



# A Bayesian Spatial Model to Predict Disease Status Using Imaging Data From Various Modalities

Wenqiong Xue<sup>1\*</sup>, F. DuBois Bowman<sup>2</sup> and Jian Kang<sup>3</sup>

<sup>1</sup> Boehringer Ingelheim Pharmaceuticals Inc., Ridgefield, CT, United States, <sup>2</sup> Department of Biostatistics, The Mailman School of Public Health, Columbia University, New York, NY, United States, <sup>3</sup> Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, United States

## OPEN ACCESS

### Edited by:

Xiaoying Tang,  
Center for Imaging Science,  
United States

### Reviewed by:

Russell Shinohara,  
University of Pennsylvania,  
United States  
Qingpeng Zhang,  
City University of Hong Kong,  
Hong Kong

### \*Correspondence:

Wenqiong Xue  
nora.xue@gmail.com

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 19 January 2017

**Accepted:** 06 March 2018

**Published:** 26 March 2018

### Citation:

Xue W, Bowman FD and Kang J  
(2018) A Bayesian Spatial Model to  
Predict Disease Status Using Imaging  
Data From Various Modalities.  
*Front. Neurosci.* 12:184.  
doi: 10.3389/fnins.2018.00184

Relating disease status to imaging data stands to increase the clinical significance of neuroimaging studies. Many neurological and psychiatric disorders involve complex, systems-level alterations that manifest in functional and structural properties of the brain and possibly other clinical and biologic measures. We propose a Bayesian hierarchical model to predict disease status, which is able to incorporate information from both functional and structural brain imaging scans. We consider a two-stage whole brain parcellation, partitioning the brain into 282 subregions, and our model accounts for correlations between voxels from different brain regions defined by the parcellations. Our approach models the imaging data and uses posterior predictive probabilities to perform prediction. The estimates of our model parameters are based on samples drawn from the joint posterior distribution using Markov Chain Monte Carlo (MCMC) methods. We evaluate our method by examining the prediction accuracy rates based on leave-one-out cross validation, and we employ an importance sampling strategy to reduce the computation time. We conduct both whole-brain and voxel-level prediction and identify the brain regions that are highly associated with the disease based on the voxel-level prediction results. We apply our model to multimodal brain imaging data from a study of Parkinson's disease. We achieve extremely high accuracy, in general, and our model identifies key regions contributing to accurate prediction including caudate, putamen, and fusiform gyrus as well as several sensory system regions.

**Keywords:** Bayesian spatial model, prediction, MCMC, posterior predictive probability, importance sampling, Parkinson's disease

## 1. INTRODUCTION

Functional and structural neuroimaging play important roles in understanding the neurological basis for major psychiatric and neurological disorders such as Parkinson's disease (PD), schizophrenia, depression, and Alzheimer's diseases. There is emerging interest in using imaging and other clinical data to forecast or blindly classify subjects into subgroups, for example, defined by disease status or more refined diagnostic categories. Classification or prediction of disease status based on imaging data remains an active area of research and holds promise for making a significant clinical impact. Prediction models may have a range of applications and be beneficial for clinical diagnosis, determining antecedents to a standard diagnosis, forecasting prognosis, and revealing the underlying neural basis of disease, thus informing the development of future treatments.

We use data from a study of PD as a motivating example for our proposed methods (see section 2). Neuroimaging has revealed various functional and structural alterations associated with PD. There have been reports of cortical thinning in PD patients determined from T1-MRI scans (Lee et al., 2013; Zarei et al., 2013; Zhang et al., 2015), decreased fractional anisotropy in the substantia nigra revealed by diffusion tensor imaging (DTI) (Vaillancourt et al., 2009), and functional connectivity, structural connectivity, and volumetric PD-related changes revealed by a multimodal imaging analysis (Bowman et al., 2016). These and other related studies suggest the utility of imaging data in revealing neuropathophysiology related the loss of dopamine producing neurons in PD and prompt the need for new methods to accommodate high-dimensional multimodal data.

Regularization and variable selection methods such as the least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996) and elastic-net (Zou and Hastie, 2005) as well as support vector classifiers are commonly used to predict a single scalar-valued outcome from high-dimensional data. Support vector classifiers, which arise from support vector machines (SVM), classify the data by constructing an optimal separating hyperplane in a high dimensional space to which the data are mapped (Cortes and Vapnik, 1995). Gaussian process (GP) models provide an alternative approach, which finds the posterior function that is closest to the training data based on Bayesian theory (Marquand et al., 2010). Ham and Kwak (2012) propose a boosted-principal component analysis (PCA) algorithm for binary classification problems that combines feature selection and classification.

Several methods have been proposed to predict follow-up imaging scans from baseline scans (Guo et al., 2008; Derado et al., 2013). Guo et al. (2008) propose a Bayesian hierarchical model for functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) data; Derado et al. (2013) extends the model by introducing both spatial correlations between voxels and temporal correlations between baseline and follow-up functional imaging scans. For structural data, Stonnington et al. (2010) propose a relevance vector regression (RVR) model to predict the clinical scores using MRI T1 weighted scans.

Predicting disease status utilizes a potentially massive number of independent variables that exhibit unknown patterns of correlation. The prediction and classification models described above do not estimate the spatial correlations in imaging scans or capture the associations between different imaging modalities. We build on ideas of spatial modeling for correlated imaging data for our prediction framework. Specifically, we propose a novel Bayesian hierarchical model to predict disease status using imaging scans of different modalities in both gray and white matter to reflect the functional as well as the structural properties of the brain. We consider a two-level brain parcellation, dividing the brain into defined regions as well as subregions within regions, and assume different spatial correlation structures between voxels within a subregion, within a region, and in different regions. We perform Markov Chain Monte Carlo (MCMC) estimations via Gibbs sampling. The predictions for disease status are conducted based on the predictive posterior

probabilities. Both whole-brain and voxel-level predictions are performed using leave-one-out cross validation (LOOCV). Also, we conduct feature selection to identify the regions that are associated with the disease based on the voxel-level prediction results. We apply our approach to a PD study and conduct simulation studies to evaluate its performance.

## 2. PARKINSON'S DISEASE DATA

This research qualifies as Research of Existing Data, Records, Specimens [Basic Exempt Criteria 45 CFR 46.101(b)(4)], and has been deemed Not Human Subjects Research (HS Code 10 in IPMAC II as reference in the manual chapter 7410) by NIH and Columbia University Medical Center Institutional Review Board (Protocol: IRB-AAAO0062).

The data were originally collected at Emory University (P50-NS071669) and were supplied to Columbia with all subjects' records de-identified. Written and informed consent was obtained from all research participants at the time of data collection.

