The Canadian Journal of Statistics
Vol. 49, No. 1, 2021, Pages 10–34
La revue canadienne de statistique

Statistical disease mapping for heterogeneous neuroimaging studies

Rongjie LIU[1], and Hongtu ZHU[2*], for the Alzheimer's Disease Neuroimaging Initiative[†]

[1]*Department of Statistics, Florida State University, Tallahassee, FL, U.S.A.*
[2]*Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, U.S.A.*

Key words and phrases: hidden Markov random field model; multivariate varying coefficient model; spatial zero-inflated Poisson model; Statistical disease mapping.

MSC 2010: Primary 62M40; secondary 62J99.

Abstract: Many cancers and neuro-related diseases display significant phenotypic and genetic heterogeneity across subjects and subpopulations. Characterizing such heterogeneity could transform our understanding of the etiology of these conditions and inspire new approaches to urgently needed prevention, diagnosis, treatment, and prognosis. However, most existing statistical methods face major challenges in delineating such heterogeneity at both the group and individual levels. The aim of this article is to propose a novel statistical disease-mapping (SDM) framework to address some of these challenges. We develop an efficient estimation method to estimate unknown parameters in SDM and delineate individual and group disease maps. Statistical inference procedures such as hypothesis-testing problems are also investigated for parameters of interest. Both simulation studies and real data analysis on the ADNI hippocampal surface dataset show that our SDM not only effectively detects diseased regions in each patient but also provides a group disease-mapping analysis of Alzheimer subgroups. *The Canadian Journal of Statistics* 49: 10–34; 2021 © 2021 Statistical Society of Canada

Résumé: De nombreux cancers et maladies neurologiques présentent une hétérogénéité phénotypique et génotypique substantielle entre les sujets et les sous-populations. Réussir à caractériser cette hétérogénéité pourrait transformer notre compréhension de l'étiologie de ces états et inspirer de nouvelles approches requises sans délai pour la prévention, le diagnostic, le traitement et le pronostic. La plupart des méthodes statistiques existantes peinent à décrire une telle hétérogénéité, autant au niveau des groupes que des individus. Les auteurs proposent un nouveau cadre de cartographie statistique de la maladie (CSM) afin de relever ce défi. Ils développent une méthode efficace d'estimation pour les paramètres inconnus de la CSM et tracent des cartes de maladie pour les individus et les groupes. Les procédures d'inférence statistique, tels que les tests d'hypothèses, sont également étudiées pour les paramètres d'intérêt. Des études de simulation et une analyse de données réelles de l'initiative d'imagerie médicale pour la maladie d'Alzheimer relatives à la surface hippocampique permettent aux auteurs de montrer que la CSM détecte non seulement les régions malades pour chaque patient, mais offre également une analyse cartographique des sous-groupes d'Alzheimer. *La revue canadienne de statistique* 49: 10–34; 2021 © 2021 Société statistique du Canada

* *Author to whom correspondence may be addressed.*
E-mail: htzhu@email.unc.edu

†Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

1. INTRODUCTION

With the rapid growth of modern technology, many large-scale biomedical studies, including the Alzheimer's Disease Neuroimaging Initiative (ADNI) study (Weiner et al., 2015) and the UK Biobank data project (Allen et al., 2014), have been conducted to collect massive datasets with large volumes of complex information from increasingly large cohorts. Despite the numerous successes of biomedical studies, it has been difficult to unravel the etiology of cancers and neuro-related disorders largely due to large disease heterogeneity at the genomic, imaging, and clinical scales. Specifically, imaging heterogeneity often presents at both the global and local scales. At the global scale, diseased regions can significantly vary across subjects and/or time in terms of their number, size, and location (Davnall et al., 2012; Brooks & Grigsby, 2013; Huang et al., 2015; Bashir et al., 2016; Spagnolo et al., 2017). At the local scale, various local imaging features can have large intra- and inter-spatial heterogeneity. Understanding such imaging heterogeneity may be critical for the development of urgently needed approaches to the prevention, diagnosis, treatment, and prognosis of those diseases, as well as precision medicine broadly.

In the presence of imaging heterogeneity, the analysis of a large number of imaging measures in those large-scale studies presents major methodological challenges for many existing statistical methods (Schnack, 2019; Varol et al., 2017). For instance, standard group analysis methods, including the voxel-based morphometry in the Statistical Parametric Mapping (SPM) framework (Penny et al., 2011), require certain signal homogeneity across subjects and/or time. Therefore, they are not suitable for capturing individual diseased regions specific to a small number of subjects. In contrast, some individual analysis methods for detecting subject-specific abnormalities often compare an individual subject's data to a known reference distribution, typically derived from a cohort of healthy controls (Shaker et al., 2017). Thus, those methods can be inefficient for borrowing common information shared among a large number of diseased subjects.

The aim of this article is to propose a novel statistical disease-mapping (SDM) framework with several formal functional data analysis tools to address the above technical challenges in delineating disease heterogeneity at both group and individual scales. Our SDM consists of two components: (i) diseased region detection at the individual level and (ii) disease map construction at the group level. In the first component, a hidden Markov random field model (HMRFM) is integrated with a multivariate varying coefficient model (MVCM), where MVCM can investigate the relationship between multivariate imaging responses and covariates of interest (e.g., age, gender) and where HMRFM is adopted to detect subject-specific diseased regions. Based on the individual diseased regions detected from the first component, a spatial zero-inflated Poisson model (SZIPM) is introduced in the second component as a disease regression model to establish the disease map at the group level. The path diagram of SDM is presented in Figure 1.

Compared to the existing statistical methods for abnormal pattern detection in the literature (Van Leemput et al., 2001; Matteoli, Diani & Corsini, 2010; Penny et al., 2011; Ramteke & Monali, 2012; Goldstein & Uchida, 2016; Shaker et al., 2017), four major distinctive contributions of this article in terms of both methodology and application are as follows:

- This article proposes a special image-on-scalar regression model that integrates HMRFM with MVCM and preserves the key features from functional data analysis tools and Markov random field models. Specifically, our SDM can not only build up the relationship between functional phenotypes and a set of covariates of interest (Zhu et al., 2011; Zhu, Li & Kong, 2012; Huang et al., 2017) but also detect individual diseased regions (Geman & Geman, 1984; Huang et al., 2015). In addition, compared to voxel-wise analysis, our SDM

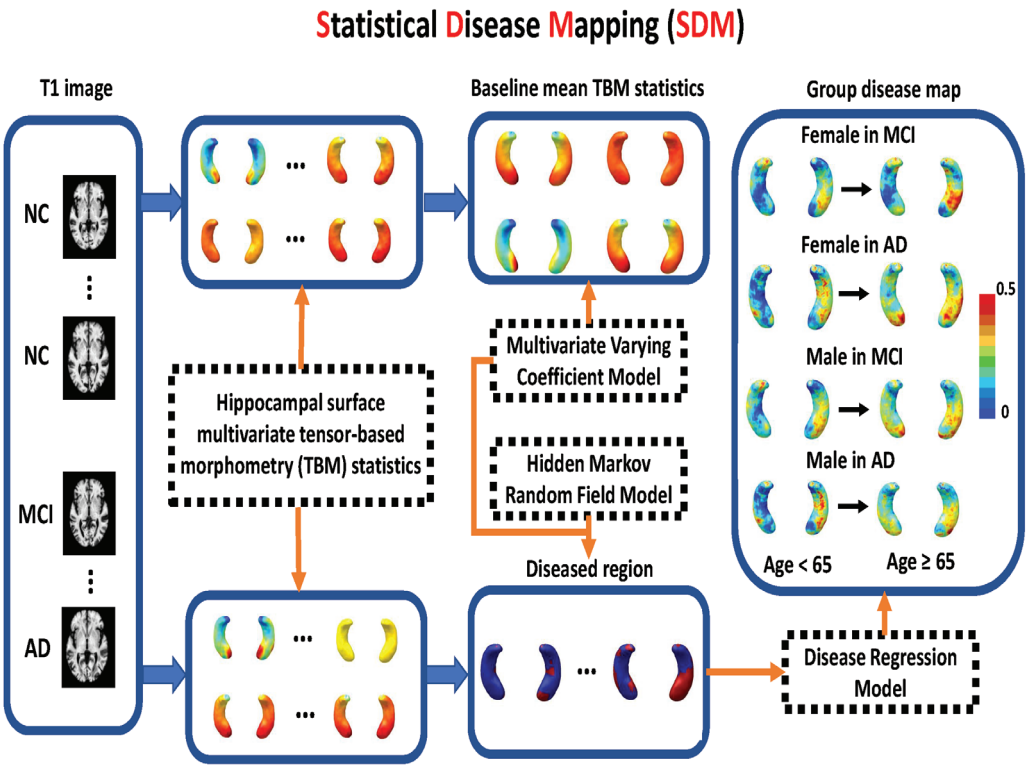

FIGURE 1: Path diagram of statistical disease mapping (SDM).

can effectively capture the spatial smoothness and correlation within the imaging signals and model the heterogeneity among multiple imaging features.

- Besides the subject-specific diseased region detection, our SDM can also build the disease map at the group level. Specifically, an SZIPM is introduced to extract shared diseased region patterns from the detected individual diseased regions within specific subgroups. The SZIPM can simultaneously handle the "nondiseased-region" (e.g., image boundary) via a generalized zero-inflated model and "potential-diseased-region" via a Poisson regression model. Different from the voxel-based morphometry, the group-level disease mapping in our SDM is not sensitive to the imaging heterogeneity using the detected individual diseased regions instead of population-level hypothesis-testing tools.
- All unknown parameters in SDM are estimated by integrating the local polynomial kernel (LPK) smoothing technique (Fan & Gijbels, 1996), a pseudo-likelihood method (Qian & Titterington, 1991), and the Expectation Maximization (EM) algorithm (Dempster, Laird & Rubin, 1977). Besides the estimation procedure, the statistical inference procedure is also considered in SDM. Specifically, both local and global hypothesis-testing problems are investigated, and the resampling method (i.e., wild bootstrap) is adopted to construct the empirical null distribution of test statistics. Furthermore, both simulation studies and real data analysis reveal that SDM can efficiently delineate imaging heterogeneity at both global and local scales.
- Our SDM has several potential applications in neuroimaging data analysis, for example, the abnormal brain region detection for different brain-related diseases, including Alzheimer's disease (Ota et al., 2004), schizophrenia (Chen et al., 2020), traumatic brain injury (Shaker et al., 2017), autism (Salmond et al., 2003), and others. Furthermore, the Python package

for our SDM is freely available online (https://github.com/BIG-S2). In particular, our SDM package can handle three types of functional phenotypes, including curves, surfaces, and volumes.

