Journal of Data Science 24 (1), 167–186
January 2026
DOI: 10.6339/25-JDS1203
Statistical Data Science

Subject-Specific Scalar-on-Image Regression

Leo Yu-feng Liu[1,†], Haixu Ma[1,†], Yufeng Liu[3,*], and Hongtu Zhu[1,2]

[1]*Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, USA*
[2]*Department of Biostatistics, University of North Carolina at Chapel Hill, USA*
[3]*Department of Statistics, University of Michigan, USA*

Abstract

Neuroimaging technology has received considerable attention in recent years. One of the key problems in the imaging data analysis is the heterogeneity among individual subjects. In particular, the relationship between the imaging biomarkers and the clinical outcomes may vary across different individuals. Popular existing statistical methodologies such as the functional linear regression and high dimensional linear regression can be inadequate because the homogeneous regression relationship is assumed for all subjects. In this paper, we propose the Subject-Specific Scalar-on-Image Regression (S3IR) model to handle heterogeneous populations. Specifically, we utilize a binary subject-specific masking image to capture the heterogeneous sparsity among individuals. The proposed S3IR model incorporates the spatial structure of the imaging data and is able to achieve both local smoothness and subject-specific sparsity of the estimated regression coefficients. Furthermore, we design an EM-type adaptive algorithm to estimate the model coefficients. Simulation studies are presented to show the superior performance of our proposed method over some existing ones in handling heterogeneity. Finally, we apply the S3IR model to analyze data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The results show that our model can effectively identify interpretable and significant disease-related regions and improve prediction performance of the cognitive scores.

Keywords *functional linear regression; heterogeneity; imaging data; reproducing kernel Hilbert space (RKHS)*

1 Introduction

The use of imaging biomarkers to predict clinical outcomes has been popular and created great impact in public health. Many studies have demonstrated that medical images can deliver clinically important information. Such information may help explore the pathophysiology of certain diseases and assist diagnosis and treatments (Fan et al., 2008; Stonnington et al., 2010; Cheng et al., 2011; Huang et al., 2016; Jiang et al., 2021). As imaging-based predictive models are increasingly used in clinical settings, ensuring their trustworthiness, such as interpretability, robustness, and alignment with clinical reasoning, has also become essential for responsible deployment in practice (Winston et al., 2002; Santos et al., 2016; Li et al., 2023).

Numerous statistical and machine learning tools have been proposed to analyze medical images in the literature. Zhu et al. (2023) provides a comprehensive review for neuroimaging data analysis. One branch of these models uses images as the covariates to predict the scalar-

†Contributed equally.
*Corresponding author. Email: yufliu@umich.edu.

Received January 9, 2025; Accepted October 9, 2025

valued clinical scores. We denote this as the scalar-on-image regression model (Zhu et al., 2017; Kang et al., 2018; Liu et al., 2017). If the linear relationship between the response and the imaging covariates is assumed, the coefficients of the model would lie on the same space as that of the covariate images. We refer such coefficients as the coefficient image. The coefficient image is of great importance in the prediction of clinical outcomes, as well as in the identification of pathologically relevant imaging biomarkers. In practice, the collected imaging data can be high-dimensional in many imaging studies, such as Magnetic Resonance Imaging (MRI) data (Raz et al., 2017; Casanova et al., 2018; Li et al., 2021). In these studies, the scalar-on-image regression problem can be viewed as a special case of the high-dimensional linear model (Bühlmann and Van De Geer, 2011), or the functional linear regression model (Ramsay and Silverman, 2005; Ferraty and Vieu, 2006), depending on how the imaging data are used in the model.

In the high dimensional linear model, one can predict the clinical score by using the high-dimensional imaging data as the covariates. In this case, each voxel of the imaging data is treated as a feature in the model. In practice, the dimension p of the voxels can be very large for medical images. For example, to analyze a typical T1 MRI image of size $256 \times 256 \times 256$, one will need to build a $16,777,216$ dimensional regression model. Moreover, due to the cost of imaging acquisitions, compared to the dimension of features, the sample size n is usually relatively small. Therefore, many regularized methods are proposed to solve this high-dimensional regression problem. For example, Carroll et al. (2009) proposed a regression model with the Elastic net penalty (Zou and Hastie, 2005) to analyze the MRI data, and Toiviainen et al. (2014) applied the Lasso regression (Tibshirani, 1996) to identify the brain regions that respond to musical stimuli. These sparse regularization methods can effectively solve the high-dimensional problem and may detect some sparse imaging voxels as the important features. However, the estimation of coefficient images of these methods does not incorporate spatial smoothness. As a result, these methods may not be able to extract interpretable imaging biomarkers. In contrast, the spatial regularization methods are attractive due to their well-structured coefficient images and improved prediction performance. For instance, Grosenick et al. (2013) proposed the GraphNet penalty as a generalized Elastic net penalty that incorporates local graphical structures; Liu et al. (2018) introduced a spatial regularization framework using the first and second order total variation penalties for both classification and regression problems. Such spatial regularization terms utilize the locally spatial structure of the imaging data and can help to yield spatially smooth coefficient images.

Another popular approach for the scalar-on-image regression is to utilize the functional linear regression techniques (Goldsmith et al., 2011; Morris, 2015; Wang et al., 2016a). Instead of fitting an ultra high dimensional regression model, these methods treat the images as functional covariates over the imaging domain, and construct the coefficient images using some basis functions. For instance, Reiss and Ogden (2007), Wang and Zhu (2017), and Feng et al. (2019) proposed the functional principal component regression and functional partial least squares approaches for the scalar-on-image regression model. Wang et al. (2014) and Reiss et al. (2015) applied the wavelet basis to construct spatially smooth regression coefficients. These methods can achieve spatially smooth coefficient images if the basis functions are chosen properly. Considering this potential advantage of functional linear regression, we develop our method under this framework.

The problem of interest in this paper is the common heterogeneity of imaging data. In the current literature, most scalar-on-image regression models assume the homogeneous regression relationship and apply a common model for all subjects. However, the homogeneity assumption may not hold in practice. Although the biological structures of human organs are relatively

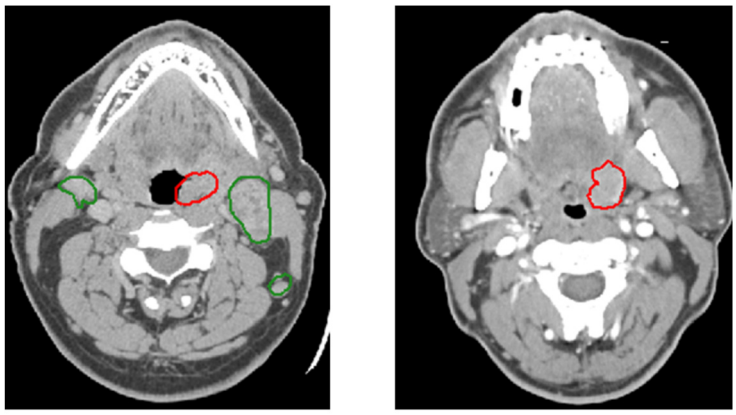

Figure 1: The brain images of two patients with head and neck cancer. The tumors are marked in the red circles.

consistent across the whole population, certain diseases can cause structural damages in different regions for different patients, thus yielding heterogeneous patterns (Maheux et al., 2023). For illustration, Figure 1 displays the brain images of two patients that have head and neck cancer. The tumors are marked by the red circles and located in different brain regions. The homogeneous models may underperform in terms of both estimation and prediction in this case. Many mixture regression models are proposed to address the heterogeneity issue. Examples include but not limited to Viele and Tong (2002), Hurn et al. (2003), Hoshikawa (2013), and Wang et al. (2016b). However, their estimation can be unstable when the number of mixture components is large. Moreover, the prediction rule of mixture regression models typically uses a weighted average of the mixture components based on the estimated posterior probabilities, which tends to ignore the heterogeneity for new patients and may cause biased prediction (Hoshikawa, 2013).

