the toy example of Figure 1A,

\[ Y = X\beta + \epsilon = eD\beta + \epsilon \triangleq \gamma + \epsilon, \]

where the first line is the total effect of the root node (\( X_1 \)), the second line is the total effect of the internal nodes (\( X_2 \) and \( X_3 \)), and the third line is the total effect of the leaf nodes (\( X_4, X_5 \) and \( X_6 \)). For a root or internal node, the total effect counts all the possible path effects to the outcome.

For visualization purposes, we consider a simplified example with \( p = 3 \) nodes, as shown in Figure 1B and demonstrate the difference between the regularity functions. Figure 2 shows the contour plots. Under a lasso regularization (\( R_2 \)), the three predictors are penalized in an equivalent way. Under \( R_1 \) and \( R \), \( X_1 \) is regularized differently from \( X_2 \) and \( X_3 \), given the fact that \( X_1 \) is the parent of \( X_2 \) and \( X_3 \).

A major drawback of the original lasso is that when the covariates are dependent, model selection is not consistent (Zou & Hastie, 2005). Taking the tree structure into consideration, \( R_1 \) (or \( R \)) circumvents this issue. This can be shown by combining models (1) and (3),

\[ Y = X\beta + \epsilon = eD\beta + \epsilon \triangleq \gamma + \epsilon, \]

where \( \gamma = D\beta \in \mathbb{R}^p \) is the new model parameter. Under this definition,

\[ R_1(\beta; D) = \| D\beta \|_1 = \| \gamma \|_1. \]
Solving the optimization problem (5) with $R_1$ as the regularization becomes equivalent to searching for an ordinary lasso solution with $\varepsilon$ being the design matrix, where the columns of $\varepsilon$ are mutually independent. When $D$ is correctly specified, the estimated active set under $R_1$ is consistent. In the next section, we show that when the design matrix is well behaved corresponding to the regularization matrix $\tilde{D}$ in $R$, the estimator of $\beta$ is consistent in $\ell_2$-norm and model selection is also consistent.

2.1 Estimation consistency and model selection consistency

Let $\beta^*$ denote the true model parameter and $M = \{ \beta \in \mathbb{R}^p \mid (D\beta)^\top S\beta = 0 \}$ denote the model space under regularization $R$, where $S$ is the support of $D\beta^*$ and $S^c$ is the complement of $S$. In order to achieve estimation consistency and model selection consistency, the following assumptions are imposed. The first assumption is on the sample Fisher information matrix, $Q = \nabla^2 \ell(\beta^*)$, where $\ell$ is the loss function and $\nabla$ is the differential operator. Under Equation (5), $\ell = \|Y - X\beta\|_2^2 / 2$. The second is also on $Q$, but with respect to the regularization matrix $D$.

Assumption 1 (Restricted strong convexity, RSC) Let $C \subset \mathbb{R}^p$ be a known convex set containing $\beta^*$. The loss function $\ell$ is RSC on $C \cap M$ when

$$\theta^\top \nabla^2 \ell(\beta) \theta \geq m \| \theta \|_2^2, \beta \in C \cap M, \theta \in (C \cap M) - (C \cap M),$$

$$\| \nabla^2 \ell(\beta) - Q \|_2 \leq L \| \beta - \beta^* \|_2, \beta \in C,$$

for some $m > 0$ and $L < \infty$.

Assumption 2 For $\tau \in (0, 1)$,

$$\| D_S X^\top (D_S X^\top)^{-1} \text{sign}(D_S \beta^*)_S \|_\infty \leq 1 - \tau,$$

where $D_S \in \mathbb{R}^{|S| \times p}$ takes the rows of $D$ in $S$, $(D_S \beta^*)_S \in \mathbb{R}^{|S|}$, and $A^-$ is the Moore–Penrose pseudoinverse of a matrix $A \in \mathbb{R}^{p \times p}$ and sign$(\cdot)$ is the sign function.

For a sparse regression problem like (5), with random Gaussian or sub-Gaussian designs, the RSC condition is satisfied, even when the predictors are dependent (Raskutti et al., 2010; Rudelson & Zhou, 2012). Assumption 2 is an irrepresentability condition that requires the active predictors (with respect to $D$) to be not overly well-aligned with the inactive predictors. The ideal scenario is that the inactive predictors are orthogonal to the active predictors, which is impossible to realize when the data are high-dimensional. Assumption 2 relaxes the orthogonality to near orthogonality. The following theorem is an adaption of Corollary 4.2 in Lee et al. (2015) to the considered regularization $R$ when $\alpha \neq 0$.