A total of 20 subjects, 11 of which are diagnosed as early to moderate PD patients and the rest are healthy controls, are included in the study. The average age is 66 (*s.d.* = 11) years, and 12 of the subjects are males. The mean duration of disease was 8.4 years (*s.d.* = 3.3). The average height is 175 cm and the average weight is 79 kg. Resting-state fMRI scans, and T1-weighted MRI scans, and diffusion tensor imaging (DTI) scans are obtained.

A Siemens Trio Tim 3T MRI scanner was used to capture all the imaging scans. MPRAGE was used to acquire the structural T1 scans ( $TR = 2,600$  ms,  $TE = 3$  ms, 192 sagittal slices at 1 mm;  $256 \times 232$  1 mm isotropic pixels). The resting-state fMRI scans were acquired using echo planar imaging (EPI) ( $TR = 3,000$  ms,  $TE = 30$  ms, 48 axial slices at 3 mm,  $128 \times 128$  2 mm isotropic pixels) for each subject. DTI data were captured using a biphasic approach with consecutive left-to-right and right-to-left phase scans. The subjects followed a DTI protocol ( $TR = 8,700$  ms,  $TE = 94$  ms, 64 axial slices at 2 mm,  $128 \times 128$  2 mm isotropic pixels) comprised of 64 directions ( $B = 1,000$  s/mm<sup>2</sup>), with three leading and three trailing B0 scans.

We extract voxel-level information from these three imaging modalities, including fractional amplitude of low-frequency fluctuation (fALFF) from resting-state fMRI scans, voxel based morphometry (VBM) from T1-weighted MRI scans, and fractional anisotropy (FA) from DTI scans. fALFF reflects the amplitude of spontaneous blood-oxygen-level-dependent (BOLD) signal fluctuations of each voxel. VBM measures the localized gray matter volume changes in each voxel after spatially normalizing all the images to a standard space, and extracting gray matter from the normalized images (Ashburner and Friston, 2000). FA has a single value for each voxel, measuring the difference in directions along different axes of the random motion of water molecules in the brain, which reflects the physical orientation of white-matter fibers at that location. In summary, fALFF provides functional information, while FA and VBM describe structural properties of the brain.

The image preprocessing was performed using statistical parametric mapping (SPM) (Wellcome Department of Cognitive

Neurology, <http://www.fil.ion.ucl.ac.uk/spm>) and FMRIB (Functional Magnetic Resonance Imaging of the Brain) Software Library (FSL) (Smith et al., 2004). Resting state preprocessing consisted of a despiking stage, slice time correction, motion correction, spatial normalization to MNI and smoothing by 6 mm FWHM. The time courses were filtered to the band 0.01–0.1 Hz.

### 3. METHODS

We propose a novel Bayesian hierarchical model to predict disease status using imaging data from different modalities, including fALFF, VBM, and FA. For resting-state fMRI scans and DTI scans, the functional and structural information lies in gray matter and white matter, respectively. Most VBM analyses focus on gray matter, which will be the focus of our upcoming data example; however, applications of VBM in white matter has also been found to be associated with psychiatric diseases such as Alzheimer’s diseases and schizophrenia (Di et al., 2009; Li et al., 2011). Potentially, our prediction model involves the voxels in gray and/or white matter for different imaging modalities.

#### 3.1. Model and Estimation

We consider a two-level brain parcellation, initially consisting of  $G = 90$  brain regions defined by the automated anatomic labeling (AAL) system (Tzourio-Mazoyer et al., 2002). In each region  $g$ , we define  $L_g$  subregions, ranging from 1 to 9, for  $g = 1, \dots, G$ . The subregions are built based on the brain parcellation algorithms described in Appendix 1 (Supplementary Material). Each subregion  $l$  is composed of  $V_l$  voxels. Let  $X_{ilg}(v)$ ,  $Y_{ilg}(v)$  and  $Z_{ilg}(v)$  respectively denote the observed fALFF, FA and VBM measures for subject  $i$  at voxel  $v$  in subregion  $l$  and region  $g$ , for  $i = 1, \dots, n$ ,  $v = 1, \dots, V_l$ ,  $l = 1, \dots, L_g$ . Let  $N_g(l) \subseteq \{1, \dots, L_g\}$  denote the neighboring subregions of subregion  $l$ , constrained to fall within region  $g$ , and  $n_{lg}$  is the number of members in  $N_g(l)$ . In our model, all the subregions in region  $g$  are considered as neighbors of subregion  $l$ ; therefore, we have  $N_g(l) = \{1, \dots, L_g\}$ , and  $n_{lg} = L_g$ . Let  $D_i \in \{0, 1\}$  denote the disease status (here, PD), where 1 indicates PD; and  $\mathbf{W}_i = (W_{i1}, \dots, W_{iQ})$  denotes the vector of  $Q$  covariates. Let  $\mathcal{B}$ ,  $\mathcal{W}$  and  $\mathcal{G}$  respectively represent the whole brain region, the white matter region and the gray matter region.

We propose a model that accounts for the spatial correlations between voxels within the same subregion, between subregions within the same region, and between regions. Building spatial correlations into our model captures associations between different brain regions and generally improves the precision of estimates by borrowing strength from other (sub)regions. First, our model assumes consistent correlations between voxels in a same subregion. Then the spatial correlations between subregions within the same AAL region are described by a conditional autoregressive (CAR) model, which allows the estimates at subregion levels to borrow strength from their neighbors within the same AAL region. In addition, we introduce unstructured spatial correlations between AAL regions.

Our model reflects the assumption that for each voxel  $v$  in the gray matter, the fALFF value  $X_{ilg}(v)$  follows a normal distribution

conditioning on the VBM value  $Z_{ilg}(v)$ ; for each voxel  $v$  in the white matter, the FA value  $Y_{ilg}(v)$  follows a normal distribution conditioning on the VBM value  $Z_{ilg}(v)$ ; and for each voxel  $v$  included in the analysis, the VBM value  $Z_{ilg}(v)$  follows a normal distribution. The proposed model has the following hierarchical structure:

$$\begin{aligned} \text{For any } v \in \mathcal{G}, [X_{ilg}(v) | Z_{ilg}(v), D_i, \mathbf{W}_i, \bullet] &\sim N\left\{\mu_{lg}^{xz}(v), \delta_{lg}^{xz}\right\}, \\ \text{for any } v \in \mathcal{W}, [Y_{ilg}(v) | Z_{ilg}(v), D_i, \mathbf{W}_i, \bullet] &\sim N\left\{\mu_{lg}^{yz}(v), \delta_{lg}^{yz}\right\}, \\ \text{for any } v \in \mathcal{B}, [Z_{ilg}(v) | D_i, \mathbf{W}_i, \bullet] &\sim N\left\{\mu_{lg}^z(v), \delta_{lg}^z\right\}, \end{aligned}$$

where

$$\begin{aligned} \mu_{lg}^{xz}(v) &= \sum_{k=0,1} [c_{klg}^{xz}(v)(Z_{ilg}(v) - \bar{Z}_{lg}(v)) + \mathbf{W}_i \boldsymbol{\gamma}_{klg}^x(v) + \beta_{klg}^x(v) \\ &\quad + \alpha_{ilg}^x + \eta_{kg}^x] I(D_i = k), \\ \mu_{lg}^{yz}(v) &= \sum_{k=0,1} [c_{klg}^{yz}(v)(Z_{ilg}(v) - \bar{Z}_{lg}(v)) + \mathbf{W}_i \boldsymbol{\gamma}_{klg}^y(v) + \beta_{klg}^y(v) \\ &\quad + \alpha_{ilg}^y + \eta_{kg}^y] I(D_i = k), \\ \mu_{lg}^z(v) &= \sum_{k=0,1} (\mathbf{W}_i \boldsymbol{\gamma}_{klg}^z(v) + \beta_{klg}^z(v) + \alpha_{ilg}^z + \eta_{kg}^z) I(D_i = k). \end{aligned}$$