2. METHODS

2.1. Data Structure

Suppose that we observe imaging data at a common template and clinical variables from n unrelated subjects. Let $\mathcal{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m\}$ be the set of m points from a compact set in $\mathbb{R}^d$, which could be grids on a curve ($d = 2$), vertices on a surface ($d = 3$), and voxels within a volume ($d = 3$). The local features at $\boldsymbol{s}_k$ for the ith subject are often represented as a J-dimensional vector, denoted as $\boldsymbol{y}_i(\boldsymbol{s}_k) = (y_{i,1}(\boldsymbol{s}_k), \ldots, y_{i,J}(\boldsymbol{s}_k))^T$, for $i = 1, \ldots, n$. Let $\mathbf{x}_i$ be a $p \times 1$ vector of predictors (e.g., age, gender) and z_i be the diagnosis status. For the diagnosis status in the ADNI study, 0 denotes normal control, 1 denotes mild cognitive impairment (MCI), and 2 denotes Alzheimer's disease (AD). For each subject, it is assumed that $\mathcal{S}$ can be decomposed into the union of normal region $\mathcal{R}_{i0}$ and diseased region $\mathcal{R}_{i1}$, that is,

$$\mathcal{S} = \mathcal{R}_{i0} \cup \mathcal{R}_{i1} \quad \text{and} \quad \mathcal{R}_{i0} \cap \mathcal{R}_{i1} = \emptyset,$$

where both $\mathcal{R}_{i0}$ and $\mathcal{R}_{i1}$ are random for each patient. In fact, due to the imaging heterogeneity, the size, shape, and location of diseased regions are all stochastic terms and can vary across patients. In order to capture the heterogeneity and randomness of diseased regions, a few further assumptions are made on $\mathcal{R}_{i0}$ and $\mathcal{R}_{i1}$: (i) Normal controls are expected to be perfectly healthy and to not have any diseased region, that is, $\mathcal{R}_{i1} = \emptyset$; (ii) $\mathcal{R}_{i1}$ may vary across diseased patients (both MCI and AD); and (iii) the expected local feature in the diseased region is lower than that in the normal region, that is, $E[\boldsymbol{y}_i(\boldsymbol{s}_k)|\boldsymbol{s}_k \in \mathcal{R}_{i1}] < E[\boldsymbol{y}_i(\boldsymbol{s}_k)|\boldsymbol{s}_k \in \mathcal{R}_{i0}]$. These assumptions are implicitly used in existing methods (Huang et al., 2015). The first assumption makes the normal controls a reference for detecting individual diseased regions, the second assumption plays a critical role in establishing spatial correspondence of diseased regions at different locations across patients, and the third assumption is an identifiability constraint that can deal with the label-switching issue in diseased region detection. Moreover, the third assumption is also reasonable for the ADNI hippocampal surface data in this article. Specifically, atrophy of the hippocampus associated with cognitive impairment in AD can be measured by the imaging responses, including radial distance and multivariate TBM statistics (Wang et al., 2011).

2.2. SDM: Statistical Disease Mapping

There are two components of statistical methods included in our SDM: (i) an image-on-scalar regression model integrating MVCM and HMRFM and (ii) a disease regression model, that is, SZIPM. In the first component, the individual diseased regions are modelled as discrete latent variables though HMRFM, and given those latent variables, the relationship between imaging signals and covariates of interest is investigated via the functional data analysis tools in MVCM. In the second component, the group-level disease map can be extracted from the individual diseased regions detected in the first component through the proposed SZIPM, where the probability that each pixel belongs to the diseased region is modelled for a certain patient group of interest. Next, the two components of our SDM will be discussed in detail.

2.2.1. Component 1: MVCM

The MVCM characterizes the spatial smoothness and correlation of local feature maps given latent disease maps. We introduce an unobserved random effect $b_i(\boldsymbol{s}_k) \in L = \{0, 1\}$ to label

$\mathcal{R}_{i0}$ and $\mathcal{R}_{i1}$ at each voxel $\boldsymbol{s}_k$ for the ith subject. Thus, $\boldsymbol{b}_i$s are label configurations of two nonoverlapping regions $\{\mathcal{R}_{i0}, \mathcal{R}_{i1}\}$. Given $\boldsymbol{b}_i$, MVCM can be written as

$$\boldsymbol{y}_i(\boldsymbol{s}_k) = \boldsymbol{B}(\boldsymbol{s}_k)\boldsymbol{x}_i + \bar{\boldsymbol{B}}\boldsymbol{x}_i b_i(\boldsymbol{s}_k) + \boldsymbol{\eta}_i(\boldsymbol{s}_k) + \boldsymbol{\epsilon}_i(\boldsymbol{s}_k), \tag{1}$$

where $\boldsymbol{B}(\boldsymbol{s}_k)$ is a $J \times p$ matrix representing main effects at voxel $\boldsymbol{s}_k$ in normal regions $\mathcal{R}_{i0}$, whereas $\bar{\boldsymbol{B}}$ is a $J \times p$ matrix of coefficients to characterize possible additional effects in the diseased region $\mathcal{R}_{i1}$. According to the definition of $b_i(\boldsymbol{s}_k)$, $\bar{\boldsymbol{B}}\boldsymbol{x}_i b_i(\boldsymbol{s}_k)$ equals zero across all voxels in normal regions. For patients, voxels in different diseased regions may have different local feature variations. Moreover, $\boldsymbol{\eta}_i(\boldsymbol{s}_k) = (\eta_{i1}(\boldsymbol{s}_k), \ldots, \eta_{iJ}(\boldsymbol{s}_k))^T$ characterizes both subject- and location-specific spatial variability, and $\boldsymbol{\epsilon}_i(\boldsymbol{s}_k) = (\epsilon_{i1}(\boldsymbol{s}_k), \ldots, \epsilon_{iJ}(\boldsymbol{s}_k))^T$ are measurement errors. It is also assumed that $\boldsymbol{\eta}_i(\boldsymbol{s}_k)$ and $\boldsymbol{\epsilon}_i(\boldsymbol{s}_k)$ are mutually independent and identical copies of $\mathrm{SP}(\boldsymbol{0}, \boldsymbol{\Sigma}_\eta)$ and $\mathrm{SP}(\boldsymbol{0}, \boldsymbol{\Sigma}_\epsilon)$, respectively, where $\mathrm{SP}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a stochastic process vector with mean function $\boldsymbol{\mu}(\boldsymbol{s})$ and covariance function $\boldsymbol{\Sigma}(\boldsymbol{s}, \boldsymbol{s}')$. Moreover, $\boldsymbol{\Sigma}_\epsilon(\boldsymbol{s}, \boldsymbol{s}')$ takes the form of $\boldsymbol{\Omega}_\epsilon(\boldsymbol{s})\mathbf{1}(\boldsymbol{s} = \boldsymbol{s}')$, where $\boldsymbol{\Omega}_\epsilon(\boldsymbol{s})$ is a symmetric matrix of functions of $\boldsymbol{s}$, and $\mathbf{1}(\cdot)$ is the indicator function.

2.2.2. Component 1: HMRFM

The HMRFM characterizes the random effects $\boldsymbol{b}_i = (b_i(\boldsymbol{s}_1), \ldots, b_i(\boldsymbol{s}_m))^T$ as follows. First, it is assumed that $\boldsymbol{b}_i$, $\{\boldsymbol{\eta}_i(\boldsymbol{s}_k)\}_{k=1}^m$ and $\{\boldsymbol{\epsilon}_i(\boldsymbol{s}_k)\}_{k=1}^m$ are mutually independent. Moreover, $\{\boldsymbol{b}_i\}_{i=1}^n$ are assumed to be independent across subjects, and each $\boldsymbol{b}_i$ for patients follows a Potts model (Besag, 1986), whose Gibbs form is given by

$$p(\boldsymbol{b}_i|\tau) = \exp\{-U(\boldsymbol{b}_i)\tau - \log C(\tau)\}, \tag{2}$$

where $U(\boldsymbol{b}_i) = \sum_l \sum_{s_k \sim s_l} (1 - \delta(b_i(\boldsymbol{s}_k) - b_i(\boldsymbol{s}_l)))$ and where τ is introduced to encourage spatial smoothness in homogeneous regions. Here, $\delta(\cdot)$ is the Dirac Delta function where $\delta(a) = 1$ when $a = 0$ and $\delta(a) = 0$ otherwise. Moreover, $C(\tau)$ is the partition function such that $p(\boldsymbol{b}_i|\tau)$ is a probability function. The notation "$\sum_{s_k \sim s_l}$" means that $\boldsymbol{s}_k$ is a neighbour of $\boldsymbol{s}_l$, and each neighbouring pair enters the summation only once. Throughout this article, we only consider the closest 2^d neighbours of each voxel, and the closest 2^d neighbours are determined according to their Euclidian distances from the voxel of interest. For the purpose of illustration, the path diagram of HMRFM is presented in Figure 2.

2.2.3. Component 2: SZIPM

According to the detected diseased region $\{\hat{b}_i(\boldsymbol{s})\}_{i=1}^n$, a disease regression model is further introduced to characterize the group disease map according to the conditional distribution of $q_k = \sum_{i=1}^n \hat{b}_i(\boldsymbol{s}_k)\zeta_i$ given $\{\boldsymbol{s}_k\}_{k=1}^m$, where ζ_i is a binary dummy variable indicating the subgroup of interest, and $n^* = \sum_{i=1}^n \zeta_i$ is the sample size of some subgroup of interest, for example, n^* is the sample size of MCI patients in the ADNI study if $\zeta_i = \mathbf{1}\{z_i = 1\}$. In particular, as some voxels are unlikely to be affected by AD, an SZIPM is considered here: For $1 \le k \le m$,

$$Pr\{q_k = j\} = \begin{cases} \pi_k + (1 - \pi_k)e^{-\lambda_k}, & j = 0, \\ (1 - \pi_k)e^{-\lambda_k}\lambda_k^j / j!, & j > 0. \end{cases} \tag{3}$$

Based on model (3), q_k is assumed to be from the point mass distribution based at zero with probability π_k and Poisson distribution $Poisson(\lambda_k)$ with probability $1 - \pi_k$. Here, λ_k and π_k are modelled by smooth functions at $\boldsymbol{s}_k$ as follows:

$$\log(\lambda_k) = \xi_\lambda(\boldsymbol{s}_k),\ \log(\pi_k/(1 - \pi_k)) = \xi_\pi(\boldsymbol{s}_k),\ 1 \le k \le m, \tag{4}$$