Theorem 1. Assume $\alpha \neq 0$. Under Assumptions 1 and 2, for some $0 < \kappa_1, \kappa_2, \kappa_3 < \infty$ and $\lambda = (8\kappa_1\sigma / \tau) \sqrt{\log p / n}$, the estimator under $R$ is unique, and with probability at least $1 - 2p^{-1}$,

1. consistent:

$$\| \hat{\beta} - \beta^* \|_2 \leq \frac{4}{m} (\kappa_3 + 4\kappa_1\kappa_2 / \tau) \sigma \sqrt{\log p / n},$$

2. model selection consistent:

$$\hat{\beta} \in M.$$

The proof of Theorem 1 and values of $\kappa_1, \kappa_2, \kappa_3$ are provided in Section A.1. When $\alpha = 0$, an analogous consistency holds, where the compatibility constants, $\kappa_1, \kappa_2, \kappa_3$, are computed with respect to the regularization function $R_1$. 

**FIGURE 2** Contour plot of the regularity functions (A) $R_1(\beta; D)$, (B) $R_2(\beta)$, and (C) $R$ (with $\alpha = 1$) for the example in (A) with $a_{12} = a_{13} = 0.5$. This figure appears in color in the electronic version of this paper, and any mention of color refers to that version.
2.2 Algorithm

As discussed above, the proposed regularizations, $R_1$ and $R$, can be solved through the generalized lasso (Tibshirani & Taylor, 2011), though in the generalized lasso literature, the proposed tree formulation was not considered. The estimating procedure is summarized in Algorithm 1. The ordinary lasso, $R_2$, can also be solved by the generalized lasso by setting the sparsity matrix to be the $p$-dimensional identity matrix. To choose the tuning parameters, the $C_p$ criterion is considered, which is defined as

$$C_p(\lambda, \alpha) = \|Y - X\hat{\beta}_{\lambda, \alpha}\|^2 - n\sigma^2 + 2\sigma^2\text{df}(X\hat{\beta}_{\lambda, \alpha})$$

where $\hat{\beta}_{\lambda, \alpha}$ is the estimate of $\beta$ under tuning parameter $(\lambda, \alpha)$ and $\text{df}$ is the degrees of freedom. An unbiased estimate of $C_p$ is provided by Tibshirani and Taylor (2011). It is suggested to choose $(\lambda, \alpha)$ that minimizes $\hat{C}_p(\lambda, \alpha)$.

3 SIMULATION STUDY

In the simulation study, we first consider binary trees, where each root or internal node has two children. A case of (1) $L = 7$ levels with $p = 127$ nodes is considered. $X$’s are first generated following (2), where the nonzero elements of the adjacency matrix are set to be one. For $\beta$, various scenarios are considered (Figure 3). In addition, a second scenario, (2) $L = 6$ levels with $p = 323$ nodes, is considered, where the tree structure is the same as the one in the ADNI dataset (Figure B.1). In this setting, four sparsity levels, 95%, 85%, 50%, 10%, of the total effect are considered, where the sparsity level is the proportion of the zero coefficient in $\beta$. The corresponding sparsity levels in $\beta$ are 91%, 75%, 37%, 8%. All errors in models (1) and (2) are independently generated from a normal distribution. Five types of regularization are considered: (i) $R_1$, (ii) $R_1 + \alpha R_2$, (iii) $R_2$ (the Lasso regularization), (iv) the elastic net (EN, Zou and Hastie, 2005), (v) the graph-constrained regularization (GCR, Li and Li, 2008; Chen et al., 2015), and (vi) the GL Yuan and Lin (2006). For (ii), various values of $\alpha$ are considered and chosen together with $\lambda$ using $C_p$ defined in (16); and for (iv), multiple choices of the $\ell^2_2$ proportion are considered. For both cases, we present the results with the tuning parameter chosen by $C_p$ introduced in Section 2.2. For all the approaches, $\lambda$ is also chosen by $C_p$. Sample sizes of $n = 100$ for (1) and $n = 500$ for (2) are considered and the simulation is repeated for 200 replications. For (1), a sample size of 100 is chosen to examine the performance under the “large $p$, small $n$” scenario; for (2), the sample size is chosen to approximate the real data. In the implementation, the value of $D$ is calculated from the adjacency matrix and imputed into the regularization function. When implementing the GCR approach, the adjacency matrix is symmetrized as the method was designed for undirected graphs. In the GL-based approach, for each internal and root node, the child nodes are defined as one group. The estimate of $\beta$ is obtained directly from the approaches. The estimate of $y$ is also obtained using the definition, $\hat{y} = X\hat{\beta}$, where $\hat{\beta}$ is the estimate of $\beta$ and $\hat{y}$ is the estimate of $y$. To evaluate the performance, the sensitivity and specificity of identifying nonzero parameters are considered, as well as the mean squared error (MSE), defined as