In this paper, we propose the Subject-Specific Scalar-on-Image Regression (S3IR) model under the functional linear regression framework to solve the heterogeneous regression problems efficiently. At the population level, the proposed S3IR model uses the homogeneous regression coefficients to characterize the common regression relationship between the biomarkers and clinical outcome. At the individual level, we further impose a subject-specific binary masking image to capture the heterogeneous sparsity across the large-scale populations. To incorporate spatial smoothness of the imaging data, we assume that the regression coefficients lie on the Reproducing Kernel Hilbert Space (RKHS) with a radial basis kernel. Hence, the overall estimate of the regression coefficients would yield spatial smoothness within certain regions and sparse spatial signals elsewhere. Moreover, we introduce an efficient adaptive algorithm to estimate the homogeneous effects and subject-specific binary masking images iteratively. For predicting the clinical outcome, we propose a novel classifier based on a weighted distance of the image covariate to determine whether to assign active or inactive regions for a new subject. The individual sparsity is determined by the corresponding individualized assignment based on this classifier. Importantly, our framework contributes to the broader goal of trustworthy machine learning, as it offers an interpretable, structured approach to model individual-level variability in imaging data. The regional masking mechanism and smooth coefficient structure enhance interpretability, while the model's individualized architecture promotes robustness to heterogeneity, which are key principles in trustworthy AI. The effective performance of the proposed S3IR model is demonstrated through several simulation studies and analysis of data from the Alzheimer's

Disease Neuroimaging Initiative (ADNI) study.

The rest of this paper is organized as follows. In Section 2, we introduce the S3IR modeling framework and its key components. The detailed algorithms for the estimation procedure, model selection, and the prediction are included in supplementary materials. We present extensive simulation experiments to illustrate the performance of S3IR in various settings in Section 3. The real application on the ADNI study is provided in Section 4. We conclude our paper in Section 5.

2 Methodology

In this section, we first introduce the homogeneous scalar-on-image regression model. Then, we extend it to our proposed S3IR model that can better deal with heterogeneity.

2.1 Data Structure and Homogeneous Models

Let $\boldsymbol{X}(t) \in \mathbb{R}$, $\forall t \in \mathcal{D}$, represent the image intensity, where $\mathcal{D}$ denotes the image domain, and t is the corresponding location index. Here, $\mathcal{D}$ can be a bounded 2-D surface or 3-D volume, and t can be a vector of length 2 or 3 respectively, according to the dimension of image domain $\mathcal{D}$. Using these notations, a scalar-on-image regression model is given as follows,

$$y_i = \beta_0 + \int_{t\in\mathcal{D}} \boldsymbol{X}_i(t)\boldsymbol{\beta}(t)dt + \epsilon_i; \qquad \text{for } i = 1, \ldots, n, \tag{1}$$

where $y_i \in \mathbb{R}$ represents the scalar responses, $\boldsymbol{X}_i$ denotes the covariate image, which is a real valued function over the domain $\mathcal{D}$, $\boldsymbol{\beta}$ is the homogeneous coefficient image of $\boldsymbol{X}_i$, $\beta_0 \in \mathbb{R}$ denotes the intercept term, ϵ_i represents the independent noise, and n denotes the sample size of the observed data. Here, model (1) is referred as the homogeneous model. It assumes a homogeneous regression relationship between the responses y_i's and the covariates $\boldsymbol{X}_i$'s, and all the subjects share a common coefficient image $\boldsymbol{\beta}$. In practice, this assumption may be violated by the heterogeneities among the samples. For example, in neuroimaging studies, the patients that belong to different subtypes of a certain disease may have different pathological patterns. Thus, they should not share one common disease map (Maheux et al., 2023). In particular, the location of affected regions in medical images may vary across different subjects, and the number of such regions may be more than one and different from one subject to another. In this case, the homogeneous model (1) may fail or underperform due to such heterogeneities across subjects.

2.2 Subject-Specific Models

Now, we propose the subject-specific model to solve the heterogeneity issue. Using the same notation of model (1), a subject-specific functional linear regression model is given by

$$y_i = \beta_0 + \int_{t\in\mathcal{D}} \boldsymbol{X}_i(t)\boldsymbol{\beta}_i(t)dt + \epsilon_i; \qquad \text{for } i = 1, \ldots, n, \tag{2}$$

where $\boldsymbol{\beta}_i$ is a subject-specific coefficient image to capture the heterogeneous relationship between $\boldsymbol{X}_i$ and y_i. This model allows each sample to perform its unique regression relationship, thus offers considerable flexibility. However, there are two major limitations of model (2). First, model (2) is not identifiable, i.e., we are unable to construct n distinct coefficient images using just n

samples. Second, model (2) may lack interpretability and prediction power due to the ignorance of the population structure.

In most neuroimaging studies, the heterogeneous disease patterns often exist in the form of sub-groups or sub-clusters. The disease maps may be similar within each group and different across groups. Thus, it is reasonable to assume that there exists a set of coefficient images, $\mathcal{B} = \{\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(M)}\}$, where $M \ll n$ is the total number of sub-groups, that covers most of the pathological patterns of the disease. Therefore, instead of using a distinct $\boldsymbol{\beta}_i$ for each subject, we may assume $\boldsymbol{\beta}_i = \boldsymbol{\beta}^{(m)}$ if subject i belongs to the m-th group. This assumption naturally leads to the consideration of a mixture regression model (Viele and Tong, 2002; Hurn et al., 2003; Hoshikawa, 2013; Wang et al., 2016b),

$$y_i = \sum_{m=1}^{M} \tau_{im} \left[\int_{t \in \mathcal{D}} \boldsymbol{X}_i(t) \boldsymbol{\beta}^{(m)}(t) dt \right] + \epsilon_i; \qquad \text{for } i = 1, \ldots, n, \tag{3}$$

where τ_{im} is the posterior probability of subject i belonging to the group m, i.e., $\tau_{im} = P(\boldsymbol{\beta}_i = \boldsymbol{\beta}^{(m)} | \boldsymbol{X}_i, y_i)$. Compared with the homogeneous model (1) and the subject-specific model (2), the mixture regression model (3) allows heterogeneity among the disease maps for different subgroups while at the same time maintains identifiability.

Although model (3) serves as a good compromise between model (1) and (2), it may underperform when the total number of the mixture components is large. Moreover, how to accurately estimating the number of mixture components can be one of the problems for the mixture model in practice. Cross validation or Bayesian information criteria (BIC) are potential methods to determine this number, however, the performance of them may not be stable. Furthermore, the prediction of mixture regression depends on the estimation of the posterior probability for each group. Such estimation requires strong assumptions on the joint distribution of the covariates and the response (Hoshikawa, 2013).

To overcome the above challenges, we introduce a hidden binary masking image $\boldsymbol{B}_i$ to characterize the heterogeneity and propose the novel S3IR model as follows,

$$y_i = \beta_0 + \int_{t \in \mathcal{D}} \boldsymbol{X}_i(t) \boldsymbol{B}_i(t) \boldsymbol{\beta}(t) dt + \epsilon_i; \qquad \text{for } i = 1, \ldots, n. \tag{4}$$

Model (4) is motivated by considering both population structure and the individual heterogeneity. At the population level, the intrinsic biological structures, i.e., brain structure, in the medical images are relatively consistent across different people. This inspires us to use a homogeneous disease map $\boldsymbol{\beta}$ to characterize the common regression relationship between the clinical responses and medical images. At the individual level, different patients may present different affected regions in the images. Hence, we introduce the subject-specific masking image $\boldsymbol{B}_i$'s to capture this heterogeneity. On one hand, model (4) can generate subject-specific disease patterns from the homogeneous disease map, i.e., $\boldsymbol{\beta}_i(t) = \boldsymbol{B}_i(t)\boldsymbol{\beta}(t), \forall t \in \mathcal{D}$. On the other hand, it can be interpreted as a subject-specific feature extraction procedure that selects different regions in the images as subject-specific covariates for different patients, i.e., $\tilde{\boldsymbol{X}}_i = \boldsymbol{X}_i \circ \boldsymbol{B}_i$. The overlap of all the $\boldsymbol{B}_i$'s characterizes the overall sparsity structure of the disease map $\boldsymbol{\beta}$. We denote the overlap as the population masking image $\boldsymbol{B}$, i.e.,

$$\boldsymbol{B}(t) = \begin{cases} 1 & \text{if one of } \boldsymbol{B}_i(t) = 1, \\ 0 & \text{if all of } \boldsymbol{B}_i(t) = 0. \end{cases} \tag{5}$$

The population masking image $\boldsymbol{B}$ serves two key purposes in our framework. First, it defines the support regions for the homogeneous disease map $\boldsymbol{\beta}$, ensuring that $\boldsymbol{\beta}(t) = 0$ when $\boldsymbol{B}(t) = 0$, which

enforces sparsity at the population level. Second, it provides the basis for constructing individual-level masking images $\boldsymbol{B}_i(t)$ by determining the candidate regions $\boldsymbol{R}_k$ that can be activated or deactivated for each subject, which will be described in detail in Section 2.2.2. Thus, $\boldsymbol{B}(t)$ acts as a template that captures the overall spatial pattern of potential disease-related regions while allowing for subject-specific heterogeneity within this template.