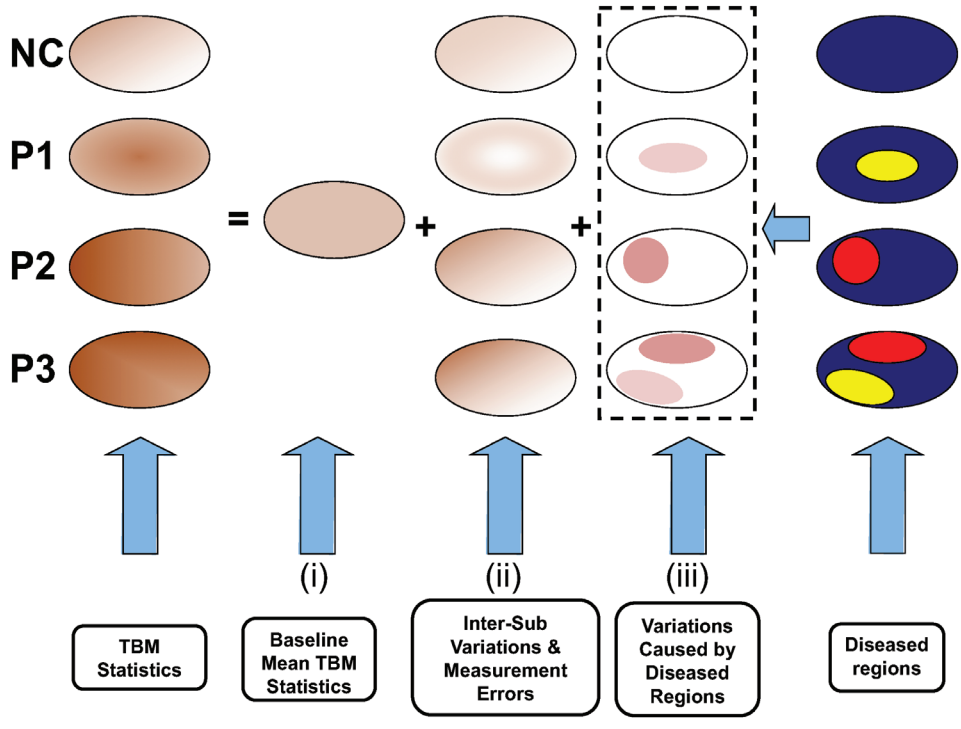

FIGURE 2: Path diagram of HMRFM.

where $\xi_\lambda(\cdot)$ and $\xi_\pi(\cdot)$ are introduced here to model both the spatial smoothness and spatial correlation within the disease map.

Given the estimates $\hat{\xi}_\lambda(\boldsymbol{s})$ and $\hat{\xi}_\pi(\boldsymbol{s})$, the conditional probability that the voxel $\boldsymbol{s}_k$ belongs to the diseased region is calculated as

$$Pr\{\boldsymbol{s}_k \text{ belongs to the diseased region}\} = \frac{e^{\hat{\xi}_\lambda(\boldsymbol{s}_k)}}{n^*(1+e^{\hat{\xi}_\pi(\boldsymbol{s}_k)})}. \tag{5}$$

This estimated disease map (5) is the map indicating the probability that a voxel belongs to the diseased region at the global level. In particular, if we focus on a group with specified gender information, age range, and diagnostic status (e.g., male AD patients with age above 65 years), we can derive the progressive changes in statistical disease mapping across age, which is of great importance in the early detection of AD.

2.3. Estimation Procedure

Our next task is to estimate the random effects $\{\boldsymbol{b}_i\}_{i=1}$ and all unknown parameters consisting of $\tau, \boldsymbol{B}(\boldsymbol{s}), \bar{\boldsymbol{B}}$, covariance functions $\boldsymbol{\Sigma}_\eta(\boldsymbol{s},\boldsymbol{s}'), \boldsymbol{\Omega}_\epsilon(\boldsymbol{s})$, and unknown functions $\xi_\lambda(\boldsymbol{s})$ and $\xi_\pi(\boldsymbol{s})$. Specifically, the whole estimation procedure is divided into three steps, including (i) Step 1 on $\boldsymbol{B}(\boldsymbol{s})$ and $\Omega_\epsilon(\boldsymbol{s})$, denoted as $\boldsymbol{\theta}$; (ii) Step 2 on $\bar{\boldsymbol{B}}$, τ, and $\{\boldsymbol{b}_i\}_{i=1}$; and (iii) Step 3 on parameter estimation in the second component. In Step 1, the maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ can be calculated by using a weighted least-squares (WLS) method (Zhang & Chen, 2007). In Step 2, the WLS estimate of $\bar{\boldsymbol{B}}$ is derived, whereas τ can be estimated by using a pseudo-likelihood method (Geman & Graffigne, 1986) as the MLE of τ is generally difficult to compute due to the normalizing part of the probability function in (2). In addition, the random effects $\{\boldsymbol{b}_i\}_{i=1}^n$ can be

estimated via the maximum a posteriori on Markov random field (MRF-MAP) method (Huang et al., 2015). In Step 3, the EM algorithm is adopted to iteratively update the estimates of unknown functions in (3) and conditional probabilities in (5). Specifically, the local weighted likelihood approach (Fan, Farmen & Gijbels, 1998) and Newton–Raphson methods are employed to derive the estimates of functions $\xi_\lambda(\boldsymbol{s})$ and $\xi_\pi(\boldsymbol{s})$. The details of each key step in the algorithm will be discussed in the following subsections. In addition, the estimation procedure will be summarized in Algorithms 1 and 2 in the Appendix.

2.3.1. Step 1: WLS method on normal control group

We first consider normal controls as, according to our assumption, only normal regions exist in these observed images. Without loss of generality, we assume that the first n_0 subjects are normal controls. In this case, MVCM (1) for normal controls is reduced to

$$\boldsymbol{y}_i(\boldsymbol{s}_k) = \boldsymbol{B}(\boldsymbol{s}_k)\boldsymbol{x}_i + \boldsymbol{\eta}_i(\boldsymbol{s}_k) + \boldsymbol{\epsilon}_i(\boldsymbol{s}_k),\ i = 1, \ldots, n_0. \tag{6}$$

The WLS method based on a local linear kernel smoothing technique can be carried out similar to the procedures in Zhu, Li & Kong (2012). Specifically, let $K(\cdot)$ be the kernel function and $\boldsymbol{H}$ be a bandwidth matrix, which is often a simple diagonal form. We also denote $K_H(\boldsymbol{s}) = |\boldsymbol{H}|^{-1}K(\boldsymbol{H}^{-1}\boldsymbol{s})$, $\boldsymbol{\psi}(\boldsymbol{s}_k - \boldsymbol{s}) = (1, (\boldsymbol{s}_k - \boldsymbol{s})^T\boldsymbol{H}^{-1})^T$ and $\boldsymbol{X}_0 = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n0})$. For each j and fixed $\boldsymbol{H}$, the WLS estimator of $\boldsymbol{B}(\boldsymbol{s})$ can be derived as

$$\hat{\boldsymbol{B}}(\boldsymbol{s}) = \left\{ (\boldsymbol{X}_0\boldsymbol{X}_0^T)^{-1}\boldsymbol{X}_0 \sum_{k=1}^{m} a_k(\boldsymbol{H}, \boldsymbol{s})\boldsymbol{Y}(\boldsymbol{s}_k) \right\}^T, \tag{7}$$

where $\boldsymbol{Y}(\boldsymbol{s}) = (\boldsymbol{y}_1(\boldsymbol{s}), \ldots, \boldsymbol{y}_{n_0}(\boldsymbol{s}))^T$ and

$$a_k(\boldsymbol{H}, \boldsymbol{s}) = (1, \boldsymbol{0}_{1\times d}) \left[\sum_{k=1}^{m} K_H(\boldsymbol{s}_k - \boldsymbol{s})\boldsymbol{\psi}_H(\boldsymbol{s}_k - \boldsymbol{s})^{\otimes 2} \right]^{-1} K_H(\boldsymbol{s}_k - \boldsymbol{s})\boldsymbol{\psi}_H(\boldsymbol{s}_k - \boldsymbol{s}).$$

Based on the estimate in (7), for fixed $\boldsymbol{H}$, the WLS estimate of $\boldsymbol{\eta}_i(\boldsymbol{s})$ is derived as

$$\hat{\boldsymbol{\eta}}_i(\boldsymbol{s}) = \sum_{k=1}^{m} a_k(\boldsymbol{H}, \boldsymbol{s})(\boldsymbol{Y}(\boldsymbol{s}_k) - \hat{\boldsymbol{B}}(\boldsymbol{s})\boldsymbol{x}_i),\ i = 1, \ldots, n_0. \tag{8}$$

Then, their empirical covariance of $\hat{\boldsymbol{\eta}}_i(\boldsymbol{s})$ can be used to estimate $\boldsymbol{\Sigma}_\eta$. Similarly, the covariance matrix $\boldsymbol{\Omega}_\epsilon(\boldsymbol{s})$ can be estimated based on the residuals.

To select the optimal bandwidth in $\hat{\boldsymbol{B}}(\boldsymbol{s})$ and $\{\hat{\boldsymbol{\eta}}_i(\boldsymbol{s})\}_{i=1}^{n_0}$, we use the generalized cross-validation (GCV) score method (Zhang & Chen, 2007). In practice, we standardize all covariates to have a mean of zero and standard deviation of one and all imaging measurements to have a comparable scale. Then, we can choose a common bandwidth for all covariates and imaging phenotypes.

2.3.2. Step 2: Iterative updates of parameter estimation related to patient group

Given the estimates $\hat{\boldsymbol{B}}(\boldsymbol{s})$ and $\hat{\boldsymbol{\Omega}}_\epsilon(\boldsymbol{s})$, we calculate $\hat{\bar{\boldsymbol{B}}}$, $\hat{\boldsymbol{\eta}}_i(\boldsymbol{s})$, and $\{b_i(\boldsymbol{s})\}_{i=1}^n$ for all patients iteratively. Specifically, given the estimates of $\bar{\boldsymbol{B}}$ and $\{b_i^{(r)}(\boldsymbol{s})\}_{i=1}^n$ at iteration r, the WLS estimate of $\boldsymbol{\eta}_i(\boldsymbol{s})$ for each patient, denoted as $\tilde{\boldsymbol{\eta}}_i^{(r+1)}(\boldsymbol{s})$, is given by

$$\sum_{k=1}^{m} a_k(\boldsymbol{H}, \boldsymbol{s})\{\boldsymbol{y}_i(\boldsymbol{s}_k) - \hat{\boldsymbol{B}}(\boldsymbol{s}_k)\boldsymbol{x}_i - \hat{\bar{\boldsymbol{B}}}^{(r)}\boldsymbol{x}_i\tilde{b}_i^{(r)}\}. \tag{9}$$

Subsequently, given the estimate of $\boldsymbol{\eta}_i(\boldsymbol{s})$ at iteration $r+1$, the update of WLS estimate $\tilde{\tilde{\boldsymbol{B}}}^{(r+1)}$ is solved by using

$$\tilde{\tilde{\boldsymbol{B}}}^{(r+1)} = \arg\min_{\bar{B}} \sum_{i=1}^{n}\sum_{k=1}^{m} \tilde{\boldsymbol{\alpha}}_{i,l}^{(r+1)^T} \hat{\boldsymbol{\Omega}}_{\epsilon}^{-1}(\boldsymbol{s}_k)\tilde{\boldsymbol{\alpha}}_{i,l}^{(r+1)}, \tag{10}$$

where $\tilde{\boldsymbol{\alpha}}_{i,l}^{(r+1)} = \boldsymbol{y}_i(\boldsymbol{s}_k) - \hat{\boldsymbol{B}}(\boldsymbol{s}_k)\boldsymbol{x}_i - \tilde{\boldsymbol{\eta}}_i^{(r+1)}(\boldsymbol{s}_k) - \bar{\boldsymbol{B}}\boldsymbol{x}_i\tilde{b}_i^{(r+1)}(s_k)$.