We assume that the probability of disease status  $P(D_i = k_i)$  is a constant, and independent of all the parameters. Also, we assume conditional independence among voxel measures of the same modality within the same subregion. The mean structure of the model is composed of several parameters, conditional on disease status.  $c_{klg}(v)$  is the slope term for centered VBM values;  $\boldsymbol{\gamma}_{klg}(v) = (\gamma_{klg1}(v), \dots, \gamma_{klgQ}(v))'$  is the parameter vector for covariates;  $\beta_{klg}(v)$ ,  $\alpha_{ilg}$ , and  $\eta_{kg}$  are the voxel-level intercept term, subregion level random effect, and region level intercept term, respectively. Each imaging modality is assumed to have the same subregion-level variance  $\delta_{lg}$  for both subject groups.

The prior beliefs about the parameters included in the likelihood function are expressed in the second or lower levels of the model.

We also assume that

$$\begin{aligned} c_{klg}^{xz}(v) &\sim N(\zeta_{klg}^{xz}, \omega_{klg}^{xz}), & \zeta_{klg}^{xz} &\sim N(a_\zeta, b_\zeta), & \omega_{klg}^{xz} &\sim \text{InvG}(a_\omega, b_\omega), \\ c_{klg}^{yz}(v) &\sim N(\zeta_{klg}^{yz}, \omega_{klg}^{yz}), & \zeta_{klg}^{yz} &\sim N(a_\zeta, b_\zeta), & \omega_{klg}^{yz} &\sim \text{InvG}(a_\omega, b_\omega), \\ \boldsymbol{\gamma}_{klgq}^m(v) &\sim N(0, s_{klg}^m), & s_{klg}^m &\sim \text{InvG}(a_s, b_s), \\ \beta_{klg}^m(v) &\sim N(\beta_{klg}^m, \lambda_{klg}^m), & \lambda_{klg}^m &\sim \text{InvG}(a_\lambda, b_\lambda), \end{aligned}$$

The slope  $c_{klg}(v)$  follows a normal distribution, whose mean and variance are drawn from noninformative hyperpriors. Each covariate parameter  $\gamma_{klgq}(v)$  is assumed to arise from a normal mean-zero distribution with variance  $s_{klg}$ , which has a noninformative hyperprior distribution. Parameters  $\beta_{klg}(v)$  that fall within the same subregion are assumed to follow normal distributions with common mean  $\beta_{klg}$ , and variance  $\lambda_{klg}$ . We assume a noninformative distribution for  $\lambda_{klg}$ , and as described

in detail below, we use a spatial prior for  $\beta_{klg}$  to incorporate spatial correlations between subregions.  $\eta_k$  follows a multivariate normal distribution with mean  $\mathbf{0}$  and covariance matrix  $\Sigma_k$  whose off-diagonal elements capture spatial dependence between AAL regions. Spatial associations between voxels within each subregion are introduced by the individualized random effect term  $\alpha_{ilg}$ , which follows a mean-zero normal distribution with variance  $\tau_{lg}^m$ , thus assuming the same spatial correlations between voxels in the same subregion.

We assume a CAR model for  $\beta_{klg}^m$  as follows:

$$[\beta_{klg}^m | \{\beta_{kl'g}^m\}_{l' \neq l, \bullet}] \sim N \left\{ \frac{\rho_g^m}{n_{lg}} \sum_{l' \in N_g(l)} \beta_{kl'g}^m, \frac{\phi_g^m}{n_{lg}} \right\},$$

$$\rho_g^m \sim U(\{0, 0.05, 0.1 \dots, 0.8, 0.81, \dots, 0.9, 0.91, \dots, 0.99\}),$$

$$\phi_g^m \sim \text{InvG}(a_\phi, b_\phi),$$

$$\alpha_{ilg}^m \sim N(0, \tau_{lg}^m), \quad \tau_{lg}^m \sim \text{InvG}(a_\tau, b_\tau),$$

$$\eta_k^m = (\eta_{k1}^m, \dots, \eta_{kG}^m)' \sim N(0, \Sigma_k^m), \quad \Sigma_k^m \sim \text{InvW}(\Lambda, \nu),$$

$$\delta_{lg}^{xz} \sim \text{InvG}(a_\delta, b_\delta), \quad \delta_{lg}^{yz} \sim \text{InvG}(a_\delta, b_\delta),$$

$$\delta_{lg}^z \sim \text{InvG}(a_\delta, b_\delta), \quad \text{where } m \in \{x, y, z\}.$$

By assuming a subregion level CAR model, we capture the spatial dependence between subregions within each AAL region. In the model,  $\rho_g$  represents the overall degree of spatial dependence in region  $g$  and  $\frac{\phi_g}{L_g}$  is the conditional variance of  $\beta_{klg}$ . The neighborhood of subregion  $l \in g$ , is defined as all the other subregions in AAL region  $g$ . The spatial neighborhood effect  $\rho_g$  is assumed to follow a discrete uniform distribution (Gelfand and Vounatsou, 2003). As we would like to identify the similarity of the neighboring subregions, we impose  $0 \leq \rho_g < 1$ . Specifically, equal mass is put on the following 36 values: 0, 0.05, 0.1, ..., 0.8, 0.81, 0.82, ..., 0.90, 0.91, 0.92, ..., 0.99, which includes a more refined set of values in the upper range of  $\rho_g$  since estimation of  $\rho_g$  for imaging data often tends toward large values.

For any disease status  $k$ , the covariance between the voxels within a same subregion  $l$  in region  $g$  is contributed by the variance from three components:  $\beta_{klg}$ ,  $\alpha_{ilg}$ , and  $\eta_{kg}$ ; the covariance between the voxels in two subregions  $l$  and  $l'$ , but the same AAL region  $g$ , comes from the covariance between  $\beta_{klg}$  and  $\beta_{kl'g}$ , and the variance of  $\eta_{kg}$ ; and the covariance between the voxels in two AAL regions  $g$  and  $g'$  is determined by the covariance of  $\eta_{kg}$  and  $\eta_{kg'}$ .