2.2.1 Homogeneous Disease Map and Its Potts Prior

The population-level disease map $\boldsymbol{\beta}$ in model (4) describes the homogeneous regression relationship between the clinical responses and the medical images, and should reveal the common pathological patterns in the images. Moreover, the biological structure of the human organs naturally yields locally smooth patterns in the medical images, and the disease-related patterns usually appear at certain locations. Due to these characteristics, the disease map should achieve both spatial smoothness and sparsity, and be able to detect the possible disease-related regions while ruling out the irrelevant ones, such as the background or normal tissues.

To incorporate spatial smoothness, we assume that the coefficient image $\boldsymbol{\beta}$ lies on the Reproducing Kernel Hilbert Space (RKHS) with a radial-based kernel. In this paper, we focus on the Gaussian kernel given as follows,

$$\mathcal{K}(s,t) = \exp\left\{-\frac{\|s-t\|^2}{\sigma^2}\right\}, \tag{6}$$

where $s, t \in \mathcal{D}$ denote two different location indices in the image domain, and σ represents the bandwidth of the kernel function. The space $\mathcal{H}$ induced by this kernel is a space of smooth functions over the image domain $\mathcal{D}$. The level of the smoothness is controlled by σ.

In general, RKHS provides smooth estimate of the coefficients over the imaging domain $\mathcal{D}$, but it does not guarantee sparse estimation. We utilize the population-level binary image $\boldsymbol{B}$ in (5) to reinforce the spatial structure and achieve the desired sparsity in the coefficient images. Specifically, the regions with $\boldsymbol{B}(t) = 1$ capture the effective regions (non-zero parts of coefficient images), and the ones with $\boldsymbol{B}(t) = 0$ represent the ineffective regions. We assume that the masking image $\boldsymbol{B}$ follows a Potts model (Besag, 1986; Zhang et al., 2001), which measures the probability of sparsity patterns in the image domain $\mathcal{D}$. Let $s_{\boldsymbol{B}}(t)$ denote the local similarity score of the binary image $\boldsymbol{B}$ at location $t \in D$, τ be a parameter that controls the level of spatial smoothness and local similarity, and $C(\tau)$ be the normalizing factor. The probability mass function of a Potts model is given as:

$$p(\boldsymbol{B}|\tau) = \exp\left\{\tau \sum_{t\in\mathcal{D}} s_{\boldsymbol{B}}(t)\right\} C(\tau). \tag{7}$$

The local similarity score is defined as:

$$s_{\boldsymbol{B}}(t) = \sum_{t'\in\mathcal{N}_t} \delta\left(\boldsymbol{B}(t'), \boldsymbol{B}(t)\right), \tag{8}$$

where $\delta(\cdot,\cdot)$ is the Kronecker function, i.e., $\delta(s,t) = 1$ if $s = t$ and 0 if $s \neq t$, and $\mathcal{N}_t$ defines all the adjacent neighborhoods of t. In particular, we use the 4 adjacent pixels and the 6 adjacent voxels as the neighborhoods of t to construct the similarity scores for the 2-D and 3-D images respectively. Given a specific $\boldsymbol{B}$, the local similarity score $s_{\boldsymbol{B}}(t)$ for $t \in \mathcal{D}$ is computed according

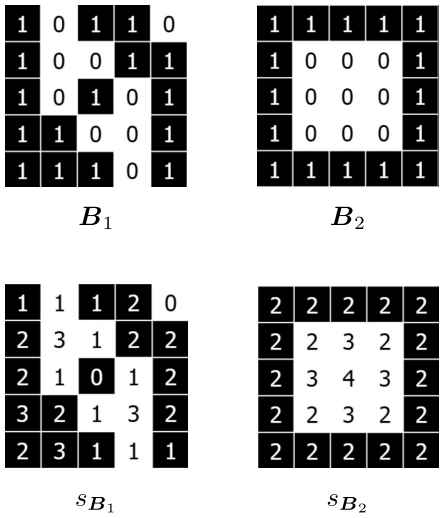

Figure 2: Examples of binary images in 2-D space and their associated local similarity scores. The first row displays the binary images $\boldsymbol{B}_1(t)$ and $\boldsymbol{B}_2(t)$, which have exactly same numbers of 0's and 1's. The second row displays their associated local similarity scores $s_{\boldsymbol{B}_1}(t)$ and $s_{\boldsymbol{B}_2}(t)$.

to Equation (8), and the corresponding Potts likelihood of $\boldsymbol{B}$ can be computed using Equation (7).

Note that compared with the $\boldsymbol{B}$ with a messy structure, the $\boldsymbol{B}$ with the spatial smooth structure is more likely to be selected in our model. For illustration, the example in Figure 2 includes two binary images $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ with exactly the same number of 0's and 1's. However, $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ have different spatial patterns, resulting in different sums of local similarity scores: $\sum_{t\in\mathcal{D}} s_{\boldsymbol{B}_1}(t) = 44$ and $\sum_{t\in\mathcal{D}} s_{\boldsymbol{B}_1}(t) = 56$. Hence, compared with $\boldsymbol{B}_1$, $\boldsymbol{B}_2$ has larger value of likelihood and is more likely to be selected. Therefore, under the Potts model assumption, we may expect that, with large probability, the binary image $\boldsymbol{B}$ would prefer partitioning the whole image domain $\mathcal{D}$ into several disjoint non-zero regions that have the more organized structures, denoted as $\mathcal{R}_k \subset \mathcal{D}$ for $k = 1, \ldots, K$, i.e.,

$$\boldsymbol{B}(t) = \begin{cases} 1 & \text{if } t \in \cup_k \mathcal{R}_k, \\ 0 & \text{if } t \in \mathcal{D}/\cup_k \mathcal{R}_k. \end{cases} \tag{9}$$

We will use these region $\mathcal{R}_k$'s to characterize the potential disease-related regions in the medical images. Since the homogeneous disease map $\boldsymbol{\beta}$ will be constructed based on the support regions in $\boldsymbol{B}$, the estimated coefficient image $\boldsymbol{\beta}$ can achieve similar spatial smoothness and spatial sparsity to that of $\boldsymbol{B}$.

2.2.2 Individual Disease Maps

In our S3IR model, the heterogeneity is characterized by the subject-specific coefficient $\boldsymbol{\beta}_i$'s, and each $\boldsymbol{\beta}_i$ is constructed by applying a binary masking image $\boldsymbol{B}_i$ on the homogeneous coefficient $\boldsymbol{\beta}$, i.e., $\boldsymbol{\beta}_i = \boldsymbol{B}_i \circ \boldsymbol{\beta}$. These heterogeneous coefficient images not only describe the intrinsic biological structures with the spatially smoothing homogeneous map $\boldsymbol{\beta}$, but also incorporate heterogeneity structure via the masking images $\boldsymbol{B}_i$'s.

Figure 3: Plots for the homogeneous coefficient images $\boldsymbol{\beta}$ in Scenarios 1, 2 and 3.

The heterogeneous masking images are constructed by activating different nonoverlapping regions detected, defined in Equation (9) for the homogeneous mask $\boldsymbol{B}$. Specifically,

$$\boldsymbol{B}_i = \sum_{k=1}^{K} I_{ik} \boldsymbol{R}_k, \tag{10}$$

where $I_{ik} \in \{0, 1\}$ indicates whether region $\mathcal{R}_k$ is active or not for the i-th subject, and $\boldsymbol{R}_k$ denotes the support function for region $\mathcal{R}_k$, i.e,

$$\boldsymbol{R}_k(t) = \begin{cases} 1 & \text{if } t \in \mathcal{R}_k, \\ 0 & \text{if } t \in \mathcal{D} \setminus \mathcal{R}_k, \end{cases} \tag{11}$$

where $\mathcal{D} \setminus \mathcal{R}_k$ denotes the set difference (i.e., the complement of $\mathcal{R}_k$).

In summary, our model contains two sets of coefficients: the population coefficients $\boldsymbol{\beta}$ and $\boldsymbol{B}$, and the individual-level coefficients $\boldsymbol{B}_i$'s. The homogeneous disease map $\boldsymbol{\beta}$ characterizes the overall regression relationship between the disease status y_i's and the medical images $\boldsymbol{X}_i$'s. The masking image $\boldsymbol{B}$ incorporates the spatial smoothness and sparsity of $\boldsymbol{\beta}$, and recovers the possible disease-related regions $\mathcal{R}_k$'s. The individual-level binary images $\boldsymbol{B}_i$'s determine the active regions for each subject, and characterize the heterogeneous structure of different individuals.