In order to update the estimate of $\{b_i(\boldsymbol{s})\}_{i=1}^n$, the MRF-MAP estimation, an efficient method for many practical applications (e.g., image segmentation (Zhang, Brady & Smith, 2001)), is adopted here. First, given the current estimate $\tilde{\boldsymbol{\theta}}^{(r)}$ and $\tilde{\tau}^{(r)}$ at iteration r, the conditional probability density function of $\boldsymbol{y}_i(\boldsymbol{s}_k)$ given $\boldsymbol{x}_i$ and $b_i(\boldsymbol{s}_k)$ is derived as $\boldsymbol{y}_i(\boldsymbol{s}_k)|\boldsymbol{x}_i, b_i(\boldsymbol{s}_k), \tilde{\boldsymbol{\theta}}^{(r)} \sim \mathcal{N}(\boldsymbol{\mu}_i^{(r)}(\boldsymbol{s}_k), \tilde{\boldsymbol{\Lambda}}^{(r)}(\boldsymbol{s}_k))$, where $\boldsymbol{\mu}_i^{(r)}(\boldsymbol{s}_k) = \hat{\boldsymbol{B}}(\boldsymbol{s}_k)\boldsymbol{x}_i + \tilde{\tilde{\boldsymbol{B}}}^{(r)}\boldsymbol{x}_i b_i(\boldsymbol{s}_k)$ and $\tilde{\boldsymbol{\Lambda}}^{(r)}(\boldsymbol{s}_k) = \tilde{\boldsymbol{\Sigma}}_{\eta}^{(r)}(\boldsymbol{s}_k, \boldsymbol{s}_k) + \hat{\boldsymbol{\Omega}}_{\epsilon}(\boldsymbol{s}_k)$. According to the MAP criterion, the estimate $\tilde{\boldsymbol{b}}_i^{(r+1)}$ is defined as

$$\tilde{\boldsymbol{b}}_i^{(r+1)} = \arg\max_{\boldsymbol{b}_i} \left\{ \prod_{k=1}^{m} f(\boldsymbol{y}_i(\boldsymbol{s}_k)|z_i, b_i(\boldsymbol{s}_k), \tilde{\boldsymbol{\theta}}^{(r)}) p(\boldsymbol{b}_i|\tilde{\tau}^{(r)}) \right\}. \tag{11}$$

To obtain the optimal solution to (11), we adopt the iterated conditional modes (ICM) algorithm (Besag, 1986), which uses a greedy iterative strategy for minimization. Convergence is achieved after only a few iterations.

Finally, in Step 2, we update the estimate of τ given the estimate $\tilde{\boldsymbol{b}}_i^{(r)}$. As τ in model (2) is not the primary parameter of interest, we use an approximate but computationally efficient method based on a pseudo-likelihood function. A key advantage of using the pseudo-likelihood function is its computational simplicity; it does not involve the intractable partition function. The pseudo-likelihood at the iteration r is a simple product of the conditional likelihood

$$PL(\tilde{\boldsymbol{b}}^{(r)}) = \prod_{\{i:\, z_i \neq 0\}} \prod_{s_k \in S - \partial S} PL(\tilde{b}_i^{(r)}(\boldsymbol{s}_k)|\tilde{\boldsymbol{b}}_i^{(r)}), \tag{12}$$

where $\partial \boldsymbol{S}$ denotes the set of points at the boundaries of $\boldsymbol{S}$, and $PL(\tilde{b}_i^{(r)}(\boldsymbol{s}_k)|\tilde{\boldsymbol{b}}_i^{(r)})$ is given by

$$\frac{\exp\left\{\tau \sum_{s_j \in N_k} \delta(\tilde{b}_i^{(r)}(\boldsymbol{s}_k) - \tilde{b}_i^{(r)}(\boldsymbol{s}_j))\right\}}{\exp\left\{\tau \sum_{s_j \in N_k} \delta(\tilde{b}_i^{(r)}(\boldsymbol{s}_j))\right\} + \exp\left\{\tau \sum_{s_j \in N_k} \delta(1 - \tilde{b}_i^{(r)}(\boldsymbol{s}_j))\right\}}.$$

The maximum pseudo-likelihood estimate $\tilde{\tau}^{(r+1)}$ can be obtained by solving

$$\frac{\partial \ln PL(\tilde{\boldsymbol{b}}^{(r)}, \tau)}{\partial \tau} = 0. \tag{13}$$

2.3.3. Step 3: EM algorithm for parameter estimation in SZIPM

Before we conduct the EM algorithm, a set of latent variables $\{\nu_k\}_{k=1}^m$ are defined as below

$$\nu_k = \begin{cases} 0, & q_k \text{ comes from the point mass distribution at 0,} \\ 1, & q_k \text{ comes from Poisson distribution Poisson}(\lambda_k). \end{cases} \tag{14}$$

Then, the complete weighted log-likelihood function is defined as

$$\begin{aligned} \log L(\xi_\pi(\boldsymbol{s}), \xi_\lambda(\boldsymbol{s}) | \{q_k\}_{k=1}^m, \{\nu_k\}_{k=1}^m) &= \log L(\xi_\pi(\boldsymbol{s}) | \{q_k\}_{k=1}^m, \{\nu_k\}_{k=1}^m) \\ &\quad + \log L(\xi_\lambda(\boldsymbol{s}) | \{q_k\}_{k=1}^m, \{\nu_k\}_{k=1}^m), \end{aligned} \tag{15}$$

where

$$\begin{aligned} \log L(\xi_\pi(\boldsymbol{s}) | \{q_k\}_{k=1}^m, \{\nu_k\}_{k=1}^m) &= \sum_{k=1}^m \nu_k (\xi_\pi(\boldsymbol{s}) - \log(1 + e^{\xi_\pi(\boldsymbol{s})})) K_H(\boldsymbol{s}_k - \boldsymbol{s}), \\ \log L(\xi_\lambda(\boldsymbol{s}) | \{q_k\}_{k=1}^m, \{\nu_k\}_{k=1}^m) &= \sum_{k=1}^m (1 - \nu_k)(q_k \xi_\lambda(\boldsymbol{s}) - e^{\xi_\lambda(\boldsymbol{s})}) K_H(\boldsymbol{s}_k - \boldsymbol{s}). \end{aligned}$$

Given the updates, $\tilde{\xi}_\lambda^{(r)}(\boldsymbol{s}_k)$ and $\tilde{\xi}_\pi^{(r)}(\boldsymbol{s}_k)$, at iteration r, in the expectation step (E-step), the conditional expectation of ν_k is derived as

$$E(\nu_k | q_k, \tilde{\xi}_\lambda^{(r)}(\boldsymbol{s}_k), \tilde{\xi}_\pi^{(r)}(\boldsymbol{s}_k)) = \begin{cases} 1/[1 + \exp\{-\tilde{\xi}_\pi^{(r)}(\boldsymbol{s}_k) - e^{\tilde{\xi}_\lambda^{(r)}(\boldsymbol{s}_k)}\}], & q_k = 0, \\ 0, & q_k > 0. \end{cases} \tag{16}$$

Then, we substitute $E(\nu_k | q_k, \tilde{\xi}_\lambda^{(r)}(\boldsymbol{s}_k), \tilde{\xi}_\pi^{(r)}(\boldsymbol{s}_k))$ from the E-step to ν_k in the complete weighted log-likelihood function (16). At iteration $r+1$, the update of $\xi_\lambda(\boldsymbol{s})$ in the maximization step (M-step) is carried out by using the Newton–Raphson method. Specifically, $\tilde{\xi}_\lambda^{(r)}(\boldsymbol{s})_{(t+1)}$ is given by

$$\tilde{\xi}_\lambda^{(r)}(\boldsymbol{s})_{(t)} + \frac{\sum_{k=1}^m (1 - E(\nu_k | q_k, \tilde{\xi}_\lambda^{(r)}(\boldsymbol{s}_k), \tilde{\xi}_\pi^{(r)}(\boldsymbol{s}_k)))(q_k - e^{\tilde{\xi}_\lambda^{(r)}(\boldsymbol{s})_{(t)}}) K_H(\boldsymbol{s}_k - \boldsymbol{s})}{\sum_{k=1}^m (1 - E(\nu_k | q_k, \tilde{\xi}_\lambda^{(r)}(\boldsymbol{s}_k), \tilde{\xi}_\pi^{(r)}(\boldsymbol{s}_k))) e^{\tilde{\xi}_\lambda^{(r)}(\boldsymbol{s})_{(t)}} K_H(\boldsymbol{s}_k - \boldsymbol{s})}.$$

Similarly, we update $\tilde{\xi}_\pi^{(r)}(\boldsymbol{s})_{(t+1)}$ as follows:

$$\tilde{\xi}_\pi^{(r)}(\boldsymbol{s})_{(t)} - \frac{\sum_{k=1}^m E(\nu_k | q_k, \tilde{\xi}_\lambda^{(r)}(\boldsymbol{s}_k), \tilde{\xi}_\pi^{(r)}(\boldsymbol{s}_k))(1 + e^{\tilde{\xi}_\pi^{(r)}(\boldsymbol{s})_{(t)}}) K_H(\boldsymbol{s}_k - \boldsymbol{s})}{\sum_{k=1}^m E(\nu_k | q_k, \tilde{\xi}_\lambda^{(r)}(\boldsymbol{s}_k), \tilde{\xi}_\pi^{(r)}(\boldsymbol{s}_k)) \tilde{\xi}_\pi^{(r)}(\boldsymbol{s})_{(t)} K_H(\boldsymbol{s}_k - \boldsymbol{s})}.$$

The $\tilde{\xi}_\lambda^{(r)}(\boldsymbol{s})_{(t+1)}$ and $\tilde{\xi}_\pi^{(r)}(\boldsymbol{s})_{(t+1)}$ are replaced by $\tilde{\xi}_\lambda^{(r+1)}(\boldsymbol{s})$ and $\tilde{\xi}_\pi^{(r+1)}(\boldsymbol{s})$ after the Newton–Raphson method obtains convergence. The E-step and M-step are repeated until the difference between $\log L(\tilde{\xi}_\pi^{(r+1)}(\boldsymbol{s}), \tilde{\xi}_\lambda^{(r+1)}(\boldsymbol{s}))$ and $\log L(\tilde{\xi}_\pi^{(r)}(\boldsymbol{s}), \tilde{\xi}_\lambda^{(r)}(\boldsymbol{s}))$ is smaller than a desired value, such as 10^{-4}.