We perform estimation using Markov chain Monte Carlo (MCMC) implemented via Gibbs sampling. The full conditional posterior distributions are shown in Appendix 2 (Supplementary Material).

### 3.2. Prediction

#### 3.2.1. Whole Brain Prediction

The objective of our model is to predict PD status, given imaging data and other covariates. To achieve this goal, we use

the posterior samples drawn from estimation to calculate the posterior predictive probability of disease status.

Let  $\theta$  denote the parameter space,  $\mathbf{B}_i = (\mathbf{X}_{ilg}, \mathbf{Y}_{ilg}, \mathbf{Z}_{ilg})$  denote the observed imaging data for subject  $i$ , and  $\mathbf{A}_i = (\mathbf{B}_i, D_i)$  denote the combination of the imaging data and disease status. Suppose we have  $n$  training subjects, and we want to predict the disease status  $D_{n+1}$  for a new subject indexed by  $n + 1$ . The posterior predictive distribution for  $D_{n+1}$  is given by

$$\begin{aligned} &P(D_{n+1} = k | \mathbf{B}_{n+1}, \{\mathbf{A}_i\}_{i=1}^n) \\ &= \frac{P(D_{n+1} = k, \mathbf{B}_{n+1} | \{\mathbf{A}_i\}_{i=1}^n)}{\sum_{k'=0,1} P(D_{n+1} = k', \mathbf{B}_{n+1} | \{\mathbf{A}_i\}_{i=1}^n)} \\ &= \frac{P(D_{n+1} = k) \int_{\theta} P(\mathbf{B}_{n+1} | D_{n+1} = k, \theta) P(\theta | \{\mathbf{A}_i\}_{i=1}^n) d\theta}{\sum_{k'=0,1} P(D_{n+1} = k') \int_{\theta} P(\mathbf{B}_{n+1} | D_{n+1} = k', \theta) P(\theta | \{\mathbf{A}_i\}_{i=1}^n) d\theta}, \end{aligned} \tag{1}$$

where

$$\begin{aligned} P(\mathbf{B}_{n+1} | D_{n+1} = k, \theta) &= \prod_{v \in \mathcal{G}} P(X_{n+1}(v) | Z_{n+1}(v), D_{n+1} = k, \theta) \\ &P(Z_{n+1}(v) | D_{n+1} = k, \theta) \\ &\prod_{v \in \mathcal{W}} P(Y_{n+1}(v) | Z_{n+1}(v), D_{n+1} = k, \theta) \\ &P(Z_{n+1}(v) | D_{n+1} = k, \theta), \end{aligned} \tag{2}$$

Suppose we draw  $T$  posterior samples, denoted  $\theta^{(t)}$ , from  $P(\theta | \{\mathbf{A}_i\}_{i=1}^n)$ , for  $t = 1, \dots, T$ . Letting  $\pi_k^{(t)} = P(\mathbf{B}_{n+1} | D_{n+1} = k, \theta^{(t)})$ , the posterior predictive probability can be expressed by

$$\hat{P}(D_{n+1} = k | \mathbf{B}_{n+1}, \{\mathbf{A}_i\}_{i=1}^n) = \frac{P(D_{n+1} = k) \sum_{t=1}^T \pi_k^{(t)}}{\sum_{k'=0,1} P(D_{n+1} = k') \sum_{t=1}^T \pi_k^{(t)}}. \tag{3}$$

Then ultimately the prediction of  $D_{n+1}$  is given by

$$\hat{D}_{n+1} = \arg \max_k \left( P(D_{n+1} = k) \sum_{t=1}^T \pi_k^{(t)} \right). \tag{4}$$

To evaluate the performance of our method, we calculate the prediction accuracy using LOOCV.

Applied directly, LOOCV is very computational expensive because it involves multiple posterior simulations with tens of thousands voxels included in the analysis. Therefore, we employ an importance sampling approach to reduce the computation for LOOCV of our model (Gelfand et al., 1992; Gelfand, 1996; Alqallaf and Gustafson, 2001; Vehtari and Lampinen, 2002). Specifically, the LOOCV predictive probabilities can be expressed by

$$P(D_i = k | \mathbf{B}_i, \mathbf{A}_{-i}) = \frac{P(D_i = k) Q_{kd_i}}{\sum_{k'=0,1} P(D_i = k') Q_{k'd_i}}, \tag{5}$$

where

$$Q_{kd_i} = \int \frac{P(\mathbf{B}_i | D_i = k, \theta)}{P(\mathbf{B}_i | D_i = d_i, \theta)} P(\theta | \{\mathbf{A}_i\}_{i=1}^n) d\theta, \tag{6}$$

and  $d_i$  is the observed disease status for subject  $i$ . Next, we provide the details of how  $Q_{kd_i}$  is derived. The posterior predictive probability can be written as follows:

$$\begin{aligned}
 P(D_i = k \mid \mathbf{B}_i, \mathbf{A}_{-i}) &= \int P(D_i = k \mid \mathbf{B}_i, \boldsymbol{\theta}) \frac{P(\boldsymbol{\theta} \mid \mathbf{B}_i, \mathbf{A}_{-i})}{P(\boldsymbol{\theta} \mid \mathbf{B}_i, D_i = d_i, \mathbf{A}_{-i})} \\
 &P(\boldsymbol{\theta} \mid \mathbf{B}_i, D_i = d_i, \mathbf{A}_{-i}) d\boldsymbol{\theta}. \tag{7}
 \end{aligned}$$

Therefore,

$$\frac{P(D_i = k \mid \mathbf{B}_i, \mathbf{A}_{-i})}{P(D_i = d_i \mid \mathbf{B}_i, \mathbf{A}_{-i})} = \frac{P(D_i = k)}{P(D_i = d_i)} Q_{kd_i}. \tag{8}$$

By using the fact that  $\sum_{k=0,1} P(D_i = k \mid \mathbf{B}_i, \mathbf{A}_{-i}) = 1$ , we have

$$P(D_i = d_i \mid \mathbf{B}_i, \mathbf{A}_{-i}) = \frac{P(D_i = d_i)}{\sum_{k=0,1} P(D_i = k) Q_{kd_i}}, \tag{9}$$

thus leading to the above LOOCV predictive probability Equation (5). For  $i = 1, \dots, n$  and  $k = 0, 1$ , compute

$$\hat{Q}_{kd_i} = \frac{1}{T} \sum_{t=1}^T \frac{P(\mathbf{B}_i \mid D_i = k, \boldsymbol{\theta}^{(t)})}{P(\mathbf{B}_i \mid D_i = d_i, \boldsymbol{\theta}^{(t)})}. \tag{10}$$

The estimate of  $D_i$  is

$$\hat{D}_i = \arg \max_k (P(D_i = k) Q_{kd_i}). \tag{11}$$

Since there are only two possible values for  $D_i$ , we only need to calculate  $P(\mathbf{B}_i \mid D_i = k, \boldsymbol{\theta}^{(t)})$  and  $P(\mathbf{B}_i \mid D_i = d_i, \boldsymbol{\theta}^{(t)})$ , where  $k \neq d_i$ , for each subject  $i$ .