$$\text{MSE}(\hat{\beta}) = \mathbb{E}\|\hat{\beta} - \beta\|^2 = \mathbb{E}\left\{\sum_{j=1}^{p}(\hat{\beta}_j - \beta_j)^2\right\}.$$
specificity of the Lasso and EN are lower, especially in identifying $\gamma$. In addition, the MSE of estimating $\beta$ and $\gamma$ is much higher than those of using $R_1$ and $R$, where the regularization is defined based on the tree structure. Compared to $R_1$, the GCR approach yields slightly higher specificity in identifying $\beta$ but lower in identifying $\gamma$. For both $\beta$ and $\gamma$, the estimated MSE of GCR is higher than that of $R_1$. This suggests that when the predictors possess a structure like a hierarchical tree, regularization based on a directed graph is more compelling in identifying the total effect. The GL approach performs the worst among all with lower sensitivity and specificity and has much higher MSE for both $\beta$ and $\gamma$ estimation.

Table 2 presents the result of Simulation (2) with $p = 323$ and $L = 6$ following the tree structure in Figure B.1. From the table, $R_1$ and $R$ outperform the Lasso and EN at sparsity levels 95%, 85%, and 50% with higher sensitivity and specificity and lower estimated MSE. When the sparsity level is as low as 10%, all methods fail with low specificity suggesting a large number of false positives. With a high level of sparsity (95% and 85%), the GCR approach yields a comparable sensitivity and specificity in estimating $\gamma$ as $R_1$ and $R$ do. However, the estimated MSE is the highest among all approaches. Similar to Simulation (1), at a high sparsity level, the GL approach performs the worst. As the sparsity level decreases, the specificity of estimating $\beta$ and $\gamma$ significantly reduces to even below 0.100 though the sensitivity yields high. The estimated MSE is also higher than the rest except the GCR approach. For all approaches, as the sparsity level decreases, the specificity of identifying $\beta$ and $\gamma$ decreases and the estimated MSE increases. Different from the rest, the GCR approach achieves high specificity, while the sensitivity is much lower and the estimated MSE is much higher when estimating $\gamma$ at low sparsity levels, suggesting that the GCR approach is more sensitive to the sparsity level of the parameters.

4 | THE ALZHEIMER’S DISEASE NEUROIMAGING INITIATIVE STUDY

We apply the proposed approach to the MRI data collected by the Alzheimer’s Disease Neuroimaging Initiative (ADNI, 2003). The ADNI study was launched in 2003 as a public–private partnership, led by Principal Investigator: Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of MCI and early AD.

A total of $n = 590$ subjects aged between 55 and 91 are included in this study (358 males and 232 females). These subjects are diagnosed with MCI (402 subjects) or AD (188 subjects) at recruitment based on the cognitive and behav-
TABLE 1  The average sensitivity, specificity, as well as the estimation mean squared error (MSE), in the simulation study with $p = 127$ and $n = 100$

| Model | Method | Sensitivity | Specificity | MSE | Sensitivity | Specificity | MSE |
|-------|--------|-------------|-------------|-----|-------------|-------------|-----|
|       | $\beta$ |             |             |     | $\gamma$   |             |     |
| $\mathcal{R}_1$ | 1.000 | 0.885 | 0.147 | 1.000 | 0.937 | 0.096 |
| $\mathcal{R}$ | 0.975 | 0.885 | 0.221 | 0.955 | 0.892 | 0.168 |
| Lasso ($\mathcal{R}_2$) | 1.000 | 0.874 | 0.237 | 1.000 | 0.615 | 0.231 |
| EN | 1.000 | 0.952 | 0.079 | 1.000 | 0.839 | 0.085 |
| GCR | 1.000 | 0.998 | 0.161 | 1.000 | 0.993 | 0.156 |
| (a) GL | 0.000 | 0.839 | 1.054 | 1.000 | 0.645 | 0.572 |
| $\mathcal{R}_1$ | 1.000 | 0.803 | 0.261 | 1.000 | 0.897 | 0.207 |
| Lasso ($\mathcal{R}_2$) | 1.000 | 0.898 | 0.144 | 1.000 | 0.691 | 0.207 |
| EN | 1.000 | 0.952 | 0.083 | 1.000 | 0.841 | 0.098 |
| GCR | 1.000 | 0.998 | 0.048 | 1.000 | 0.997 | 0.130 |
| (b) GL | 1.000 | 0.868 | 0.274 | 1.000 | 0.723 | 0.204 |
| $\mathcal{R}_1$ | 1.000 | 0.669 | 0.554 | 1.000 | 0.823 | 0.435 |
| Lasso ($\mathcal{R}_2$) | 1.000 | 0.952 | 0.027 | 1.000 | 0.838 | 0.086 |
| EN | 1.000 | 0.952 | 0.037 | 1.000 | 0.862 | 0.086 |
| GCR | 0.990 | 0.999 | 0.044 | 0.998 | 0.999 | 0.135 |
| (c) GL | 1.000 | 0.958 | 0.097 | 1.000 | 0.899 | 0.135 |
| $\mathcal{R}_1$ | 1.000 | 0.888 | 0.182 | 1.000 | 0.939 | 0.091 |
| Lasso ($\mathcal{R}_2$) | 0.972 | 0.882 | 0.271 | 0.970 | 0.857 | 0.159 |
| EN | 0.999 | 0.813 | 0.265 | 1.000 | 0.492 | 0.385 |
| GCR | 0.990 | 0.815 | 0.668 | 1.000 | 0.527 | 0.456 |
| (d) GL | 0.830 | 0.966 | 0.628 | 0.965 | 0.836 | 0.277 |
| $\mathcal{R}_1$ | 0.907 | 0.777 | 0.882 | 0.995 | 0.511 | 0.475 |
| Lasso ($\mathcal{R}_2$) | 0.970 | 0.801 | 0.360 | 1.000 | 0.890 | 0.209 |
| EN | 0.994 | 0.750 | 0.886 | 0.952 | 0.767 | 0.544 |
| GCR | 0.990 | 0.722 | 0.845 | 1.000 | 0.422 | 0.625 |
| (e) GL | 0.927 | 0.964 | 1.076 | 1.000 | 0.477 | 0.656 |
| $\mathcal{R}_1$ | 0.879 | 0.726 | 1.886 | 1.000 | 0.461 | 0.903 |