2.4. Inference Procedure

After all the parameters are estimated, we carry out formal statistical inference on $\boldsymbol{B}(\boldsymbol{s})$ and $\bar{\boldsymbol{B}}$ in MVCM. In real applications, we are interested in testing (i) whether covariates of interest locally (or globally) are associated with any region for normal controls and (ii) whether there is

any additional disease effect in specific regions. In each voxel, the hypothesis-testing problem (i) can be written in the following general form:

$$H_0(\boldsymbol{s}) : \boldsymbol{C}\text{vec}(\boldsymbol{B}^T(\boldsymbol{s})) = \boldsymbol{0} \quad \text{vs.} \quad H_1(\boldsymbol{s}) : \boldsymbol{C}\text{vec}(\boldsymbol{B}^T(\boldsymbol{s})) \neq \boldsymbol{0}, \tag{17}$$

where $\boldsymbol{C}$ is a $J \times Jp$ matrix with rank J, and $vec(\cdot)$ is the vectorization operator. A sequence of Wald tests can be used here. The test statistic $T_B(\boldsymbol{s})$ for (20) can be written as

$$(\boldsymbol{C}\text{vec}(\hat{\boldsymbol{B}}^T(\boldsymbol{s})))^T[\boldsymbol{C}(\hat{\boldsymbol{\Sigma}}_\eta(\boldsymbol{s},\boldsymbol{s}) \otimes (\boldsymbol{X}_0\boldsymbol{X}_0^T)^{-1})\boldsymbol{C}^T]^{-1}\boldsymbol{C}\text{vec}(\hat{\boldsymbol{B}}^T(\boldsymbol{s})), \tag{18}$$

where $\hat{\boldsymbol{\Sigma}}_\eta(\boldsymbol{s},\boldsymbol{s})$ can be obtained by the empirical covariance of $\{\hat{\boldsymbol{\eta}}_i(\boldsymbol{s})\}_{i=1}^{n_0}$. The corresponding P-values can be derived based on the asymptotic properties of the test statistics under H_0. In particular, under the null hypothesis, when the sample size is large enough, $T_B(\boldsymbol{s})$ approximately follows a χ^2 distribution with J degrees of freedom. The false discovery rate (FDR) adjustment method (Yekutieli & Benjamini, 1999) is also employed here to calculate the adjusted P-values corrected for the multiple comparison problems (20).

To investigate the global effect of covariates of interest, a global test statistic T_B is derived from the local test statistics in (21):

$$T_B = \int_S T_B(\boldsymbol{s})d\boldsymbol{s}. \tag{19}$$

The asymptotic distribution of T_B under H_0 is quite complicated, and it is difficult to directly approximate the percentiles of T_B under the null hypothesis (Zhu, Li & Kong, 2012). Instead, the wild bootstrap method is considered here to obtain critical values of T_B. The method consists of the following four steps:

- Fit model (6) under H_0 for all voxels on $\boldsymbol{X}_0$ and $\{\boldsymbol{Y}(\boldsymbol{s}_k)\}_{k=1}^m$, yielding $\hat{\boldsymbol{B}}(\boldsymbol{s}_k)$, $\hat{\boldsymbol{\eta}}_i(\boldsymbol{s}_k)$, $\hat{\boldsymbol{\epsilon}}_i(\boldsymbol{s}_k)$, and the global test statistic T_B;
- Generate random variables $\tau_i^{(l)}$ and $\tau_i^{(l)}(\boldsymbol{s}_k)$ independently from the standard normal distribution $N(0, 1)$ for $k = 1, \ldots, m$, and then construct

$$\boldsymbol{y}_i^{(l)}(\boldsymbol{s}_k) = \hat{\boldsymbol{B}}(\boldsymbol{s}_k)\boldsymbol{x}_i + \tau_i^{(l)}\hat{\boldsymbol{\eta}}_i(\boldsymbol{s}_k) + \tau_i^{(l)}(\boldsymbol{s}_k)\hat{\boldsymbol{\epsilon}}_i(\boldsymbol{s}_k),\ i = 1, \ldots, n_0;$$

- Based on $\boldsymbol{X}_0$ and $\{\boldsymbol{Y}^{(l)}(\boldsymbol{s}_k)\}_{k=1}^m$, recalculate the global test statistic $T_B^{(l)}$;
- Repeat the previous two steps L times to obtain $\{T_B^{(1)}, \ldots, T_B^{(L)}\}$, which yields the empirical P-value as $p = \sum_{l=1}^L \mathbf{1}(T_B^{(l)} > T_B)/L$.

In this article, 1,000 bootstrap replications are generated, which have been shown to be generally sufficient for type-I error control (Zhu, Li & Kong, 2012).

The hypothesis-testing problem (ii) can be written in the following general form:

$$H_0 : \bar{\boldsymbol{C}}\text{vec}(\bar{\boldsymbol{B}}) = \boldsymbol{0} \text{ vs. } H_1 : \bar{\boldsymbol{C}}\text{vec}(\bar{\boldsymbol{B}}) \neq \boldsymbol{0}, \tag{20}$$

where $\bar{\boldsymbol{C}}$ is a $J \times Jp$ matrix with rank J. A Wald test statistic can be written as

$$T_{\bar{B}} = (\bar{\boldsymbol{C}}\text{vec}(\hat{\bar{\boldsymbol{B}}}))^T[\bar{\boldsymbol{C}}\text{Var}[\text{vec}(\hat{\bar{\boldsymbol{B}}})]\bar{\boldsymbol{C}}^T]^{-1}\bar{\boldsymbol{C}}\text{vec}(\hat{\bar{\boldsymbol{B}}}), \tag{21}$$

which approximately follows a χ^2 distribution with J degrees of freedom under the null hypothesis when the sample size is large enough. Here, the variance term $\mathrm{Var}[\mathrm{vec}(\hat{\tilde{\boldsymbol{B}}})]$ in $T_{\tilde{B}}$ can be obtained via the wild bootstrap method as described in Huang et al. (2015), which proceeds as follows:

1. Fit the model with the original data and retain the fitted mean values $\hat{\boldsymbol{y}}_i(\boldsymbol{s}_k) = \hat{\boldsymbol{B}}(\boldsymbol{s}_k)\boldsymbol{x}_i + \hat{\tilde{\boldsymbol{B}}}\boldsymbol{x}_i\hat{b}_i(\boldsymbol{s}_k)$, individual values $\hat{\boldsymbol{\eta}}_i(\boldsymbol{s}_k)$, and residuals $\hat{\boldsymbol{e}}_i(\boldsymbol{s}_k)$ for $i = 1, \ldots, n$ and $k = 1, \ldots, m$;
2. Generate random variables $\tau_i^{(l)}$ and $\tau_i^{(l)}(\boldsymbol{s}_k)$ independently from the standard normal distribution $N(0,1)$ for $k = 1, \ldots, m$ and then construct $\boldsymbol{y}_i(\boldsymbol{s}_k)^{(l)} = \hat{\boldsymbol{y}}_i(\boldsymbol{s}_k) + \tau_i^{(l)}\hat{\boldsymbol{\eta}}_i(\boldsymbol{s}_k) + \tau_i^{(l)}(\boldsymbol{s}_k)\hat{\boldsymbol{e}}_i(\boldsymbol{s}_k), i = 1, \ldots, n$;
3. Refit the model using the synthetic response variables $\boldsymbol{y}_i(s_k)^{(l)}$ and retain the estimates $\hat{\tilde{\boldsymbol{B}}}^{(l)}$;
4. Repeat Steps 2 and 3 independently L times to derive $\hat{\tilde{\boldsymbol{B}}}^{(1)}, \ldots, \hat{\tilde{\boldsymbol{B}}}^{(L)}$;
5. The bootstrap covariance matrix of $\mathrm{vec}(\hat{\tilde{\boldsymbol{B}}})$ is estimated by

$$\frac{1}{L-1}\sum_{l=1}^{L}\left(\mathrm{vec}(\hat{\tilde{\boldsymbol{B}}}^{(l)} - \hat{\tilde{\boldsymbol{B}}}^{a})\right)\left(\mathrm{vec}(\hat{\tilde{\boldsymbol{B}}}^{(l)} - \hat{\tilde{\boldsymbol{B}}}^{a})\right)^T, \quad \hat{\tilde{\boldsymbol{B}}}^{a} = L^{-1}\sum_{l=1}^{L}\hat{\tilde{\boldsymbol{B}}}^{(l)}.$$

It has been shown that 100 bootstrap replications ($L = 100$ in this article) are generally sufficient for standard error estimation (Efron & Tibshirani, 1994).

3. RESULTS

In this section, we will examine the finite-sample performance of SDM for diseased region detection through both simulation studies and real data analysis. Specifically, the ADNI hippocampal surface data (Huang et al., 2017) are considered, where the data description and processing procedures are described as below.

3.1. ADNI Data Description

Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging, National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations as a $60 million, 5-year public–private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early ADAD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians in developing new treatments and monitoring their effectiveness, as well as lessening the time and cost of clinical trials. The principal investigator of this initiative is Michael W. Weiner, MD, at the VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the United States and Canada. The goal was to recruit 800 subjects, but the initial study (ADNI-1) has been followed by ADNI-GO and ADNI-2. To date, these three protocols have recruited over 1,500 adults, aged 55–90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow-up duration of each group is specified in the protocols for ADNI-1, ADNI-2,

and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

3.2. Data Processing

In this data analysis, we included 798 MRI scans from healthy controls and individuals with AD or MCI (186 AD, 388 MCI, and 224 healthy controls) from ADNI-1. The scans (from 336 men and 462 women, aged 75.42 ± 6.83 years), which were performed on a variety of 1.5 Tesla MRI scanners with protocols individualized for each scanner, include standard T1-weighted images obtained using volumetric three-dimensional sagittal MPRAGE or equivalent protocols with varying resolutions. The typical protocol includes: repetition time = 2,400 ms, inversion time = 1,000 ms, flip angle = 8^o, and field of view = 24 cm, with a $256 \times 256 \times 170$ acquisition matrix in the $x-$, $y-$, and $z-$dimensions, which yields a voxel size of $1.25 \times 1.26 \times 1.2$ mm^3. We processed the MRI data by using standard steps, including anterior commissure and posterior commissure correction, skull-stripping, cerebellum removing, intensity inhomogeneity correction, segmentation, and registration. Subsequently, we carried out automatic regional labeling by labeling the template and by transferring the labels following the deformable registration of subject images. After labeling 93 regions of interests (ROIs), we were able to compute volumes for each ROI for each subject.