3 Simulation Studies

In this section, we present the numerical performance of the S3IR model through multiple simulation studies conducted for various scenarios. The covariate images $\boldsymbol{X}_i$'s are generated with a 2D Gaussian random field, with zeros constrained on the boundary. The population-level coefficient image $\boldsymbol{\beta}$ is generated based on Gaussian functions centered at various locations in the image domain. The number of non-zero regions is set to be 2, 3 and 5 in Scenarios 1, 2 and 3, respectively. The demonstration of these coefficient images is illustrated in Figure 3.

In each scenario, every individual-level coefficient $\boldsymbol{\beta}_i$ has at least one active region chosen randomly from the non-zero regions in the homogeneous coefficient images. The remaining regions are assigned to be active with a probability 0.2. Within each active region $\mathcal{R}_k$, we introduce regional signals to the covariate images, denoted as S_k. These signals are generated from the corresponding regional signal in the homogeneous coefficient image. The strength of the signals is controlled by the Regional Signal-To-Noise Ratio (RSNR), i.e., $S_k = \text{RSNR} * \text{Unif}(1, 1.5) * \boldsymbol{\beta} * \boldsymbol{R}_k$. These additional signals introduce the heterogeneity among the covariates, with the RSNR determining the level of heterogeneity. A higher RSNR indicates more distinguishable patterns

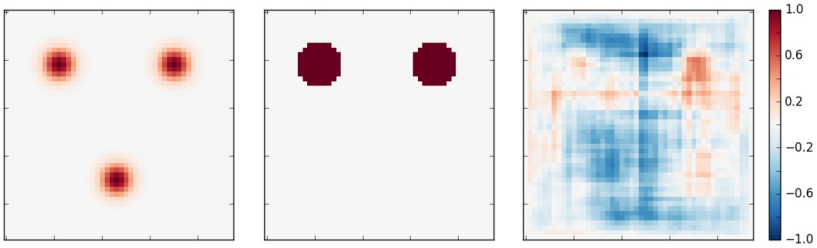

Figure 4: An example of the synthetic image in the case with moderate heterogeneity. The left panel is the coefficient image with 3 non-zero regions. The middle panel is the masking image, indicating that the top two regions are active. The right panel is the corresponding covariate image $\boldsymbol{X}$ with RSNR $= 0.5$.

between active and inactive regions. In our simulation studies, we conduct experiments in each scenario with the value of RSNR increasing from 0, 0.25, 0.5, 0.75 to 1. This covers a wide range of heterogeneity, from homogeneity to weak, moderate, and strong heterogeneity. An example of the coefficient image and its associated covariate image in the case of moderate heterogeneity (RSNR $= 0.5$) is shown in Figure 4. The homogeneous coefficient image has 3 non-zero regions, with the top two being active in the covariate image.

We compare our model's performance with existing high-dimensional regression models, including Ridge Regression (RR) (Hoerl and Kennard, 1970), Lasso regression (Lasso) (Tibshirani, 1996), Elastic-Net regression (EN) (Zou and Hastie, 2005), Functional Linear Regression (FLR) (Yuan and Cai, 2010), and Spatial REGularized regression (SREG) (Liu et al., 2018).

For each simulation setting, we generate 200 training samples and an independent test set of 200 images. The comparison methods are evaluated using 10-fold cross validation for model selection. To evaluate the performance of different methods, we use the Relative Estimation Error (REE) of the homogeneous disease map $\boldsymbol{\beta}$, and the Root Mean Squared Prediction Error (RMSPE) calculated on the test samples. The REE and RMSPE are defined as $\text{REE} = \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2/\|\boldsymbol{\beta}\|^2$, and $\text{RMSPE} = \sqrt{1/n^* \sum_{i=1}^{n^*} \left(\hat{y}_i^* - y_i^*\right)^2}$, where * indicates the test samples and n^* is the test sample size.

The experiments are repeated 50 times, and the means and standard errors of the REE for $\boldsymbol{\beta}$ and RMSPE for prediction are reported in Tables 1 and 2. Figure 5 presents the boxplots of REE and RMSPE, displaying the median, 0.25 quantile, and 0.75 quantile. As shown in Table 1 and Figure 5, the proposed S3IR method consistently achieves highly competitive performance for the estimation of $\boldsymbol{\beta}$ compared to homogeneous methods. Specifically, we provide a visualization of the estimated coefficients from one replication in Figure 6. As shown in Figure 6, Lasso and Elastic-Net demonstrate an extremely sparse estimation characterized by isolated pixels, while ridge regression produces an overall shrinkage effect on the estimated coefficients. The estimated coefficients using SREG and FLR employ the specific clustering pattern. However, they also exhibit numerous false positives for the inactive regions. Furthermore, these methods attenuate the signals in active regions and thus, lead to biased estimation because they neglect the subject heterogeneity. In addition, as the RSNR increases, the accuracy of estimation improves for all six methods. Notably, the S3IR method capitalizes on the regional signals and the underlying heterogeneity, resulting in a more pronounced enhancement in estimation accuracy. Moreover, we observe stable empirical convergence of the S3IR method. The estimated coefficients stabilize rapidly. For instance, in a representative simulation with RSNR $= 0.5$, the relative change in

Table 1: REE of $\boldsymbol{\beta}$ in simulation studies. The means of the 50 iterations are reported, with standard errors in parentheses. The values of the best methods are in bold.

	Scenario 1				
RSNR	0	0.25	0.5	0.75	1
RR	0.95 (0.01)	0.90 (0.01)	0.76 (0.02)	0.61 (0.02)	0.48 (0.02)
EN	0.97 (0.03)	0.84 (0.05)	0.57 (0.06)	0.46 (0.04)	0.44 (0.04)
Lasso	0.99 (0.03)	0.87 (0.07)	0.60 (0.06)	0.51 (0.06)	0.49 (0.05)
SREG	0.35 (0.10)	0.17 (0.09)	0.16 (0.06)	0.18 (0.08)	0.18 (0.06)
FLR	0.67 (0.04)	0.49 (0.05)	0.30 (0.03)	0.22 (0.02)	0.18 (0.02)
S3IR	**0.15** (0.09)	**0.14** (0.08)	**0.08** (0.06)	**0.06** (0.04)	**0.06** (0.03)
	Scenario 2				
RSNR	0	0.25	0.5	0.75	1
RR	0.97 (0.01)	0.89 (0.01)	0.73 (0.02)	0.57 (0.03)	0.44 (0.02)
EN	0.99 (0.02)	0.87 (0.04)	0.71 (0.06)	0.63 (0.05)	0.60 (0.06)
Lasso	1.00 (0.01)	0.97 (0.05)	0.77 (0.07)	0.72 (0.07)	0.71 (0.08)
SREG	0.59 (0.12)	0.24 (0.08)	0.21 (0.10)	0.20 (0.06)	0.20 (0.07)
FLR	0.75 (0.04)	0.50 (0.04)	0.30 (0.03)	0.23 (0.02)	0.21 (0.02)
S3IR	**0.22** (0.17)	**0.09** (0.03)	**0.06** (0.01)	**0.05** (0.01)	**0.05** (0.01)
	Scenario 3				
RSNR	0	0.25	0.5	0.75	1
RR	0.97 (0.01)	0.90 (0.02)	0.73 (0.03)	0.57 (0.03)	0.44 (0.03)
EN	1.00 (0.01)	0.90 (0.04)	0.84 (0.12)	0.84 (0.10)	0.84 (0.05)
Lasso	1.00 (0.01)	1.02 (0.04)	1.05 (0.07)	1.05 (0.07)	1.06 (0.08)
SREG	0.78 (0.10)	0.37 (0.12)	0.25 (0.08)	0.26 (0.07)	0.28 (0.07)
FLR	0.80 (0.04)	0.55 (0.04)	0.32 (0.03)	0.26 (0.03)	0.23 (0.02)
S3IR	**0.70** (0.26)	**0.13** (0.09)	**0.09** (0.03)	**0.11** (0.03)	**0.10** (0.02)

MSE of $\hat{\boldsymbol{\beta}}$ falls below 10^{-3} within 15 iterations.

For the prediction accuracy under the independent test sets, as shown in Table 2 and Figure 5, the homogeneous models exhibit overall inferior performance for prediction due to biased estimation. Moreover, these models apply uniform coefficients to each subject, disregarding the fact that the subjects are heterogeneous. In contrast, the proposed S3IR model identifies active and inactive patterns in the training data and utilizes them to determine the active regions for the test samples through the pattern matching procedure defined in Equation (S.8) in the supplementary materials. Under moderate and strong RSNR settings, the prediction accuracy of S3IR shows significant improvement. However, in the homogeneous setting where RSNR equals to 0, although S3IR may yield a better estimation of the population disease map $\boldsymbol{\beta}$, its prediction performance can be worse than that of the homogeneous models. This is possibly due to the insufficient information available to accurately identify the active regions in the absence of regional heterogeneity.