Note: The results are the averages over 200 replications. EN, elastic net; GCR, graph-constrained regularization; GL, group lasso.

The hierarchical data structure. From Level 0 to Level 5, there are $p = 323$ regions: 1 in Level 0, 7 in Level 1, 13 in Level 2, 44 in Level 3, 108 in Level 4, and 150 in Level 5. The multi-level volumetric data satisfy a compositional property. That is, the volume of root/internal nodes is the sum of the volume of their children. Based on this property, after standardizing the data, the $(i, j)$ element in the adjacency matrix is the ratio of the standard deviation of region $j$ over region $i$. In this study, ADNI_MEM, which is a composite score of memory, is considered as the outcome ($Y$) to study the association between brain volume composition and memory. This composite score has been validated in Crane et al. (2012). It uses data from the ADNI neuropsychological battery following item response...
### Table 2

The average sensitivity, specificity, as well as the estimation mean squared error (MSE), in the simulation study with \( p = 323 \) and \( n = 500 \) under different sparsity levels in \( y \), where the tree structure is the same as in the ADNI dataset in Section 4.

| Sparsity | Method | Sensitivity | Specificity | MSE | Sensitivity | Specificity | MSE |
|----------|--------|-------------|-------------|-----|-------------|-------------|-----|
| \( R_1 \) | GL     | 0.710       | 0.536       | 0.710 | 0.983 | 0.447 | 0.545 |
| \( R \)  | EN     | 0.996       | 0.609       | 0.378 | 1.000 | 0.679 | 0.384 |
| Lasso (\( R_2 \)) | EN | 0.826       | 0.585       | 0.673 | 0.999 | 0.441 | 0.545 |
| EN      | GCR    | 0.815       | 0.579       | 0.680 | 0.999 | 0.437 | 0.559 |
| 95% GL  |        | 0.696       | 0.840       | 0.734 | 0.997 | 0.662 | 0.572 |
| \( R \)  | EN     | 0.996       | 0.620       | 0.447 | 0.999 | 0.675 | 0.387 |
| Lasso (\( R_2 \)) | EN | 0.826       | 0.585       | 0.673 | 0.999 | 0.441 | 0.545 |
| EN      | GCR    | 0.815       | 0.579       | 0.680 | 0.999 | 0.437 | 0.559 |
| 85% GL  |        | 0.982       | 0.017       | 1.370 | 1.000 | 0.012 | 1.232 |
| \( R \)  | EN     | 0.996       | 0.355       | 1.372 | 0.999 | 0.361 | 1.176 |
| Lasso (\( R_2 \)) | EN | 0.949       | 0.280       | 1.606 | 0.999 | 0.205 | 1.342 |
| EN      | GCR    | 0.953       | 0.248       | 1.654 | 0.999 | 0.181 | 1.364 |
| 50% GL  |        | 0.993       | 0.012       | 1.888 | 0.999 | 1.012 | 1.294 |
| \( R \)  | EN     | 0.999       | 0.122       | 2.179 | 0.999 | 0.120 | 1.826 |
| Lasso (\( R_2 \)) | EN | 0.987       | 0.132       | 2.128 | 0.999 | 0.102 | 1.794 |
| EN      | GCR    | 0.993       | 0.068       | 2.063 | 0.999 | 0.055 | 1.681 |
| 10% GL  |        | 0.995       | 0.003       | 2.496 | 0.999 | 0.002 | 2.303 |

Note: The results are the averages over 200 replications. EN, elastic net; GCR, graph-constrained regularization; GL, group lasso.