We adopted a hippocampal subregional analysis package based on surface fluid registration (Shi et al., 2013) that uses isothermal coordinates and fluid registration to generate one-to-one hippocampal surface registration for computing the surface statistics. Then, we computed the various surface statistics on the registered surface, such as multivariate statistics, which retain the full tensor information of the deformation Jacobian matrix, together with the radial distance, which retains information on the deformation along the surface normal direction. More details can be found in Wang et al. (2011).

3.3. Simulation Studies

We generated the data from the following model:

$$\begin{aligned} \boldsymbol{y}_i(\boldsymbol{s}) \;=\; & \boldsymbol{B}_0(\boldsymbol{s}) + \boldsymbol{B}_1(\boldsymbol{s})x_{i1} + \boldsymbol{B}_2(\boldsymbol{s})x_{i2} \\ & + (\bar{\boldsymbol{B}}_0(\boldsymbol{s}) + \bar{\boldsymbol{B}}_1(\boldsymbol{s})x_{i1} + \bar{\boldsymbol{B}}_2(\boldsymbol{s})x_{i2})b_i(\boldsymbol{s}) + \boldsymbol{\eta}_i(\boldsymbol{s}) + \boldsymbol{\epsilon}_i(\boldsymbol{s}), \end{aligned} \tag{22}$$

where the image responses include both (i) the determinant of the deformation Jacobian matrix and (ii) the radial distance at each vertex from the left hippocampal surface, while the predictors include gender (x_{i1}) and normalized age (x_{i2}). In order to mimic the hippocampal surface data, the true values of parameters in (26) were learned from the real data itself. Specifically, we first fitted the MVCM below to the left hippocampal surface data of all the normal controls in ADNI-1,

$$\boldsymbol{y}_i(\boldsymbol{s}) = \boldsymbol{B}_0(\boldsymbol{s}) + \boldsymbol{B}_1(\boldsymbol{s})x_{i1} + \boldsymbol{B}_2(\boldsymbol{s})x_{i2} + \boldsymbol{\eta}_i(\boldsymbol{s}) + \boldsymbol{\epsilon}_i(\boldsymbol{s}). \tag{23}$$

Then, we used the obtained parameter estimators of coefficient functions ($\boldsymbol{B}_j(\boldsymbol{s}), j = 0, 1, 2$) and covariance functions ($\boldsymbol{\Sigma}_\eta(\boldsymbol{s}, \boldsymbol{s}')$) and $\boldsymbol{\Sigma}_\epsilon(\boldsymbol{s}, \boldsymbol{s}')$)) as their true values in (26). Next, the covariates x_{i1} and x_{i2} in (26) were generated according to their data types, that is, x_{i1} was generated from a Bernoulli distribution with parameter $p = 0.5$, while x_{i2} was generated from a uniform distribution $U(0, 1)$. For the stochastic terms in (26), the individual function $\boldsymbol{\eta}_i(\boldsymbol{s})$ and measurement error function $\boldsymbol{\epsilon}_i(\boldsymbol{s})$ were generated from the Gaussian process with mean function $\mathbf{0}$ and covariance function $\boldsymbol{\Sigma}_\eta(\boldsymbol{s}, \boldsymbol{s}')$ and $\boldsymbol{\Sigma}_\epsilon(\boldsymbol{s}, \boldsymbol{s}')$, respectively. Moreover, for the additional

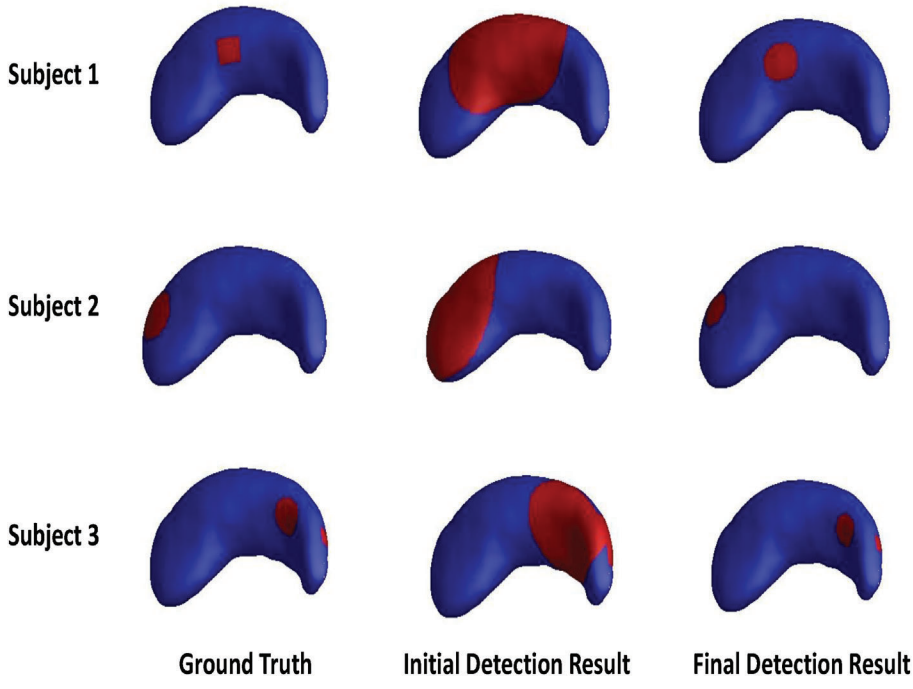

FIGURE 3: Predefined diseased regions, initial detection result (K-means clustering), and final detection result (SDM) for three randomly selected subjects.

effects in diseased regions in (26), the elements in matrix $\bar{\boldsymbol{B}}$ were generated from a uniform distribution $U(-0.5,-0.1)$. For the simulated patterns in diseased regions, to mimic the AD disease region pattern on hippocampal surfaces, the predetermined diseased regions located in the subregions are more likely to be affected by AD (e.g., CA1 subregion, see Frisoni et al. (2008)). Specifically, the diseased patients were divided into three subgroups in terms of their diseased region patterns: For each patient in subgroup 1, one diseased region with a square shape was designated in the CA1 subregion, and the corresponding side length was generated from $U(3,6)$; for each patient in subgroup 2, one diseased region with a circle shape was designated in the tail subregion, and the corresponding diameter was generated from $U(3,6)$; for each patient in subgroup 3, two diseased regions with a circle shape were designated in the head subregion, and the corresponding diameters were independently generated from $U(3,6)$. For all the three groups, the number of vertices within each diseased region is approximately between 250 and 900. Thus, the number, location, shape, and size of diseased regions varied across different subgroups. In column 1 of Figure 3, the predefined diseased regions for subjects randomly selected from three different subgroups were presented. For the patients within the same subgroup, the center of each diseased region was randomly shifted with a unit vertex size at each direction. Finally, we generated 60 diseased patients and 60 normal controls, and the corresponding signal-to-noise ratio (SNR) is 12.235. Compared to the SNR for the left hippocampal surface data, (13.541), it means that our simulated data are close to the real data in terms of SNR.

We applied SDM to detect the diseased regions in each subject, where the K-means clustering method was adopted to obtain the initial detection results. Specifically, for each subject, the observed features were resized into an $m \times J$ matrix. Then, the classical K-means was applied to this matrix with the number of clusters prespecified as 2. The K-means clustering results for the three selected subjects were shown in column 2 from Figure 3. The final detection

results for the subjects selected are presented in column 3 of Figure 3. For all the selected subjects, the diseased regions can be successfully detected with few misclassifications in the results.

Apart from the K-means method, we also include two other approaches for comparison: (i) vertex-wise DSM that adopts linear regression model instead of MVCM and (ii) hidden Markov model (HMM). For the first approach, all the coefficients in the first components of SDM were estimated at each vertex, while the HMRFM was still used to model the individual diseased regions. For the second approach, a linear regression model was considered at each voxel for all normal subjects as follows:

$$\boldsymbol{y}_i(\boldsymbol{s}_k) = \boldsymbol{B}(\boldsymbol{s}_k)\boldsymbol{x}_i + \boldsymbol{\epsilon}_i(\boldsymbol{s}_k),\ k = 1, \ldots, m, \tag{24}$$

where errors $\boldsymbol{\epsilon}_i(\boldsymbol{s}_k)$s' were assumed to be independent across subjects and voxels with mean $\mathbf{0}$ and finite homogeneous variance. After fitting model (28), voxel-wised residual terms for each diseased patient were calculated via

$$\tilde{\boldsymbol{\epsilon}}_i(\boldsymbol{s}_k) = \boldsymbol{y}_i(\boldsymbol{s}_k) - \tilde{\boldsymbol{B}}(\boldsymbol{s}_k)\boldsymbol{x}_i,\ i = n_0 + 1, \ldots, n. \tag{25}$$

Furthermore, for diseased patients, residual terms were modelled by using a Gaussian HMM as follows:

$$\tilde{\boldsymbol{\epsilon}}_i(\boldsymbol{s}_k) \sim N(\boldsymbol{\mu} b_i(s_k), \sigma^2 \boldsymbol{I}), \tag{26}$$

where $\boldsymbol{b}_i$ follow the Potts model in (2) with parameter τ. Here, $\boldsymbol{\mu}$, τ, and b_i were estimated, respectively, via the EM algorithm, the pseudo-likelihood method, and the MAP estimate method described in Section 2. To compare the average detection performance of K-means, HMM, vertex-wise SDM, and SDM, the adjusted Rand index (ARI) was calculated for each subject. For the adjusted Rand index, a higher value indicates a more accurate detection result. The detailed definition of ARI is included in the Appendix. Simulation results based on all the subjects are presented in Figure 4. In terms of both mean value and variability of ARI, SDM outperforms the other three methods, which indicates that it performs very well in the diseased region detection. In particular, due to the functional data analysis tools in our SDM, the estimation variability was reduced compared to that in vertex-wise SDM, leading to higher performance in terms of the average ARI.