Table 2: RMSPE for prediction in simulation studies. The means from 50 iterations are reported, with standard errors in parentheses. The values of the best methods are in bold.

	Scenario 1				
RSNR	0	0.25	0.5	0.75	1
RR	2.51 (0.14)	3.34 (0.16)	4.42 (0.19)	5.03 (0.28)	5.08 (0.30)
EN	2.52 (0.14)	3.20 (0.18)	3.21 (0.24)	2.96 (0.20)	2.83 (0.17)
Lasso	2.53 (0.14)	3.23 (0.22)	3.24 (0.24)	3.01 (0.21)	2.89 (0.17)
SREG	1.94 (0.14)	2.04 (0.14)	2.19 (0.18)	2.25 (0.16)	2.28 (0.18)
FLR	2.28 (0.12)	2.57 (0.15)	2.63 (0.17)	2.56 (0.15)	2.40 (0.16)
S3IR	**2.20** (0.14)	**2.08** (0.17)	**1.50** (0.18)	**1.19** (0.09)	**1.16** (0.08)
	Scenario 2				
RSNR	0	0.25	0.5	0.75	1
RR	2.71 (0.14)	4.08 (0.18)	5.59 (0.34)	6.20 (0.38)	6.24 (0.42)
EN	2.71 (0.15)	4.03 (0.19)	4.41 (0.29)	4.15 (0.31)	3.99 (0.27)
Lasso	2.71 (0.15)	4.18 (0.25)	4.46 (0.30)	4.27 (0.32)	4.13 (0.28)
SREG	**2.38** (0.16)	2.66 (0.20)	2.94 (0.22)	3.02 (0.22)	3.07 (0.24)
FLR	2.59 (0.13)	3.14 (0.16)	3.32 (0.18)	3.22 (0.18)	3.22 (0.21)
S3IR	2.74 (0.19)	**2.67** (0.21)	**1.91** (0.30)	**1.36** (0.09)	**1.29** (0.10)
	Scenario 3				
RSNR	0	0.25	0.5	0.75	1
RR	3.00 (0.15)	5.08 (0.29)	7.27 (0.51)	8.20 (0.50)	8.14 (0.57)
EN	2.99 (0.16)	5.08 (0.32)	6.69 (0.55)	6.43 (0.53)	6.07 (0.36)
Lasso	2.99 (0.17)	5.39 (0.29)	6.76 (0.53)	6.64 (0.51)	6.41 (0.38)
SREG	**2.90** (0.20)	**3.79** (0.28)	4.18 (0.30)	4.24 (0.29)	4.32 (0.28)
FLR	2.98 (0.15)	4.10 (0.20)	4.53 (0.29)	4.45 (0.25)	4.33 (0.24)
S3IR	3.48 (0.23)	4.14 (0.29)	**3.61** (0.45)	**2.42** (0.21)	**2.02** (0.15)

4 Real Data Application

We apply the proposed S3IR model to the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. The ADNI study was initiated in 2003 by National Institute on Aging, National Institute of Biomedical Imaging and Bioengineering, Food and Drug Administration, private pharmaceutical companies and non-profit organizations. The study collects various data, including Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) images, CerebroSpinal Fluid (CSF), blood biomarkers, and neuropsychological assessments, to investigate the combination of biological markers and neuropsychological tests for measuring the progression of Mild Cognitive Impairment (MCI) and early Alzheimer's Disease (AD). More information about the ADNI study can be found at the ADNI website (http://www.adni-info.org/).

We aim to utilize the T1-weighted MRI images collected at baseline to predict the cognitive scores of patients, including Mini–Mental State Examination (MMSE) scores, Alzheimer's Disease Assessment Scale-cognitive score with 11 test items (ADAS-cog11), and scores with 13 items (ADAS-cog13). The MMSE is a brief 30-point questionnaire used to evaluate the cog-

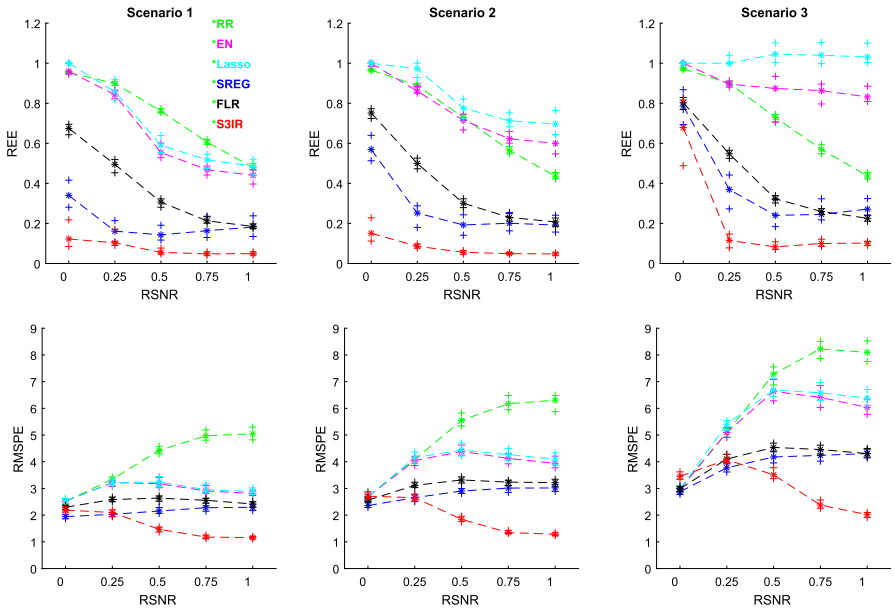

Figure 5: Boxplots of REE for $\boldsymbol{\beta}$ and RMSPE for prediction in simulation studies. The asterisk signs show the medians from 50 iterations, and the cross signs denote the 0.25/0.75 percentiles.

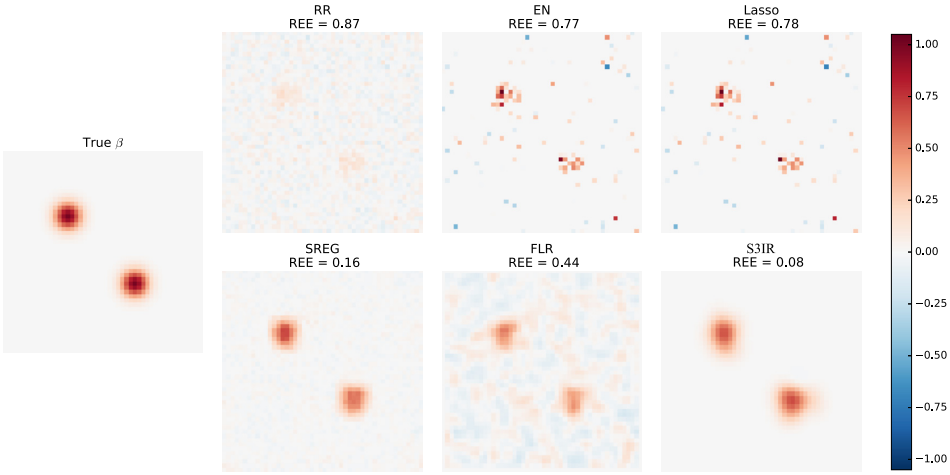

Figure 6: Plots of estimated coefficients in Scenario 1 with RSNR = 0.25. The image on the most left shows the true coefficients. The top row displays the estimation from Ridge Regression (RR), Elastic-Net (EN) and Lasso, and the bottom row demonstrates the estimation from Spatial REGularized regression (SREG), Functional Linear Regression (FLR) and the proposed Subject-Specific Scalar-on-image Regression (S3IR). The Relative Estimation Error (REE) for these methods are 0.87, 0.77, 0.78, 0.16, 0.44 and 0.08 respectively.

nitive impairment. It can be used to assess the patient's arithmetic, memory and orientation condition. In general, any score greater than or equal to 27 indicates normal cognition, while scores below this threshold indicate varying degrees of cognitive impairment. In addition, the ADAS-Cog score is important for evaluating AD pathology stage and predicting future pro-

Table 3: Demographical information of all subjects in data analysis. The mean values of different variables are reported, with standard deviations in parentheses.