...theory methods, including different word lists in the Rey Auditory Verbal Learning Test and the ADAS-Cog, and by Logical Memory I data missing by design. A higher ADNI_MEM outcome indicates better performance in the tests. We take the ADNI_MEM score acquired on the same day as the MRI scan or the first post-imaging measurement as the outcome. We apply the proposed approach to investigate the association between global/local brain volume and memory.

Based on the simulation results, regularization \( R \) is employed, where tuning parameters, \( \alpha \) and \( \lambda \), are chosen based on the estimated \( C_p \). Figure 4A presents the brain regions with a nonzero direct effect estimate and their parent brain region. In the figure, arrows in gray inform the hierarchical structure between regions. (This figure appears in color in the electronic version of this article, and any mention of color refers to that version.) The brain maps are colored corresponding to the segmentation level. A red arrow to ADNI_MEM indicates a positive direct effect and a blue arrow indicates a negative direct effect. Figure 4B shows the total effect of the regions included in Figure 4A.

The estimate of the effects is presented in Table B.1. Compared to the results from the lasso (\( R_2 \)) and the elastic net (Figure B.2), a more clear hierarchical structure is observed under \( R \). Most of the regions identified by the lasso and the elastic net are Level-4 and Level-5 regions, where the effect of some regions can be aggregated into the parent region in a higher level.

Regional brain atrophy is observed in normal aging and AD, which has been found to be associated with cognitive impairments, such as memory declination. A stereotypical pattern of neurodegeneration suggests that the atrophy occurs early in the medial temporal lobe and soon after spreads to the rest of the cortical areas following a trajectory of temporal–parietal–frontal, while motor areas are not generally impacted until the later stages of the disease (Pini et al., 2016). From Figure 4A, nonzero direct effects are observed in the limbic system (the part of the cerebral cortex that is beneath the temporal lobe and is involved in multiple complex functions, particularly in...
Figure 4 Relevant brain regions (A) with a nonzero direct effect and (B) with a nonzero total effect. Brain maps are colored by level. The gray arrows inform the hierarchical structure. A red arrow indicates a positive effect and a blue one indicates a negative effect. The width of the lines is proportional to the magnitude of the effect. This figure appears in color in the electronic version of this paper, and any mention of color refers to that version.
emotional and behavioral responses, Purves et al., 2004),
the temporal and occipital lobes, and the lateral ventricles.
Figure 4B presents the estimated total effect using the pre-
specified adjacency matrix. From Figure 4B, nonzero total
effects are mainly observed in Level 4 and Level 5 regions
suggesting localized impacts of atrophy on memory.

Based on the segmentation, the Level 4 limbic area con-
sists of the parahippocampal and entorhinal cortices in
Level 5. Together with the amygdala, these are all key
AD markers, repeatedly verified in the existing litera-
ture (De Leon et al., 2004; St J et al., 2006; Jones et al.,
2006; Barnes et al., 2006). Positive direct and total effects
are estimated using the proposed approach suggesting
the association between atrophy in these areas and mem-
ory decline. The gray matter loss in the lateral temporal
cortex, dorsal parietal and frontal cortex occurs during
the progression from incipient to mild AD. During this
period, cognitive deficits are observed in both memory and
non-memory domains including language, visuo-spatial,
and executive function (Frisoni et al., 2009). For the sen-
sorimotor and visual cortices, atrophy is observed until
later stages of AD (Pini et al., 2016). The left anterior
insula/frontal operculum complex (IFO) also has a strong
positive direct total effect on the memory outcome.
The association between atrophy in the insular cortex and cog-
nitive deficits, such as memory, in AD, has been reported in
the existing literature (Foundas et al., 1997; Lin et al., 2017).
The proposed approach identifies a positive direct total
effect of the right superior fronto-occipital fasciculus (SFO)
on memory (Figure 4 shows the location of the core of
the SFO). In the human brain, the SFO can either be an
isolated fasciculus or a branch of the superior longitudi-
nal fasciculus (Bao et al., 2017). It plays a major role in
speech and language, as well as the top-down modulation
of visual processing and spatial aspects of cognitive pro-
cessing (Bar et al., 2006; Schmahmann et al., 2007), and
was found to be associated with cognitive decline in the
aging population (Price et al., 2020). Negative direct total
effects are mainly observed in the lateral ventricles. Due
to the sharp contrast between the CSF in the ventricles and
surrounding tissue in T1-weighted images, measurement
of ventricular volume is amenable to robust automatic seg-
mentation. Thus, the ventricles are among the study focus
in the research of brain tissue atrophy (Nestor et al., 2008).
Particularly in AD research, ventricular enlargement, as
a measurement of hemispheric atrophy rates, has been
repeatedly reported as a marker of AD progression (Nestor
et al., 2008; Fjell et al., 2009; Kruthika et al., 2019).