### 3.2.2. Voxel-Level Prediction

We also consider the use of imaging data  $\mathbf{B}_i(v) = (X_{ilg}(v), Y_{ilg}(v), Z_{ilg}(v))$  for subject  $i$  at voxel  $v$  to predict the disease status  $D_i$ . Similar to Equation (5), the voxel-level LOOCV predictive probabilities can be expressed by

$$P(D_i = k \mid \mathbf{B}_i(v), \mathbf{A}_{-i}) = \frac{P(D_i = k) Q_{kd_i}}{\sum_{k'=0,1} P(D_i = k') Q_{k'd_i}}, \tag{12}$$

where

$$Q_{kd_i} = \int \frac{P(\mathbf{B}_i(v) \mid D_i = k, \boldsymbol{\theta}) / \sum_{k'=0,1} P(\mathbf{B}_i(v) \mid D_i = k', \boldsymbol{\theta}) P(D_i = k')}{P(\mathbf{B}_i \mid D_i = d_i, \boldsymbol{\theta}) / \sum_{k'=0,1} P(\mathbf{B}_i \mid D_i = k', \boldsymbol{\theta}) P(D_i = k')} P(\boldsymbol{\theta} \mid \{\mathbf{A}_i\}_{i=1}^n) d\boldsymbol{\theta}, \tag{13}$$

which is estimated by

$$\hat{Q}_{kd_i} = \frac{1}{T} \sum_{t=1}^T \frac{P(\mathbf{B}_i(v) \mid D_i = k, \boldsymbol{\theta}^{(t)}) / \sum_{k'=0,1} P(\mathbf{B}_i(v) \mid D_i = k', \boldsymbol{\theta}^{(t)}) P(D_i = k')}{P(\mathbf{B}_i \mid D_i = d_i, \boldsymbol{\theta}^{(t)}) / \sum_{k'=0,1} P(\mathbf{B}_i \mid D_i = k', \boldsymbol{\theta}^{(t)}) P(D_i = k')}. \tag{14}$$

Then the estimate of  $D_i$  is

$$\hat{D}_i = \arg \max_k (P(D_i = k) Q_{kd_i}), \tag{15}$$

which is equivalent to

$$\hat{D}_i = \arg \max_k \left( P(D_i = k) \frac{1}{T} \sum_{t=1}^T P(\mathbf{B}_i(v) \mid D_i = k, \boldsymbol{\theta}^{(t)}) \right). \tag{16}$$

$Q_{kd_i}$  is derived in the similar way as in the whole brain analysis. The derivation of Equation (14) is described in Appendix 3 (Supplementary Material).

The voxel-level prediction result can be used as a way to select the regions that are highly associated with PD if the prediction accuracy is high in these regions. An alternative approach to perform feature selection using our model is discussed in section 5.

## 4. RESULTS

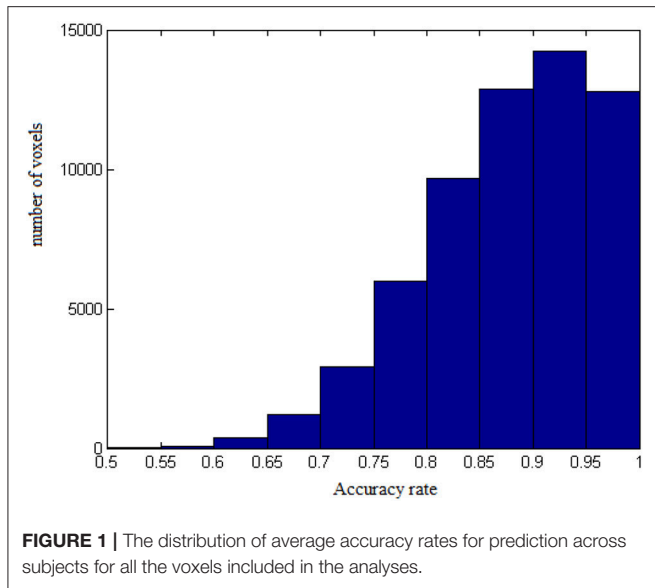
### 4.1. Parkinson's Disease Data

We applied our proposed Bayesian spatial model to PD data, which has T1 and resting-state fMRI images available; therefore, our model reduces to one which includes two imaging modalities, VBM and fALFF, and only considers data in the gray matter. We generate predictions of PD based on multimodal imaging data aggregated across the whole brain, and we provide voxel-level predictions as well. By evaluating the prediction accuracy at each voxel, we are able to identify brain regions that are highly associated with Parkinson's disease as an alternative to performing feature selection.

In the estimation procedure, the hyperparameters for the prior distribution are set to provide vague information to ensure that the results are dominated by the information in the data. Specifically, all the hyperparameters in the inverse-gamma distribution are set to  $10^{-3}$  (Spiegelhalter et al., 1994/2003), the normal prior for  $\zeta_{klg}$  is assumed to have mean  $a_\zeta = 0$  and variance  $b_\zeta = 10^5$ . In the inverse-Wishart distribution, the degrees of freedom  $\nu$  should be greater than  $G - 1$  to build a proper distribution, so we set  $\nu = G$ , which provides the least information based on our data. The scale matrix  $\Lambda$  is set as  $10^{-3} \times \mathbf{I}_G$ , where  $\mathbf{I}_G$  is a  $G \times G$  identity matrix.

We perform a total of 10,000 MCMC iterations including 5,000 burn-in iterations, and store the results thinning by 10. Due

to the massive number of parameters in our model, we randomly check trace plots for parameters at the voxel-level, subregion-level, and region-level, respectively. We provide some examples in Appendix 4 (Supplementary Material).



**FIGURE 1 |** The distribution of average accuracy rates for prediction across subjects for all the voxels included in the analyses.

**TABLE 1 |** Summary of average accuracy rates for prediction across subjects.

Accuracy rate	[80%, 85%)	[85%, 90%)	[90%, 95%)	[95%, 100%)	100%
Number of voxels	5,993	9,663	12,878	14,236	12,764
(Percentage)	(9.97)	(16.07)	(21.42)	(23.68)	(21.23)

After estimating the model parameters, we perform a whole-brain and voxel-level prediction using posterior samples based on procedures described in section 3.2. Here, we have a total of 500 posterior samples after thinning. By assuming an equal probability for classification as a PD patient and a control subject, our model achieves 100% accuracy from the whole-brain prediction based on LOOCV.