Diagnosis	Male	Female	Age	MMSE	ADAS-cog11	ADAS-cog13
AD	86	79	75.47(7.33)	23.32(1.98)	18.32(6.02)	28.65(7.44)
MCI	238	128	74.78(7.23)	26.98(1.78)	11.55(4.48)	18.66(6.32)
NC	115	103	75.90(5.04)	29.10(1.01)	6.19(2.96)	9.44(4.22)

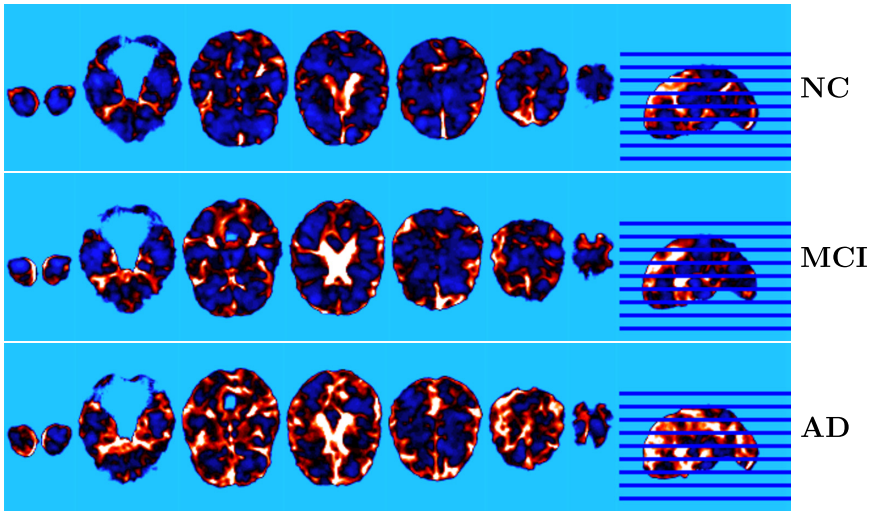

Figure 7: Three typical patterns of the covariate images from patients with diagnostic labels as NC, MCI and AD respectively (from top to bottom).

gression. The ADAS-cog11 test consists of 70 points while the ADAS-cog13 test consists of 85 points. Higher scores indicate more severe disease progression. For further details, please refer to Mungas (1991), Rosen et al. (1984), and Petersen et al. (2005).

Structural MRI is a powerful imaging technique for diagnosing AD. It is intriguing to use MRI data to predict cognitive scores, which can assist in diagnosing the current disease status of AD. In our analysis, we exclude subjects with missing measurements or low image quality and select 749 participants from the ADNI study, including 165 AD patients, 366 MCI patients, and 218 healthy controls. Table 3 summarizes the demographic information of the subjects.

To preprocess the MRI images, we perform several steps, including anterior commissure and posterior commissure correction, N2 bias field correction, skull-stripping, intensity inhomogeneity correction, cerebellum removal, segmentation, and registration. By applying these steps, we generate RAVENS-maps of the whole brain for each subject, resulting in 749 images of size $128 \times 128 \times 128$. Figure 7 provides examples of typical covariate images from patients with diagnostic labels as NC (Normal Control), MCI, and AD.

For model training and evaluation, we employ a training-test splitting scheme. The entire dataset is stratified based on diagnostic results and is divided into a 80% training set and a 20% test set. We train the models on the training set and evaluated their performance on the test set. For models requiring tuning parameter selection, we utilize a 5-fold cross validation procedure on the training set for model selection. We replicate this entire process with 30 times.

Table 4: Mean values of the Root Mean Square Prediction Error (RMSPE) in the ADNI data analysis among 30 random splits. The smallest values are displayed in bold and underlined, and the second smallest values are shown in bold only.

	MMSE	ADAS-cog11	ADAS-cog13
Ridge	**0.93**	**<u>0.92</u>**	**<u>0.89</u>**
Elastic	1.13	1.16	1.21
Lasso	0.99	1.00	0.96
SREG	1.14	1.12	1.12
FLR	0.98	1.02	0.97
S3IR	**<u>0.89</u>**	**0.93**	**0.93**

Table 5: Mean values of the correlation between predicted scores and observed scores in the ADNI data analysis among 30 random splits. The largest values are displayed in bold and underlined, and the second largest values are shown in bold only.

	MMSE	ADAS-cog11	ADAS-cog13
Ridge	0.66	0.66	0.70
Elastic	0.53	0.51	0.50
Lasso	0.61	0.59	0.64
SREG	**0.68**	**0.67**	**0.71**
FLR	0.64	0.60	0.65
S3IR	**<u>0.71</u>**	**<u>0.70</u>**	**<u>0.72</u>**

To predict each type of cognitive scores, we fit our proposed S3IR model and other comparison methods used in the simulation studies based on image covariates. In order to facilitate comparison across different cognitive measurements, we standardize the observed and estimated responses using the means and standard deviations calculated from the response values in the training set. We then compute the RMSPE using the rescaled values. Additionally, we evaluate the correlation between the predicted responses and the observed values, denoted as $corr(\hat{Y}, Y)$.

The RMSPE and the correlation results are summarized in Tables 4 and 5. We can see that the prediction results show consistent pattern for the three different responses across all models compared. Ridge regression achieves an overall better prediction error (around 0.90), although its correlation performance is not as good as other spatial methods. The sparse methods, Elastic Net and Lasso, yield high prediction errors and low correlations primarily due to their inconsistency in variable selection and extreme sparsity in the estimated coefficient images. SREG achieves low prediction accuracies but high correlations overall, partly due to the underlying spatial structure in its coefficient images and its capability to identify disease-related regions in the estimated coefficients. FLR delivers moderate prediction error and correlation, mainly due to the spatial smoothness induced by the Gaussian kernel in its coefficient images. The proposed S3IR achieves either the best or second-best prediction error and the highest correlation among all methods, demonstrating the advantage of modeling heterogeneity. The boxplots of RMSPE and the correlation in Figure 8 further confirm these results. Note that the Elastic Net and S3IR methods exhibit more variation across different iterations. For Elastic Net, this is primarily due to inconsistent variable selection by ignoring the heterogeneity and the complexity of the tuning

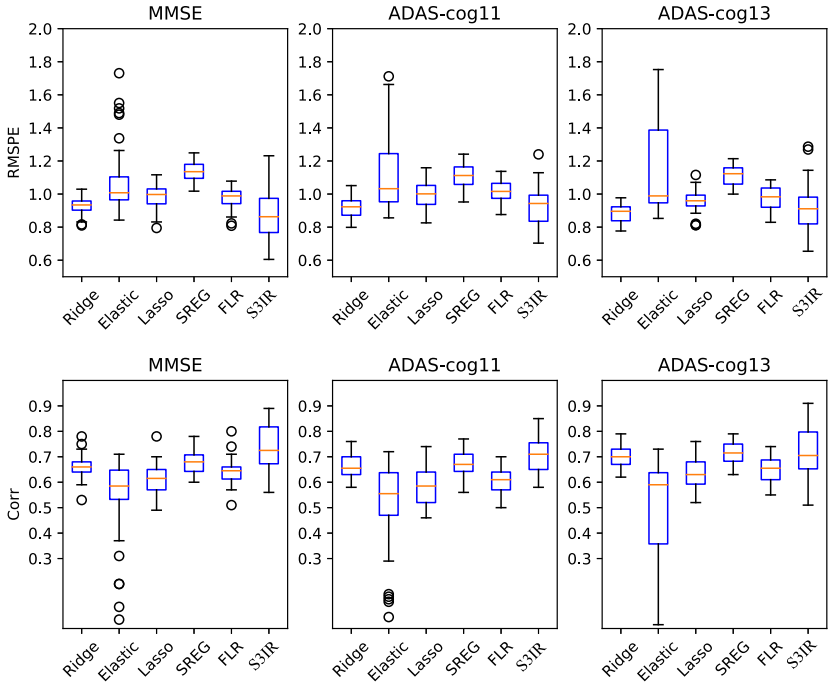

Figure 8: Boxplots of the Root Mean Square Prediction Error (RMSPE) and the correlation between predicted scores and observed scores in the ADNI data analysis among 30 random splits.

procedure, while for S3IR, the variation arises from the region estimations and assignments.

The coefficient images from one typical replication for the three different responses MMSE, ADAS-cog11, and ADAS-cog13 are displayed in Figure 9. We can observe consistent patterns across the coefficient images from ridge regression, SREG, FLR, and S3IR. The sparse methods, Elastic Net and Lasso, yield ultra-sparse disease maps that are challenging to interpret biologically. Ridge regression and SREG provide spatially clustered estimations that roughly capture the disease-related regions. FLR produces smoother coefficient images with larger patchy signals compared to ridge regression and SREG. The proposed S3IR not only captures spatial smoothing signals but also yields moderate sparsity in the disease maps, benefitting from the use of both the Gaussian kernel and the Potts prior. By comparing the coefficients in Figure 9 (a) with those in (b) and (c), we observe that the effects of the signals are opposite to each other. This discrepancy is due to the negative correlation between MMSE and disease severity, whereas the ADAS-cog scores exhibit a positive correlation.