5 | DISCUSSION

In this study, we propose an $\ell_1$-type regularization for
predictors following a hierarchical tree structure. Under
the concept of a path diagram, the proposed penalty reg-
ulates the total effect of each predictor on the outcome.
With regularity conditions, it is shown that under the
proposed regularization, the estimators of the model coef-
ficients are consistent in $\ell_2$-norm and the model selection
is also consistent. By applying to a brain structural imag-
ing dataset acquired from the ADNI study, the proposed
approach identifies brain regions associated with memory
decline. With regularization on the total effects, the
findings suggest that the impact of atrophy on memory
deficits is from small brain regions. The current analysis
is at the level of ROI. In neuroimaging studies, analyses
at the voxel level are also widely considered, where the
number of predictors ($p$) increases much faster than the
sample size ($n$). Theoretical modifications under the set-
ting of ultra-high dimension are a possible topic for future
research.

When the predictors follow a hierarchical structure, the
ordinary lasso regularization may lead to biased results, as
the predictors can be strongly correlated (Zou & Hastie,
2005). The proposed approach circumvents this issue by
introducing the influence matrix as the penalty matrix in
the regularization. We show that this is equivalent to apply-
ing the ordinary lasso regularization on the independent
latent factors that generate the predictors. In this study, we
focus on the estimation and interpretation of the model
parameters and the total effects. An important follow-up
question is to perform inference on the parameters and
the total effects, which we leave to future work. Though
this study is motivated by structural neuroimaging data,
it can be generalized to any other area of research where
hierarchical compositions are informative predictors.

ACKNOWLEDGMENTS

Data used in preparation of this paper were obtained
from the Alzheimer’s Disease Neuroimaging Initiative
(ADNI) database (adni.loni.usc.edu). As such, the
investigators within the ADNI contributed to the
design and implementation of ADNI and/or provided
data but did not participate in analysis or writing
of this paper. A complete list of ADNI investigators
can be found at: http://adni.loni.usc.edu/wp-content/
uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.
Zhao, Caffo, and Luo were partially supported by NIH grant
R01MH126970; Zhao by NIH grants P30AG072976 and U54AG065181; Caffo by NIH grants
R01EB029977, P41EB031771, and U54DA049110; Luo by
NIH grant R01EB022911.

DATA AVAILABILITY STATEMENT

The data that support the findings in this paper are
openly available in the Alzheimer’s Disease Neuroimaging
Initiative (ADNI) at adni.loni.usc.edu.
This article has earned an Open Materials badge for making publicly available the components of the research methodology needed to reproduce the reported procedure and analysis. All materials are available at http://re3data.org/.

**ORCID**

Yi Zhao [https://orcid.org/0000-0003-4766-5934](https://orcid.org/0000-0003-4766-5934)

Bingkai Wang [https://orcid.org/0000-0002-9349-2336](https://orcid.org/0000-0002-9349-2336)

**REFERENCES**

ADNI (2003). The Alzheimer’s Disease Neuroimaging Initiative (ADNI): [https://adni.loni.usc.edu/](https://adni.loni.usc.edu/) [Accessed 25th September 2020].

Bao, Y., Wang, Y., Wang, W. & Wang, Y. (2017) The superior fronto-occipital fasciculus in the human brain revealed by diffusion spectrum imaging tractography: an anatomical reality or a methodological artifact? *Frontiers in Neuroanatomy*, 11, 119.

Bar, M., Kassam, K.S., Ghuman, A.S., Boshyan, J., Schmid, A.M., Dale, A.M., Hämäläinen, M.S., Marinkovic, K., Schacter, D.L. & Rosen, B.R. (2006) Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, 103, 449–454.

Barnes, J., Whitwell, J.L., Frost, C., Josephs, K.A., Rossor, M. & Fox, N.C. (2006) Measurements of the amygdala and hippocampus in pathologically confirmed Alzheimer disease and frontotemporal lobar degeneration. *Archives of Neurology*, 63, 1434–1439.

Chen, L., Liu, H., Kocher, J.-P.A., Li, H. & Chen, J. (2015) glmgraph: an R package for variable selection and predictive modeling of structured genomic data. *Bioinformatics*, 31, 3991–3993.