Next, we investigated the Type I error rate and power of test statistics (22) and (24). First, for the covariate "age" (x_{i2}), we set the corresponding coefficient function to $\boldsymbol{B}_2^*(\boldsymbol{s}) = a\boldsymbol{B}_2(\boldsymbol{s}_k)$, where $\boldsymbol{B}_2(\boldsymbol{s}_k)$ is the obtained parameter estimators that were used as the true values in previous simulation setting, and a is a scalar to be specified later. Here, we were interested in the following hypothesis-testing problem

$$H_0 :\ \mathrm{vec}(\boldsymbol{B}_2^*(\boldsymbol{s})) = \mathbf{0}\ \forall\ \boldsymbol{s}, \text{vs. } H_1 :\ \mathrm{vec}(\boldsymbol{B}_2^*(\boldsymbol{s})) \neq \mathbf{0} \text{ for at lease one } \boldsymbol{s}. \tag{27}$$

We set $a = 0$ to assess the Type I error rate for T_B in (22) and set $a = 0.2, 0.4, 0.6, 0.8$, and 1 to examine the power of T_B. Second, for the diseased region-related coefficient, we set it as $\bar{\boldsymbol{B}}^* = c\bar{\boldsymbol{B}}^c$, where the elements in $\bar{\boldsymbol{B}}^c$ were generated from the uniform distribution $U(-0.5, -0.1)$, and c is a scalar to be specified later. Here, another hypothesis-testing problem was considered as follows:

$$H_0 :\ \mathrm{vec}(\bar{\boldsymbol{B}}^*) = \mathbf{0} \text{ vs. } H_1 :\ \mathrm{vec}(\bar{\boldsymbol{B}}^*) \neq \mathbf{0}. \tag{28}$$

We set $c = 0$ to assess the Type I error rate for $T_{\bar{B}}$ in (24) and set $c = 0.2, 0.4, 0.6, 0.8$, and 1 to examine the power of $T_{\bar{B}}$. For each situation, the significance levels were set to $\alpha = 0.05$ and

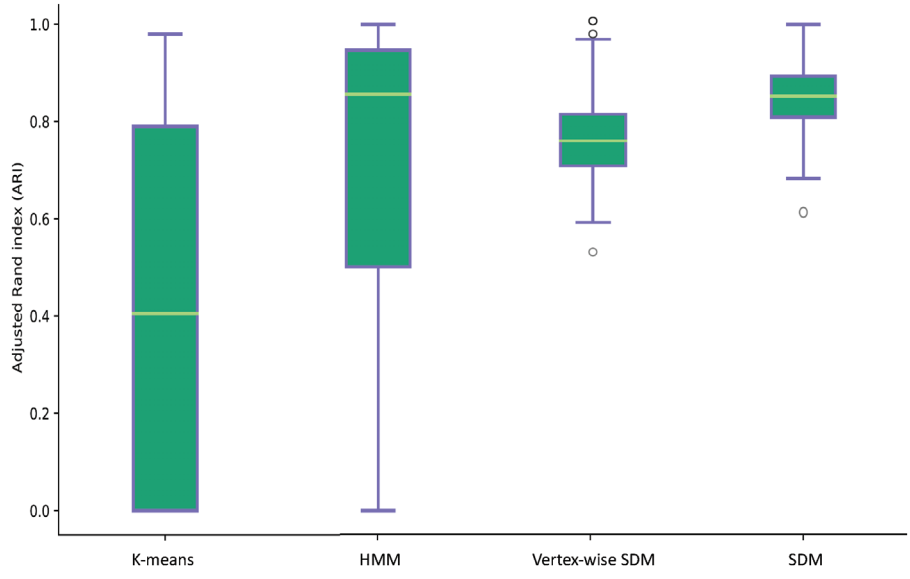

FIGURE 4: Adjusted Rand index (ARI) of detection results for four different methods: K-means clustering method, HMM, vertex-wise SDM, and SDM.

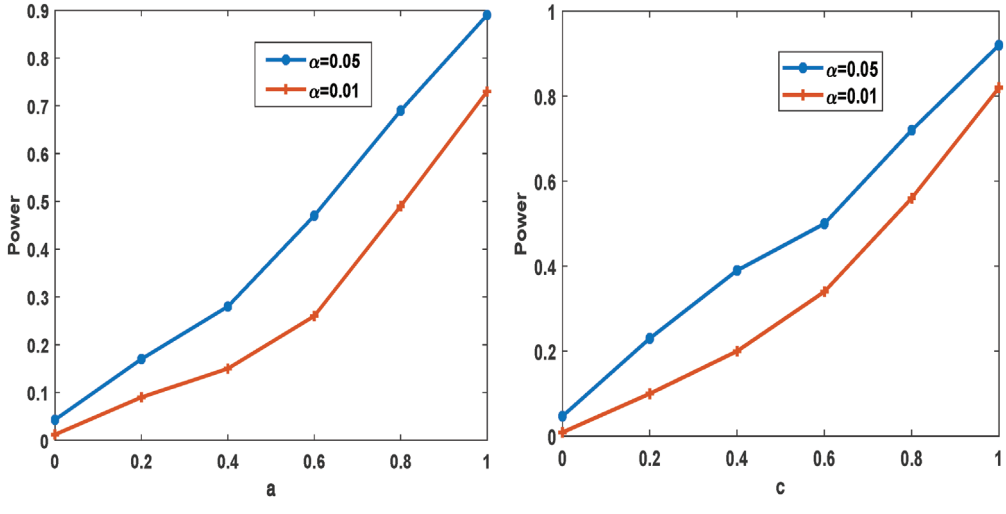

FIGURE 5: Simulation results: power curve for testing problem (31) with different choices of c in $\boldsymbol{B}_2^*(\boldsymbol{s})$ (left panel) and testing problem (32) with different choices of c in $\bar{\boldsymbol{B}}^*$ (right panel).

$\alpha = 0.01$, and 500 bootstrap replications were generated to construct the empirical distribution of T_B under H_0.

Figure 5 depicts the power curves for both test statistics. It can be seen that the Type I error rates for T_B and $T_{\bar{B}}$ are accurate at both significance levels ($\alpha = 0.01$ or 0.05), while the testing power is reasonable for both test statistics when the corresponding effect size becomes larger.

Finally, besides the individual diseased region detection, our SDM was also applied to construct the disease map at the whole population level, which is presented in Figure 6 (a). For the purpose of comparison, two alternative approaches are considered here: (i) vertex-wise SDM

 DOI: 10.1002/cjs.11595

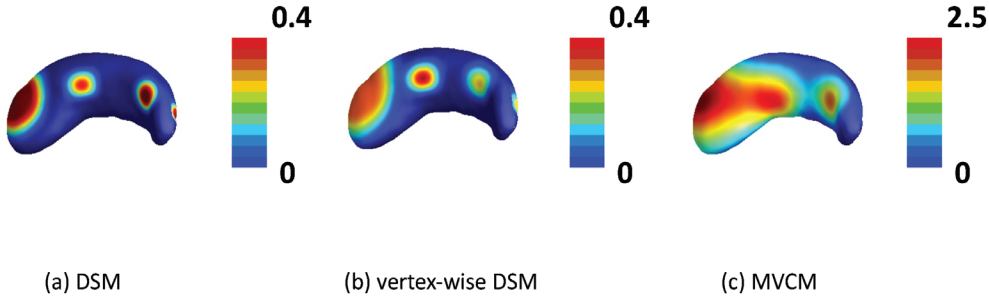

FIGURE 6: Group-level disease map for three different methods: (a) disease probability from SDM, (b) disease probability from vertex-wise SDM, and (c) FDR-adjusted $-\log_{10}(p)$ value map for hypothesis testing (34) based on MVCM.

and (ii) local hypothesis testing in MVCM. The group-level disease map based on vertex-wise SDM was calculated and presented in Figure 6 (b). According to the disease maps, both SDM and vertex-wise SDM successfully detected the disease pattern in each of the three groups. However, our SDM shows better performance as the disease probability in the predefined diseased pattern is higher than that based on vertex-wise SDM and is very close to the empirical probability, that is, 0.33. For the second approach, we introduced a dummy variable x_{i3} indicating whether the ith subject is a patient or not. Then, the varying coefficient model is written as

$$\boldsymbol{y}_i(\boldsymbol{s}) = \boldsymbol{B}_0(\boldsymbol{s}) + \boldsymbol{B}_1(\boldsymbol{s})x_{i1} + \boldsymbol{B}_2(\boldsymbol{s})x_{i2} + \boldsymbol{B}_3(\boldsymbol{s})x_{i3} + \boldsymbol{\eta}_i(\boldsymbol{s}) + \boldsymbol{\epsilon}_i(\boldsymbol{s}), \tag{29}$$

where $\boldsymbol{\eta}_i(\boldsymbol{s})$ and $\boldsymbol{\epsilon}_i(\boldsymbol{s})$ have the same assumptions as those in model (1). In order to derive the group-level disease map, the following local hypothesis-testing problem is considered at each $\boldsymbol{s}_k, k = 1, \ldots, m$,

$$H_0 : \ \text{vec}(\boldsymbol{B}_3(\boldsymbol{s})) = \boldsymbol{0}\ , \text{vs. } H_1 : \ \text{vec}(\boldsymbol{B}_3(\boldsymbol{s})) \neq \boldsymbol{0}. \tag{30}$$

The FDR-adjusted $-\log_{10}(p)$ values across all vertices based on (33) are shown in Figure 6 (c). Compared to the previous two approaches, the group-level disease map based on the local hypothesis testing fails to accurately detect all the predefined group disease patterns. Therefore, our SDM also performs very well in building the group-level disease map.

3.4. ADNI Hippocampal Surface Data Analysis

The proposed SDM was used to detect the diseased regions in this dataset. In SDM, we set $\boldsymbol{x}$ =(1, Gender, Age). Four local features were considered, including the logged radial distance and multivariate TBM statistics (details in Shi et al. (2013)). The local feature maps of logged radial distance and multivariate TBM statistics for three randomly selected subjects (one normal control, one MCI, and one AD) are presented in Figure 7.

After extracting the local feature maps, we applied MVCM across all normal control subjects, and the estimates of coefficient functions associated with the covariates gender and age are presented in Figure 8. In order to test whether the covariates of interest (gender and age) locally or globally affect the regions for normal controls, the local and global Wald test statistics were calculated. The adjusted $-\log_{10}(p)$ values across all vertices are shown in Figure 9, while the global P-value related to each covariate is listed as well. To calculate the distribution percentile of the global test statistic under the null hypothesis, the wild bootstrap method was adopted to construct the empirical distribution of test statistic under H_0. It indicates that, compared to the gender effect, the age effect is more significant in terms of the local features across the left and right hippocampal surfaces.