The results from voxel-level prediction provide interesting information as well. The highest voxel-level accuracy rate is 100%, and the lowest is near 50%. **Figure 1** shows the distribution of the average accuracy rate across subjects for all the voxels included in the analysis. **Table 1** gives the number of voxels (percentage) achieving accuracy rates higher than 80%. Also, an average whole-brain prediction map based on the results from voxel-level prediction across subjects are presented in **Figure 2**.

To identify the regions which are predictive for disease status, we compute the average accuracy rates across voxels within a region, and **Table 2** shows the regions that have accuracy rate above 95%. **Table 2** also shows the percentage of voxels exceeding 90% accuracy rates for those regions. The right rectus gyrus, which is associated with cognitive impairment in PD patients, and is shown to have different gray matter density between PD and controls (Nagano-Saito et al., 2005), is identified in our analysis. The precentral gyrus, which is part of the primary motor cortex, is identified among the most accurate brain regions, and its performance is consistent with the involvement of this region in planning and initiating motor movements, which is critically impaired in patients with PD. We also find the

**TABLE 2 |** List of regions with above 95% average accuracy rate across voxels.

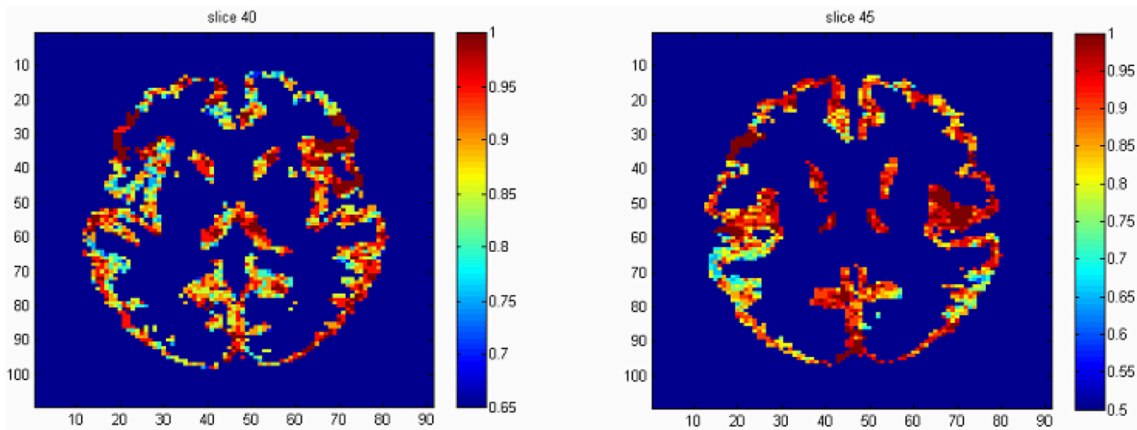
Region	Accuracy rate	Percentage of voxels with accuracy rate > 90%
Left postcentral	99.9%	42.6%
Right rectus	99.3%	52.2%
Left inferior parietal	99.3%	90.2%
Right superior medial frontal	99.0%	61.1%

bilateral caudate and the left putamen as regions with accurate predictions. The caudate and putamen, two regions comprising the dorsal striatum, exhibit marked pathologic changes from PD, linked to the loss of dopaminergic neurons in the substantia nigra which projects to striatal neurons in the caudate nucleus and putamen (Spencer et al., 1992). The right fusiform gyrus, which is believed to related to impaired ability to correctly identify negative facial expressions (Geday et al., 2006), and the left inferior parietal lobule which is involved in the perception of emotions in facial stimuli, may play a role of differentiating healthy controls and PD patients as well. Other regions which are involved in face perception such as the right mid-temporal pole are also identified. The left postcentral gyrus, the left superior parietal lobule, and the right superior medial frontal gyrus also stand out since all of them are part of the sensory system. A region-level prediction map based on the average accuracy rates across voxels within a region is shown in **Figure 3**.

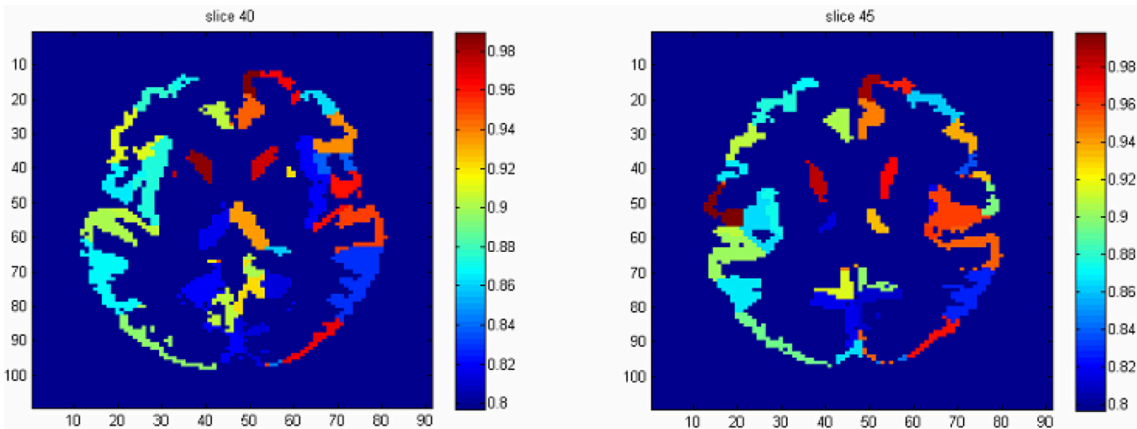
### 4.2. Simulation Studies

We conduct a simulation study to evaluate the performance of our proposed model. The purpose of this simulation study is to show that the MCMC generated samples from our model accurately target the true values and that the whole-brain prediction is accurate. In addition, we demonstrate that our model can distinguish regions that are predictive of disease status.

We assume that the imaging data are generated from the likelihood function of our model. We simulate data for 25 subjects from three AAL regions, the number of subregions within an AAL region has a mean and variance of 3, and the number of voxels within a subregion has a mean and variance of 50. We specify the true values for the parameters in the likelihood function, i.e.,  $c_{klg}(v)$ ,  $\gamma_{klg}(v)$ ,  $\beta_{klg}(v)$ ,  $\alpha_{ilg}$ ,  $\eta_{kg}$ , and  $\delta_{lg}$ , which are the most relevant parameters for voxel-level inference and future prediction. In this way, we can compare our posterior estimations with specified true values. All the other parameters are updated from the posterior distributions. And the hyperparameters are set to be the same as in data from PD study. We select some subregions to be the ones that are associated with PD, and a region is classified into this category if it contains those selected subregions. We set different true values of parameters for disease and non-disease group if they are within the pre-specified regions and otherwise assume that the true values are the same for two groups. A total of 100 data sets are drawn in the simulation study. The programming is implemented in Matlab, and the computation is performed on a Linux cluster with 16 GB of RAM. Execution time is approximately 3–4 h for one data set.