Crane, P.K., Carle, A., Gibbons, L.E., Insel, P., Mackin, R.S., Gross, A., Jones, R.N., Mukherjee, S., Curtis, S.M. & Harvey, D. (2012) Development and assessment of a composite score for memory in the Alzheimer’s Disease Neuroimaging Initiative (ADNI). *Brain Imaging and Behavior*, 6, 502–516.

De Leon, M., DeSanti, S., Zinkowski, R., Mehta, P., Pratico, D., Segal, S., Clark, C., Kerkman, D., DeBernardis, I. & Li, J. (2004) MRI and CSF studies in the early diagnosis of Alzheimer’s disease. *Journal of Internal Medicine*, 256, 205–223.

Djamanakova, A., Tang, X., Li, X., Faria, A.V., Ceritoglu, C., Oishi, K., Hillis, A.E., Albert, M., Lyketsos, C. & Miller, M.I. (2014) Tools for multiple granularity analysis of brain MRI data for individualized image analysis. *Neuroimage*, 101, 168–176.

Doshi, J., Erus, G., Ou, Y., Resnick, S.M., Gur, R.C., Gur, R.E., Satterthwaite, T.D., Furth, S., Davatzikos, C. & Alzheimer’s Neuroimaging Initiative, (2016) MUSE: Multi-atlas region Segmentation utilizing Ensembles of registration algorithms and parameters, and locally optimal atlas selection. *Neuroimage*, 127, 186–195.

Fjell, A., Amlien, I., Westlye, L. & Walhovd, K. (2009) Mini-mental state examination is sensitive to brain atrophy in Alzheimer’s disease. *Dementia and Geriatric Cognitive Disorders*, 28, 252–258.

Foundas, A.L., Leonard, C.M., Mahoney, S.M., Agee, O.F. & Hellman, K.M. (1997) Atrophy of the hippocampus, parietal cortex, and insula in Alzheimer’s disease: a volumetric magnetic resonance imaging study. *Neuropsychiatry, Neuropsychology, & Behavioral Neurology*, 10, 81–89.

Frisoni, G.B., Prestia, A., Rasser, P.E., Bonetti, M. & Thompson, P.M. (2009) In vivo mapping of incremental cortical atrophy from incipient to overt Alzheimer’s disease. *Journal of Neurology*, 256, 916–924.

Jones, B.P., Barnes, J., Uylings, H.B., Fox, N.C., Frost, C., Witter, M.P. & Scheltens, P. (2006) Differential regional atrophy of the cingulate gyrus in Alzheimer disease: a volumetric MRI study. *Cerebral Cortex*, 16, 1701–1708.

Kruithof, K., Maheshappa, H. & Initiative, A.D.N. (2019) Multistage classifier-based approach for Alzheimer’s disease prediction and retrieval. *Informatics in Medicine Unlocked*, 14, 34–42.

Lee, J.D., Sun, Y. & Taylor, J.E. (2015) On model selection consistency of regularized M-estimators. *Electronic Journal of Statistics*, 9, 608–642.

Li, C. & Li, H. (2008) Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics*, 24, 1175–1182.

Lin, F., Ren, P., Lo, R.Y., Chapman, B.P., Jacobs, A., Baran, T.M., Porsteinsson, A.P. & Foxe, J.J. (2017) Insula and inferior frontal gyms’ activities protect memory performance against Alzheimer’s disease pathology in old age. *Journal of Alzheimer’s Disease*, 55, 669–678.

Liu, M., Zhang, D. & Shen, D. (2014) Identifying informative imaging biomarkers via tree structured sparse learning for AD diagnosis. *Neuroinformatics*, 12, 381–394.

Mori, S., Wu, D., Ceritoglu, C., Li, Y., Kolasny, A., Vaillant, M.A., Faria, A.V., Oishi, K. & Miller, M.I. (2016) MRICloud: delivering high-throughput MRI neuroinformatics as cloud-based software as a service. *Computing in Science & Engineering*, 18, 21–35.

MRICloud (2016) The MRICloud platform: [https://mricloud.org/](https://mricloud.org/) [Accessed 3rd March 2021].

Nestor, S.M., Rupsingh, R., Borrie, M., Smith, M., Accomazzi, V., Wells, J.L., Fogarty, J., Bartha, R. & Initiative, A.D.N. (2008) Ventricular enlargement as a possible measure of Alzheimer’s disease progression validated using the Alzheimer’s disease neuroimaging initiative database. *Brain*, 131, 2443–2454.

Pan, W., Xie, B. & Shen, X. (2010) Incorporating predictor network in penalized regression with application to microarray data. *Biometrics*, 66, 474–484.

Pini, L., Pievani, M., Bocchetta, M., Allomare, D., Bosco, P., Cavedo, E., Galluzzi, S., Marizzoni, M. & Frisoni, G.B. (2016) Brain atrophy in Alzheimer’s disease and aging. *Ageing Research Reviews*, 20, 35–48.