To further investigate our results, we superimpose all the obtained coefficient images onto the Montreal Neurological Institute (MNI)-152 template (Fonov et al., 2011). Several regions are identified as significant and disease related in the generated disease maps, including the mid frontal gyrus, hippocampus, and temporal gyrus. These regions have been extensively discussed in previous research papers and are believed to be closely associated with the development of MCI and AD, and possibly functionally linked to cognitive abilities. For instance, the superior and middle frontal gyrus, which are involved in logical thinking and planning, undergo progressive damage during the progression of the disease (Hirono et al., 1998). Additionally, atrophy in the medial temporal lobe is predominantly observed in the early stages of Alzheimer's disease,

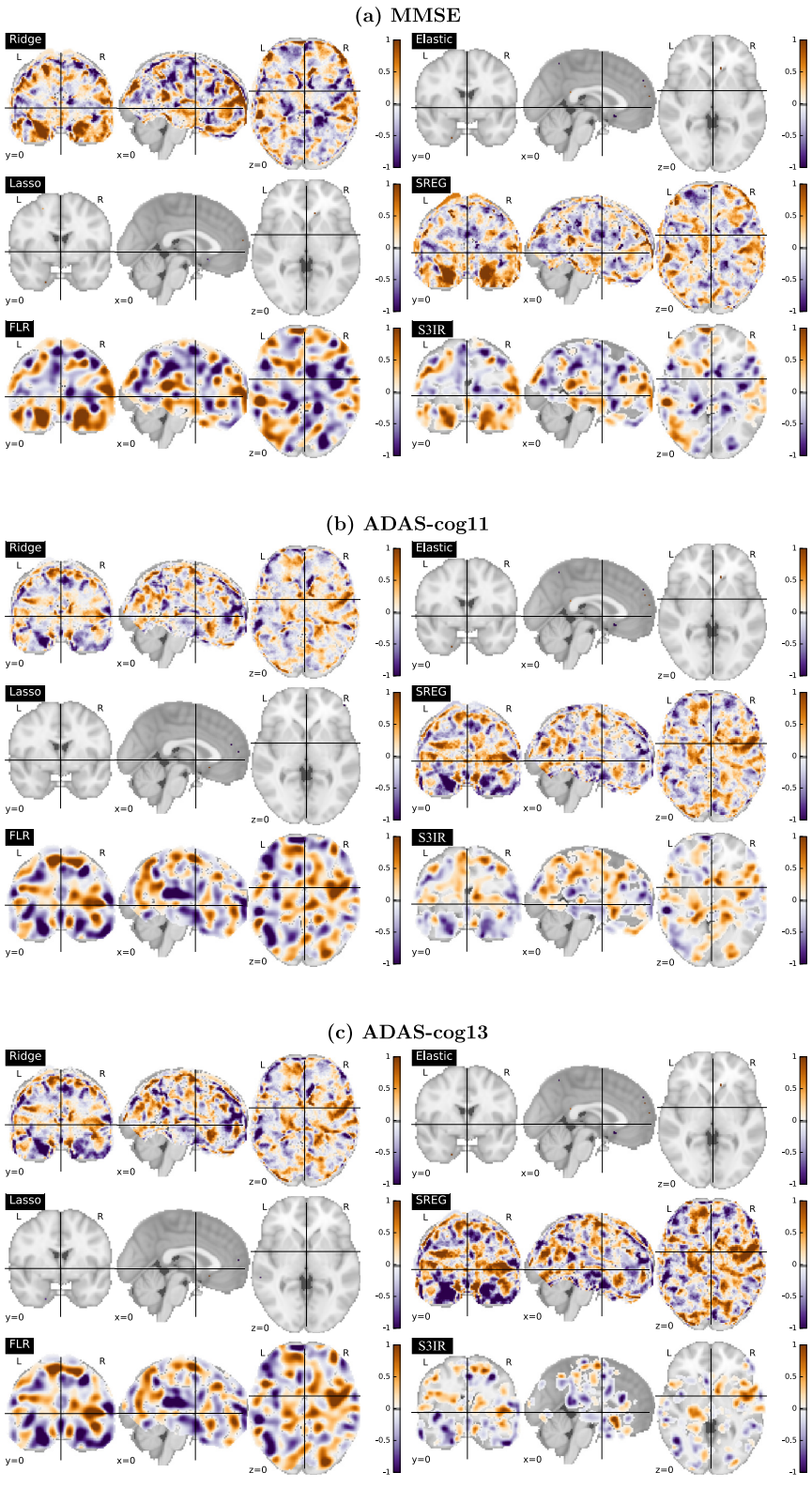

Figure 9: Plots of the coefficient images from all comparison methods for different response variables: (a) MMSE, (b) ADAS-cog11, and (c) ADAS-cog13. Each panel shows results from Ridge, Elastic Net, Lasso, SREG, FLR, and S3IR methods (from top-left to bottom-right).

while widespread temporal lobe and global cerebral atrophy are characteristic of advanced AD (Killiany et al., 1993; Chan et al., 2001). Our analysis is consistent with the above findings in the literature on Alzheimer's disease. While the performance gain of S3IR in the ADNI study is more modest compared to simulations, it remains consistent with expectations given the likely lower degree of regional heterogeneity in the real data. The ADNI results suggest that even in less heterogeneous settings, S3IR can offer reliable and stable performance, while its advantages become more pronounced as regional variation increases.

5 Discussion

In this paper, we introduce a new approach called Subject-Specific Scalar-on-Image Regression (S3IR) model to address the challenges posed by heterogeneous imaging data. The proposed S3IR model leverages the Gaussian reproducing kernel to maintain the spatial smoothness of the underlying biological structure during the estimation process. To further enhance spatial smoothness and exclude irrelevant regions, we incorporate a hidden masking image with a Potts prior. Our S3IR model captures both population-level and individual-level coefficients. On one hand, the population-level coefficients represent the homogeneous regression relationship between clinical responses and covariate images across the entire population. On the other hand, the individual-level coefficients take into account the heterogeneous structure and generate distinct patterns of active and inactive regions for each individual. By incorporating the heterogeneous structure among subjects, our S3IR model improves the prediction performance based on the individual characteristics and variations in the imaging data.

From a trustworthy machine learning perspective (Li et al., 2023), S3IR contributes along several dimensions. First, the subject-specific masks offer interpretable visual summaries of which brain regions are actively contributing to prediction for each individual, enhancing transparency in biomedical applications. Second, by explicitly modeling heterogeneity, the approach supports robust generalization to subpopulations and out-of-distribution individuals, which is critical in real-world deployment.

Most scenarios considered in the paper focus on the continuous outcome. One interesting future direction is to generalize the S3IR model to deal with various types of outcomes, such as discrete outcome and survival outcome. Another possible future work is to extend our method to deal with data with multiple outcomes.

Supplementary Material

The detailed algorithms for the estimation procedure, model selection, and the prediction are included in supplementary materials.

References

Besag J (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 259–302. https://doi.org/10.1111/j.2517-6161.1986.tb01412.x

Bühlmann P, Van De Geer S (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer Science & Business Media.

Carroll MK, Cecchi GA, Rish I, Garg R, Rao AR (2009). Prediction and interpretation of distributed neural activity with sparse models. *NeuroImage*, 44(1): 112–122. https://doi.org/10.1016/j.neuroimage.2008.08.020

Casanova R, Barnard RT, Gaussoin SA, Saldana S, Hayden KM, Manson JE, et al. (2018). Using high-dimensional machine learning methods to estimate an anatomical risk factor for Alzheimer's disease across imaging databases. *NeuroImage*, 183: 401–411. https://doi.org/10.1016/j.neuroimage.2018.08.040

Chan D, Fox NC, Scahill RI, Crum WR, Whitwell JL, Leschziner G, et al. (2001). Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Annals of Neurology*, 49(4): 433–442. https://doi.org/10.1002/ana.92

Cheng B, Zhang D, Chen S, Shen D (2011). Predicting clinical scores using semi-supervised multimodal relevance vector regression. In: *Machine Learning in Medical Imaging*. Suzuki K, Wang F, Shen D, Yan P (eds). *Lecture Notes in Computer Science*, Vol. 7009. Springer Berlin Heidelberg, Berlin, Heidelberg. 241–248. https://doi.org/10.1007/978-3-642-24319-6_30

Fan Y, Gur RE, Gur RC, Wu X, Shen D, Calkins ME, et al. (2008). Unaffected family members and schizophrenia patients share brain structure patterns: A high-dimensional pattern classification study. *Biological Psychiatry*, 63(1): 118–124. https://doi.org/10.1016/j.biopsych.2007.03.015

Feng X, Li T, Song X, Zhu H (2019). Bayesian scalar on image regression with nonignorable nonresponse. *Journal of the American Statistical Association*, 115(532): 1574–1597. https://doi.org/10.1080/01621459.2019.1686391

Ferraty F, Vieu P (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, New York, NY, USA.

Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, 54(1): 313–327. https://doi.org/10.1016/j.neuroimage.2010.07.033

Goldsmith J, Bobb J, Crainiceanu CM, Caffo B, Reich D (2011). Penalized functional regression. *Journal of Computational and Graphical Statistics*, 20(4): 830–851. https://doi.org/10.1198/jcgs.2010.10007

Grosenick L, Klingenberg B, Katovich K, Knutson B, Taylor JE (2013). Interpretable whole-brain prediction analysis with GraphNet. *NeuroImage*, 72: 304–321. https://doi.org/10.1016/j.neuroimage.2012.12.062

Hirono N, Mori E, Ishii K, Ikejiri Y, Imamura T, Shimomura T, et al. (1998). Frontal lobe hypometabolism and depression in Alzheimer's disease. *Neurology*, 50(2): 380–383. https://doi.org/10.1212/WNL.50.2.380

Hoerl AE, Kennard RW (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1): 55–67. https://doi.org/10.1080/00401706.1970.10488634

Hoshikawa T (2013). Mixture regression for observational data, with application to functional regression models. arXiv preprint: https://arxiv.org/abs/1307.0170.

Huang L, Jin Y, Gao Y, Thung KH, Shen D (2016). Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest. *Neurobiology of Aging*, 46: 180–191. https://doi.org/10.1016/j.neurobiolaging.2016.07.005

Hurn M, Justel A, Robert CP (2003). Estimating mixtures of regressions. *Journal of Computational and Graphical Statistics*, 12(1): 55–79. https://doi.org/10.1198/1061860031329

Jiang Z, He W, Kirby M, Asiri S, Yan D (2021). Weakly supervised spatial deep learning based on imperfect vector labels with registration errors. In: *Proceedings of the 27th ACM SIGKDD*

Conference on Knowledge Discovery & Data Mining (KDD '21). Association for Computing Machinery, New York, NY, USA. 767–775. https://doi.org/10.1145/3447548.3467301

Kang J, Reich BJ, Staicu AM (2018). Scalar-on-image regression via the soft-thresholded Gaussian process. *Biometrika*, 105(1): 165–184. https://doi.org/10.1093/biomet/asx075

Killiany RJ, Moss MB, Albert MS, Sandor T, Tieman J, Jolesz F (1993). Temporal lobe regions on magnetic resonance imaging identify patients with early Alzheimer's disease. *Archives of Neurology*, 50(9): 949–954. https://doi.org/10.1001/archneur.1993.00540090052010

Li B, Qi P, Liu B, Di S, Liu J, Pei J, et al. (2023). Trustworthy AI: From principles to practices. *ACM Computing Surveys*, 55(9): 177. https://doi.org/10.1145/3555803

Li C, Wang X, Du G, Chen H, Brown G, Lewis MM, et al. (2021). Folded concave penalized learning of high-dimensional MRI data in Parkinson's disease. *Journal of Neuroscience Methods*, 357: 109157. https://doi.org/10.1016/j.jneumeth.2021.109157

Liu LYF, Liu Y, Zhu H (2018). SMAC: Spatial multi-category angle-based classifier for high-dimensional neuroimaging data. *NeuroImage*, 175: 230–245. https://doi.org/10.1016/j.neuroimage.2018.03.040

Liu Y, Yan B, Merikangas K, Shou H (2017). Total variation regularized tensor-on-scalar regression. arXiv preprint: https://arxiv.org/abs/1703.05264.

Maheux E, Koval I, Ortholand J, Birkenbihl C, Archetti D, Bouteloup V, et al. (2023). Forecasting individual progression trajectories in Alzheimer's disease. *Nature Communications*, 14(1): 761. https://doi.org/10.1038/s41467-022-35712-5

Morris JS (2015). Functional regression. *Annual Review of Statistics and Its Application*, 2: 321–359. https://doi.org/10.1146/annurev-statistics-010814-020413

Mungas D (1991). In-office mental status testing: A practical guide. *Geriatrics*, 46(7): 54–8.

Petersen RC, Thomas RG, Grundman M, Bennett D, Doody R, Ferris S, et al. (2005). Vitamin E and donepezil for the treatment of mild cognitive impairment. *New England Journal of Medicine*, 352(23): 2379–2388. https://doi.org/10.1056/NEJMoa050151

Ramsay JO, Silverman BW (2005). *Functional Data Analysis*. Wiley, Hoboken, NJ. https://doi.org/10.1002/0470013192.bsa239

Raz G, Svanera M, Singer N, Gilam G, Cohen MB, Lin T, et al. (2017). Robust inter-subject audiovisual decoding in functional magnetic resonance imaging using high-dimensional regression. *NeuroImage*, 163: 244–263. https://doi.org/10.1016/j.neuroimage.2017.09.032

Reiss PT, Huo L, Zhao Y, Kelly C, Ogden RT (2015). Wavelet-domain regression and predictive inference in psychiatric neuroimaging. *The Annals of Applied Statistics*, 9(2): 1076–1101. https://doi.org/10.1214/15-AOAS829

Reiss PT, Ogden RT (2007). Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association*, 102(479): 984–996. https://doi.org/10.1198/016214507000000527

Rosen W, Mohs R, Davis K (1984). A new rating scale for Alzheimer's disease. *The American Journal of Psychiatry*, 141(11): 1356–1364. https://doi.org/10.1176/ajp.141.11.1356

Santos S, Almeida I, Oliveiros B, Castelo-Branco M (2016). The role of the amygdala in facial trustworthiness processing: A systematic review and meta-analyses of fMRI studies. *PLOS ONE*, 11(11): 1–28. https://doi.org/10.1371/journal.pone.0167276

Stonnington CM, Chu C, Klöppel S, Jack Jr CR, Ashburner J, Frackowiak RS, et al. (2010). Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. *Neuroimage*, 51(4): 1405–1413. https://doi.org/10.1016/j.neuroimage.2010.03.051

Tibshirani R (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal*

Statistical Society: Series B (Methodological), 58(1): 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Toiviainen P, Alluri V, Brattico E, Wallentin M, Vuust P (2014). Capturing the musical brain with Lasso: Dynamic decoding of musical features from fMRI data. *NeuroImage*, 88: 170–180. https://doi.org/10.1016/j.neuroimage.2013.11.017

Viele K, Tong B (2002). Modeling with mixtures of linear regressions. *Statistics and Computing*, 12(4): 315–330. https://doi.org/10.1023/A:1020779827503

Wang JL, Chiou JM, Müller HG (2016a). Functional data analysis. *Annual Review of Statistics and Its Application*, 3: 257–295. https://doi.org/10.1146/annurev-statistics-041715-033624

Wang S, Huang M, Wu X, Yao W (2016b). Mixture of functional linear models and its application to CO_2–GDP functional data. *Computational Statistics & Data Analysis*, 97: 1–15. https://doi.org/10.1016/j.csda.2015.11.008

Wang X, Nan B, Zhu J, Koeppe R (2014). Regularized 3D functional regression for brain image data via Haar wavelets. *The Annals of Applied Statistics*, 8(2): 1045–1064. https://doi.org/10.1214/14-AOAS736

Wang X, Zhu H (2017). Generalized scalar-on-image regression models via total variation. *Journal of the American Statistical Association*, 112(519): 1156–1168. https://doi.org/10.1080/01621459.2016.1194846

Winston JS, Strange BA, O'Doherty J, Dolan RJ (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, 5(3): 277–283. https://doi.org/10.1038/nn816

Yuan M, Cai TT (2010). A reproducing kernel Hilbert space approach to functional linear regression. *The Annals of Statistics*, 38(6): 3412–3444. https://doi.org/10.1214/09-AOS772

Zhang Y, Brady M, Smith S (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1): 45–57. https://doi.org/10.1109/42.906424

Zhu H, Li T, Zhao B (2023). Statistical learning methods for neuroimaging data analysis with applications. *Annual Review of Biomedical Data Science*, 6(1): 73–104. https://doi.org/10.1146/annurev-biodatasci-020722-100353

Zhu H, Shen D, Peng X, Liu LY (2017). MWPCR: Multiscale Weighted Principal Component Regression for High-Dimensional Prediction. *Journal of the American Statistical Association*, 112(519): 1009–1021. https://doi.org/10.1080/01621459.2016.1261710

Zou H, Hastie T (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2): 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x