**FIGURE 2** | The average prediction map based on the voxel-level prediction results across subjects.



**FIGURE 3** | The region-level prediction map based on the average accuracy rates across voxels within a region.

First, we evaluate the posterior estimates by comparing the posterior means to the true values. Instead of examining a total of five thousand parameters which have known true values separately, we calculate the mean structure and variance of the likelihood function from posterior samples and compare them to the truth since they are the most essential for inferences and predictions. The average bias (percentage change) in mean structure is  $3.52 \times 10^{-2}$  (0.54%), and in variance is  $1.04 \times 10^{-5}$  (1.04%). Secondly, we calculate the accuracy of whole-brain prediction. The LOOCV achieves 100% for the whole-brain prediction for all 100 simulated data sets. Thirdly, we identify the regions that are highly associated with disease status by evaluating the voxel-level accuracy rates for prediction. We compare the average accuracy for voxel-level prediction between the pre-specified regions and the others. Within the pre-specified regions, the average accuracy rate is 99.8%; for voxels which are in the other regions, the average accuracy rate is 71.7%. Here, we can see an improvement in prediction when voxels are from the pre-specified regions.

In comparison, we apply the elastic-net model to the simulated data as described above, and the LOOCV achieves an average of 86% accuracy rate for the whole-brain prediction.

In summary, our model accurately performs posterior estimation with small bias, provides accurate prediction of PD status using whole-brain imaging data, and correctly identifies the regions that are highly associated with disease.

## 5. DISCUSSION

We propose a Bayesian spatial model to predict PD using different modalities of imaging data, including fALFF, VBM, and FA in gray and white matter. Our framework performs voxel-level estimation for imaging data and conducts whole-brain and voxel-level prediction of disease status based on posterior predictive probabilities. Our model estimates both the mean and covariance structures of imaging data, predicting disease status using whole-brain imaging data, and identifying the regions which are highly associated with the disease based on voxel-level prediction results.

In our framework, we consider spatial correlations at voxel level, subregion level, and region level, and specify different correlation structures such as exchangeable, CAR, and unstructured correlation matrices for them. The rich hierarchical spatial correlation structures captured by our model extends previous spatial modeling frameworks by Bowman et al. (2008) and Derado et al. (2013). The intra-subregion correlation in our model is described by a single value within each subregion; the inter-subregion correlation is modeled by a CAR model which borrows information from the subregions within the same parent AAL region; the inter-region correlation is assumed to have a unstructured correlation matrix.

We derive the posterior predictive probability using whole brain data and data from a single voxel. Due to the complexity of computation, we adopt an importance sampling strategy to conduct LOOCV. The importance sampling techniques estimate the LOOCV error rate based only on one-model fitting using all samples and produces very accurate estimate on the LOOCV error rate. We evaluate the accuracy rate of the whole-brain prediction and identify the regions that are predictive for disease based on the results from voxel-level accuracy rates. Our model accounts for spatial correlations embedded in the data; however, additional multiple testing strategies could be explored to account for potential dependence inherent in the data. Our model increases localization compared to some approaches by offering voxel-level predictions. While we incorporate information from multiple modalities, we are unable to dissociate the relative predictive accuracy generated by each modality.

One weakness of our method is computational time since we use a joint model that performs estimation at the voxel-level. However, by applying the importance sampling strategy, we only need to perform the posterior estimation once, and then the posterior predictive probabilities can be computed fairly efficiently.

Compared to the existing feature selection methods, e.g., LASSO or elastic-net, our model uses a different modeling strategy and different criteria for selection. LASSO and elastic-net model the probability of PD, while our method starts from modeling the imaging data. This distinction leads to an important advantage that we are able to estimate and borrow strength from the spatial correlations in the data, whereas highly correlated predictors often lead to poor performance of the LASSO and related methods. Also, we use posterior predictive probability as the criteria to select the features, which is the exact target of prediction problems; on the other hand, LASSO and elastic-net, from a Bayesian perspective, use posterior modes to perform feature selection. Our model also has interpretive advantages over SVM and GP models by identifying particular voxels, subregions, or regions that contribute significantly to accurate prediction. Compared to the methodology of scalar-on-image regression

(Goldsmith et al., 2014; Reiss et al., 2015; Kang et al., 2016; Wang et al., 2017), our method models the images as the response, which is a natural generative process, and then we predict the disease distribution given the imaging scans.

In summary, the advantages of our proposed Bayesian method are three fold. First, it is more straightforward to incorporate prior knowledge regarding brain function and structure, which is extremely useful to improve the prediction accuracy and to provide a better understanding of the etiology. Second, it yields estimates and inference from the full posterior distribution, e.g., rather than point estimates. In particular, it can provide measures of uncertainty of the predictions based on the posterior predictive distribution. In addition, the posterior computation based on the MCMC algorithm is more robust to complex imaging data, while the optimization algorithms for other frequentist prediction methods are more likely to be trapped at the local modes, which may reduce the prediction accuracy.

In our method, we select features based on the posterior predictive probability of each voxel; ideally, we would like to identify the voxels  $v \in \mathcal{V}$  s.t.

$$P(D_i = k \mid \{\mathbf{B}_i(v)\}_{v \in \mathcal{V}}, \mathbf{A}_{-i}) = P(D_i = k \mid \mathbf{B}_i, \mathbf{A}_{-i}), \quad (17)$$

which could be a possible extension of our proposed approach.

For Parkinson's disease, our model may not immediately supplant current clinical standards to diagnose patients at or near the manifestation of motor symptoms. However, our model stands to provide insights into the useful information for the diagnosis of PD, underlying neurophysiological basis of the disease, potentially early pre-motor alterations, and effective strategies to design studies examining potential neuroprotective treatments with consideration of the cost and complexity as well as extensive validation and comparison to current standards.

## AUTHOR CONTRIBUTIONS

WX: methodology, simulation, and writing. FB and JK: supervising and editing the manuscript.