Price, L.R., Deardorff, R., Wu, Y.-C., West, J.D., McDonald, B.C., Risacher, S.L. & Saykin, A.J. (2020) A novel MRI contrast weighted ratio method for measuring myelin in older adults at risk for Alzheimer’s disease: neuroimaging/new imaging methods. *Alzheimer’s & Dementia*, 16, e046297.

Purves, D., Augustine, G., Fitzpatrick, D., Hall, W., LaMantia, A., McNamara, J. & White, L. (2004) *Neuroscience*. Sinauer Associates.

Raskutti, G., Wainwright, M.J. & Yu, B. (2010) Restricted eigenvalue properties for correlated Gaussian designs. *The Journal of Machine Learning Research*, 11, 2241–2259.

Rudelson, M. & Zhou, S. (2012) Reconstruction from anisotropic random measurements. In: Shie, M., Nathan, S. and Robert, C.W. (Eds.) *Conference on learning theory*. pp. 10.1–10.24.
Schmahmann, J.D., Pandya, D.N., Wang, R., Dai, G., D’Arceuil, H.E., de Crespigny, A.J. & Wedeen, V.J. (2007) Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. Brain, 130, 630–653.

Shojaie, A. & Michailidis, G. (2009) Analysis of gene sets based on the underlying regulatory network. Journal of Computational Biology, 16, 407–426.

Shojaie, A. & Michailidis, G. (2010) Penalized likelihood methods for estimation of sparse high-dimensional directed acyclic graphs. Biometrika, 97, 519–538.

St J, T., Pruessner, J.C., Faltraco, F., Born, C., Rocha-Unold, M., Evans, A., Möller, H.-J. & Hampel, H. (2006) Comprehensive dissection of the medial temporal lobe in AD: measurement of hippocampus, amygdala, entorhinal, perirhinal and parahippocampal cortices using MRI. Journal of Neurology, 253, 794–800.

Tang, X., Oishi, K., Faria, A.V., Hillis, A.E., Albert, M.S., Mori, S. & Miller, M.I. (2013) Bayesian parameter estimation and segmentation in the multi-atlas random orbit model. PloS One, 8, e65591.

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

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. & Knight, K. (2005) Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67, 91–108.

Tibshirani, R.J. & Taylor, J. (2011) The solution path of the generalized lasso. The Annals of Statistics, 39, 1335–1371.

Vemuri, P. & Jack, C.R. (2010) Role of structural MRI in Alzheimer’s disease. Alzheimer’s Research & Therapy, 2, 23.

Wang, T. & Zhao, H. (2017) Constructing predictive microbial signatures at multiple taxonomic levels. Journal of the American Statistical Association, 112, 1022–1031.

Wang, T. & Zhao, H. (2017) Structured subcomposition selection in regression and its application to microbiome data analysis. The Annals of Applied Statistics, 11, 771–791.

Wu, G., Kim, M., Sanroma, G., Wang, Q., Munsell, B.C., Shen, D. & Alzheimer’s Disease Neuroimaging Initiative, (2015) Hierarchical multi-atlas label fusion with multi-scale feature representation and label-specific patch partition. NeuroImage, 106, 34–46.

Yan, X. & Bien, J. (2021) Rare feature selection in high dimensions. Journal of the American Statistical Association, 116, 887–900.

Yuan, M. & Lin, Y. (2006) Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68, 49–67.

Yuan, M. & Lin, Y. (2007) Model selection and estimation in the Gaussian graphical model. Biometrika, 94, 19–35.

Zou, H. (2006) The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 101, 1418–1429.

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

SUPPORTING INFORMATION

Web Appendices, Tables, and Figures referenced in Sections 1–4, along with R code, are available with this paper at the Biometrics website on Wiley Online Library. R code is also available at Github website: https://github.com/zhaoyi1026/TreeLasso.

Figure B.1: The hierarchical tree structure of the brain segmentation. Brain regions are colored by level.

Table B.1: Estimated direct (\( \hat{\beta} \)) and total (\( \hat{\gamma} \)) effect of the relevant brain regions in Figure 4

Figure B.2: The identified brain regions related to the ADNI MEM outcome using (A) the lasso (\( R_1^2 \)), (B) the elastic net, (C) the graph-constrained regularization (GCR), and (D) the GL putting into the tree diagram

Data S1
Data S2

How to cite this article: Zhao, Y., Wang, B., Liu, C.-F., Faria, A.V., Miller, M.I., Caffo, B.S. et al. (2023) Identifying brain hierarchical structures associated with Alzheimer’s disease using a regularized regression method with tree predictors. Biometrics, 79, 2333–2345. https://doi.org/10.1111/biom.13775