## FUNDING

This research was funded by a grant from the NINDS (U18 NS082143) at NIH as part of the Parkinson's Disease Biomarker Program.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2018.00184/full#supplementary-material>

## REFERENCES

- Alqallaf, F., and Gustafson, P. (2001). On cross-validation of Bayesian models. *Can. J. Stat.* 29, 333–340. doi: 10.2307/3316081
- Ashburner, J., and Friston, K. J. (2000). Voxel-based morphometry—the methods. *NeuroImage* 11, 805–821. doi: 10.1006/nimg.2000.0582

- Bowman, F. D., Caffo, B. S., Spear Bassett, S., and Kilts, C. (2008). A Bayesian hierarchical framework for spatial modeling of fMRI data. *NeuroImage* 39, 146–156. doi: 10.1016/j.neuroimage.2007.08.012
- Bowman, F. D., Drake, D. F., and Huddleston, D. (2016). Multimodal imaging signatures of Parkinson's disease. *Front. Neurosci.* 10:131. doi: 10.3389/fnins.2016.00131



- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018
- Derado, G., Bowman, F. D., and Zhang, L. (2013). Predicting brain activity using a Bayesian spatial model. *Stat. Methods Med. Res.* 22, 382–397. doi: 10.1177/0962280212448972
- Di, X., Chan, R. C. K., and Gong, Q. (2009). White matter reduction in patients with schizophrenia as revealed by voxel-based morphometry: an activation likelihood estimation meta-analysis. *Prog. Neuro-psychopharmacol. Biol. Psychiatry* 33, 1390–1394. doi: 10.1016/j.pnpbp.2009.08.020
- Geday, J., Ostergaard, K., and Gjedde, A. (2006). Stimulation of subthalamic nucleus inhibits emotional activation of fusiform gyrus. *NeuroImage* 33, 706–714. doi: 10.1016/j.neuroimage.2006.06.056
- Gelfand, A. E. (1996). “Inference and monitoring convergence,” in *Markov Chain Monte Carlo in Practice*, eds W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Chapman & Hall), 131–144.
- Gelfand, A. E., Dey, D. K., and Chang, H. (1992). “Model determination using predictive distributions with implementation via sampling-based methods (with discussion),” in *Bayesian Statistics*, Vol. 4, eds J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Oxford University Press), 147–167.
- Gelfand, A. E., and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* 4, 11–15. doi: 10.1093/biostatistics/4.1.11
- Goldsmith, J., Huang, L., and Crainiceanu, C. M. (2014). Smooth scalar-on-image regression via spatial Bayesian variable selection. *J. Comput. Graph. Stat.* 23, 46–64. doi: 10.1080/10618600.2012.743437
- Guo, Y., Bowman, F. D., and Kilts, C. (2008). Predicting the brain response to treatment using a Bayesian hierarchical model with application to a study of schizophrenia. *Hum. Brain Mapp.* 29, 1092–1109. doi: 10.1002/hbm.20450
- Ham, S. L., and Kwak, N. (2012). “Boosted-PCA for binary classification problems,” in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 1219–1222.
- Kang, J., Reich, B. J., and Staicu, A. (2016). Scalar-on-Image Regression via the Soft-Thresholded Gaussian Process. arXiv:1604.03192.
- Lee, E. Y., Sen, S., Eslinger, P. J., Wagner, D., Shaffer, M. L., Kong, L., et al. (2013). Early cortical gray matter loss and cognitive correlates in non-demented Parkinson’s patients. *Parkinsonism Relat. Disord.* 19, 1088–1093. doi: 10.1016/j.parkreldis.2013.07.018
- Li, J. P., Pan, P. L., Huang, R., and Shang, H. F. (2011). A meta-analysis of voxel-based morphometry studies of white matter volume alterations in Alzheimer’s disease. *Neurosci. Biobehav. Rev.* 36, 757–763. doi: 10.1016/j.neubiorev.2011.12.001
- Marquand, A., Howard, M., Brammer, M., Chu, C., Coen, S., and Mourão-Miranda, J. (2010). Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. *NeuroImage* 49, 2178–2189. doi: 10.1016/j.neuroimage.2009.10.072
- Nagano-Saito, A., Washimi, Y., Arahata, Y., Kachi, T., Lerch, J. P., Evans, A. C., et al. (2005). Cerebral atrophy and its relation to cognitive impairment in Parkinson disease. *Neurology* 64, 224–229. doi: 10.1212/01.WNL.0000149510.41793.50
- Reiss, P. T., Huo, L., Zhao, Y., Kelly, C., and Ogden, R. T. (2015). Wavelet-domain regression and predictive inference in psychiatric neuroimaging. *Ann. Appl. Stat.* 9:1076. doi: 10.1214/15-AOAS829
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23(Suppl. 1), S208–S219. doi: 10.1016/j.neuroimage.2004.07.051
- Spencer, D. D., Robbins, R. J., Naftolin, F., Marek, K. L., Vollmer, T., Leranth, C., et al. (1992). Unilateral transplantation of human fetal mesencephalic tissue into the caudate nucleus of patients with Parkinson’s disease. *New Engl. J. Med.* 327, 1541–1548. doi: 10.1056/NEJM199211263272201
- Spiegelhalter, D. J., Thomas, A., Best, N. G., Gilks, W. R., and Lunn, D. (1994/2003). *BUGS: Bayesian Inference Using Gibbs Sampling*. Cambridge: MRC Biostatistics Unit. Available online at: <http://www.mrc-bsu.cam.ac.uk/bugs/>
- Stonnington, C. M., Chu, C., Klöppel, S., Jack, C. R. Jr., Ashburner, J., Frackowiak, R. S. J., et al. (2010). Predicting clinical scores from magnetic resonance scans in Alzheimer’s disease. *NeuroImage* 51, 1405–1413. doi: 10.1016/j.neuroimage.2010.03.051
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* 58, 267–288.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* 15, 273–289. doi: 10.1006/nimg.2001.0978
- Vaillancourt, D. E., Spraker, M. B., Prodoehl, J., Abraham, I., Corcos, D. M., Zhou, X. J., et al. (2009). High-resolution diffusion tensor imaging in the substantia nigra of *de novo* Parkinson’s disease. *Neurology* 72, 1378–1384. doi: 10.1212/01.wnl.0000340982.01727.6e
- Vehtari, A., and Lampinen, J. (2002). Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Comput.* 14, 2439–2468. doi: 10.1162/08997660260293292
- Wang, X., Zhu, H., and Alzheimer’s disease Neuroimaging Initiative. (2017). Generalized scalar-on-image regression models via total variation. *J. Amer. Stat. Assoc.* 112, 1156–1168. doi: 10.1080/01621459.2016.1194846
- Zarei, M., Ibarretxe-Bilbao, N., Compta, Y., Hough, M., Junque, C., Bargallo, N., et al. (2013). Cortical thinning is associated with disease stages and dementia in Parkinson’s disease. *J. Neurol. Neurosurg. Psychiatry* 84, 875–881. doi: 10.1136/jnnp-2012-304126
- Zhang, L., Wang, M., Sterling, N., Lee, E., Eslinger, P., Wagner, D., et al. (2015). Cortical thinning and cognitive impairment in Parkinson’s disease without dementia. *IEEE/ACM Trans. Comput. Biol. Bioinformatics.* doi: 10.1109/TCBB.2015.2465951
- Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* 67, 301–320. doi: 10.1111/j.1467-9868.2005.00503.x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Xue, Bowman and Kang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.