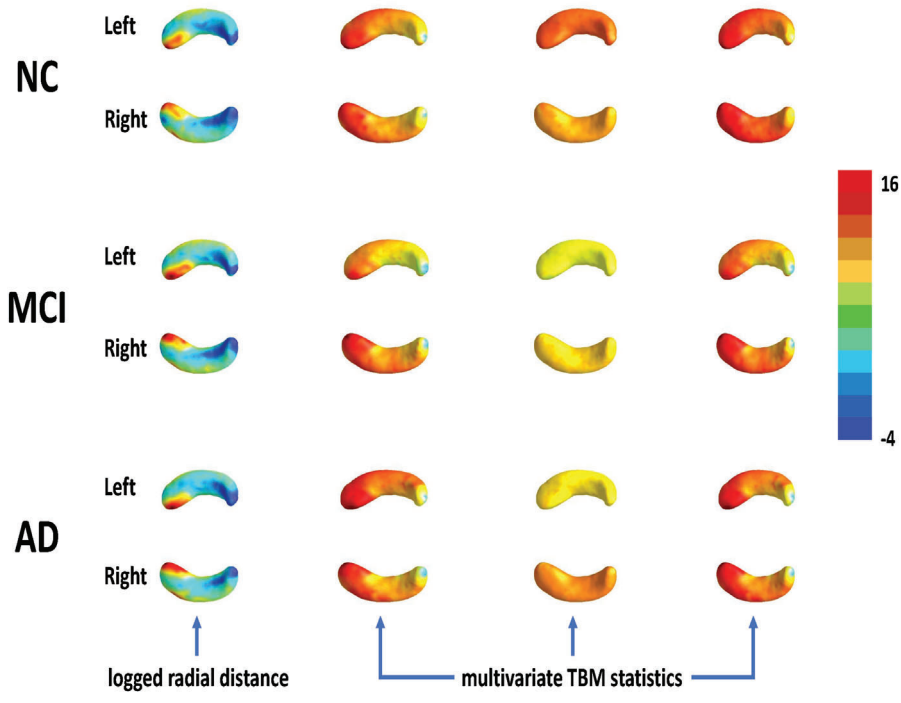

FIGURE 7: ADNI hippocampal surface data analysis: local feature maps of logged radial distance and multivariate TBM statistics for three randomly selected subjects (one normal control, one MCI, and one AD).

Then, we applied SDM to the local feature maps from both MCI and AD patients. The inference results of $\tilde{\boldsymbol{B}}$ were presented in Table 1. For the diseased regions, the detection results of one randomly selected MCI patient and one randomly selected one AD patient were plotted in Figure 10, in which the red area indicates the detected diseased region. The disease maps for some subgroups were also estimated. Note that AD is considered to be younger-onset Alzheimer's if it affects a person under 65 years of age (Batsch & Mittelman, 2015). In this article, it is critical to investigate the difference between younger-onset AD patients and other AD patients in terms of their group-level disease map. Actually, eight different subgroups were considered here: (i) male MCI with age below 65 years, (ii) male MCI with age above 65 years, (iii) male AD with age below 65 years, (iv) male AD with age above 65 years, (v) female MCI with age below 65 years, (vi) female MCI with age above 65 years, (vii) female AD with age below 65 years, and (viii) female AD with age above 65 years.

Figure 11 presents the estimated disease maps. It can be found that: (a) the disease patterns are different between MCI and AD groups; (b) the disease patterns are different between younger-onset Alzheimer's and ADs; and (c) the disease patterns are different between left and right hippocampal surfaces, indicating the hippocampal asymmetry (Shi et al., 2009).

For comparison, here, we also calculated the disease map based on the hypothesis-testing problem (34) for model (33). Both the MCI group and the AD group were compared with the normal controls, and the FDR-adjusted $-\log_{10}(p)$ values across all vertices are presented in Figure 12. In order to further compare those two approaches, we investigated specific subregions corresponding to the disease mapping based on the cytoarchitectonic subregions (Figure 13) mapped on blank MR-based models at 3T of the hippocampal formation (Duvernoy, 2005; Frisoni et al., 2008). Compared to the results based on the hypothesis-testing method, all the

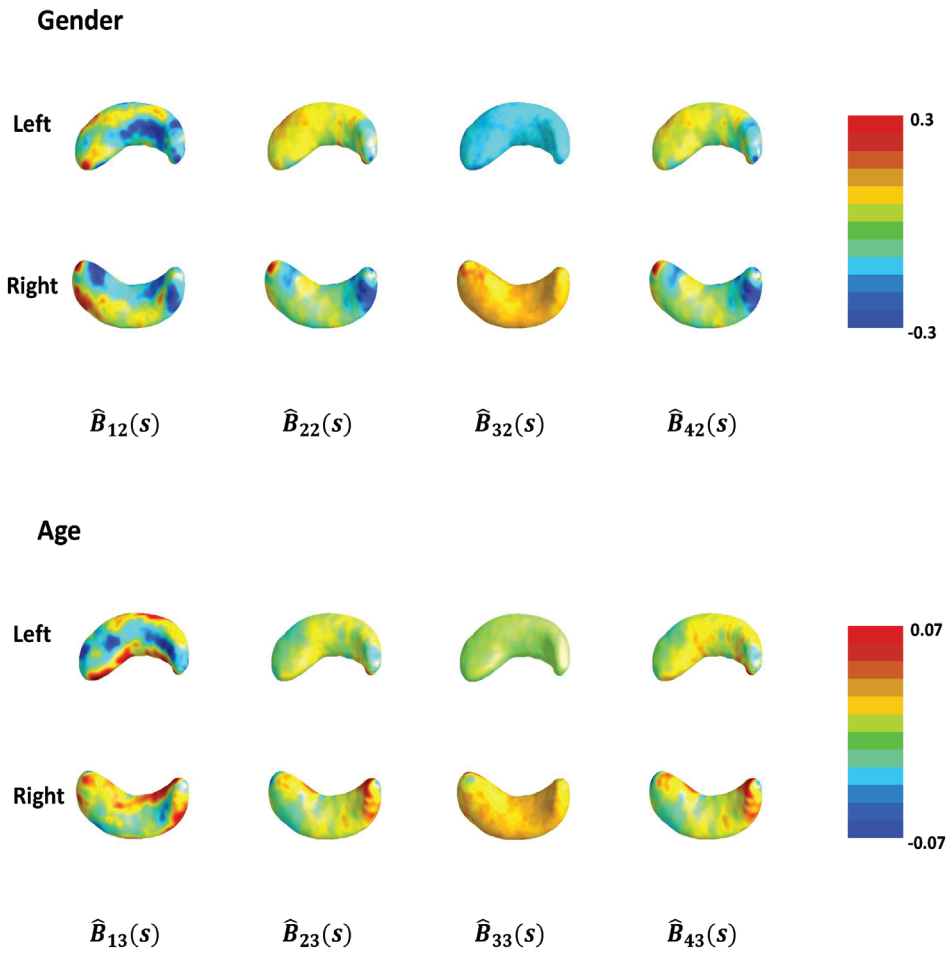

FIGURE 8: ADNI hippocampal surface data analysis: parameter estimation for covariates of interest (top: gender; bottom: age).

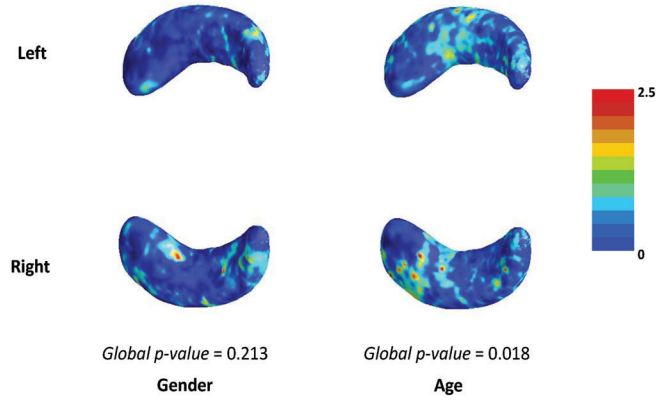

FIGURE 9: ADNI hippocampal surface data analysis: local adjusted $-\log_{10}(p)$ values associated to covariates of interest (column 1: gender; column 2: age). The global P-values are listed below the local P-value maps.

TABLE 1: Hypothesis testing results of $\bar{\boldsymbol{B}}$

	Left hippocampal surface			
Covariate	Parameter	Estimation	Standard error	$-\log_{10}(p)$ value
Gender	$\bar{B}_{12}$	−0.0269	0.0033	1.9324
	$\bar{B}_{22}$	−0.0677	0.0081	1.2257
	$\bar{B}_{32}$	−0.0318	0.0029	0.8901
	$\bar{B}_{42}$	0.0374	0.0027	0.9062
Age	$\bar{B}_{13}$	−0.0788	0.0075	4.3138
	$\bar{B}_{23}$	−0.0874	0.0094	3.9866
	$\bar{B}_{33}$	−0.0264	0.0026	2.7935
	$\bar{B}_{43}$	−0.0336	0.0030	3.5263
	Right hippocampal surface			
Covariate	Parameter	Estimation	Standard error	$-\log_{10}(p)$ value
Gender	$\bar{B}_{12}$	−0.0291	0.0033	1.0035
	$\bar{B}_{22}$	0.0022	0.0005	0.2717
	$\bar{B}_{32}$	−0.0011	0.0009	0.1990
	$\bar{B}_{42}$	0.0044	0.0011	0.3167
Age	$\bar{B}_{13}$	−0.0690	0.0048	3.9711
	$\bar{B}_{23}$	−0.0092	0.0013	0.4323
	$\bar{B}_{33}$	−0.0020	0.0008	0.5292
	$\bar{B}_{43}$	−0.0050	0.0007	0.3308

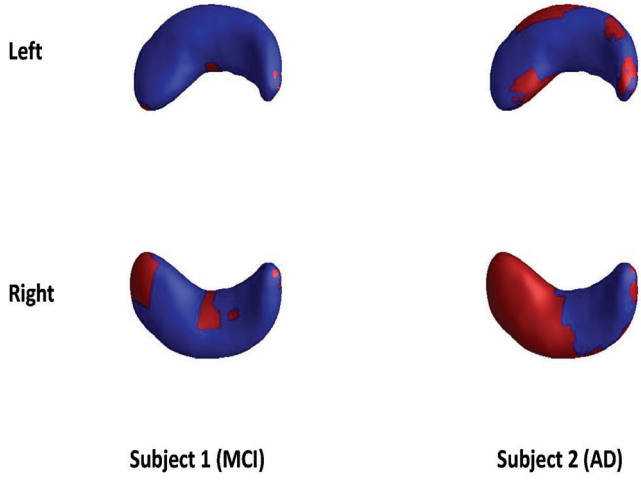

FIGURE 10: ADNI hippocampal surface data analysis: detection results obtained from two randomly selected subjects (column 1: MCI patient; column 2: AD patient).

 DOI: 10.1002/cjs.11595

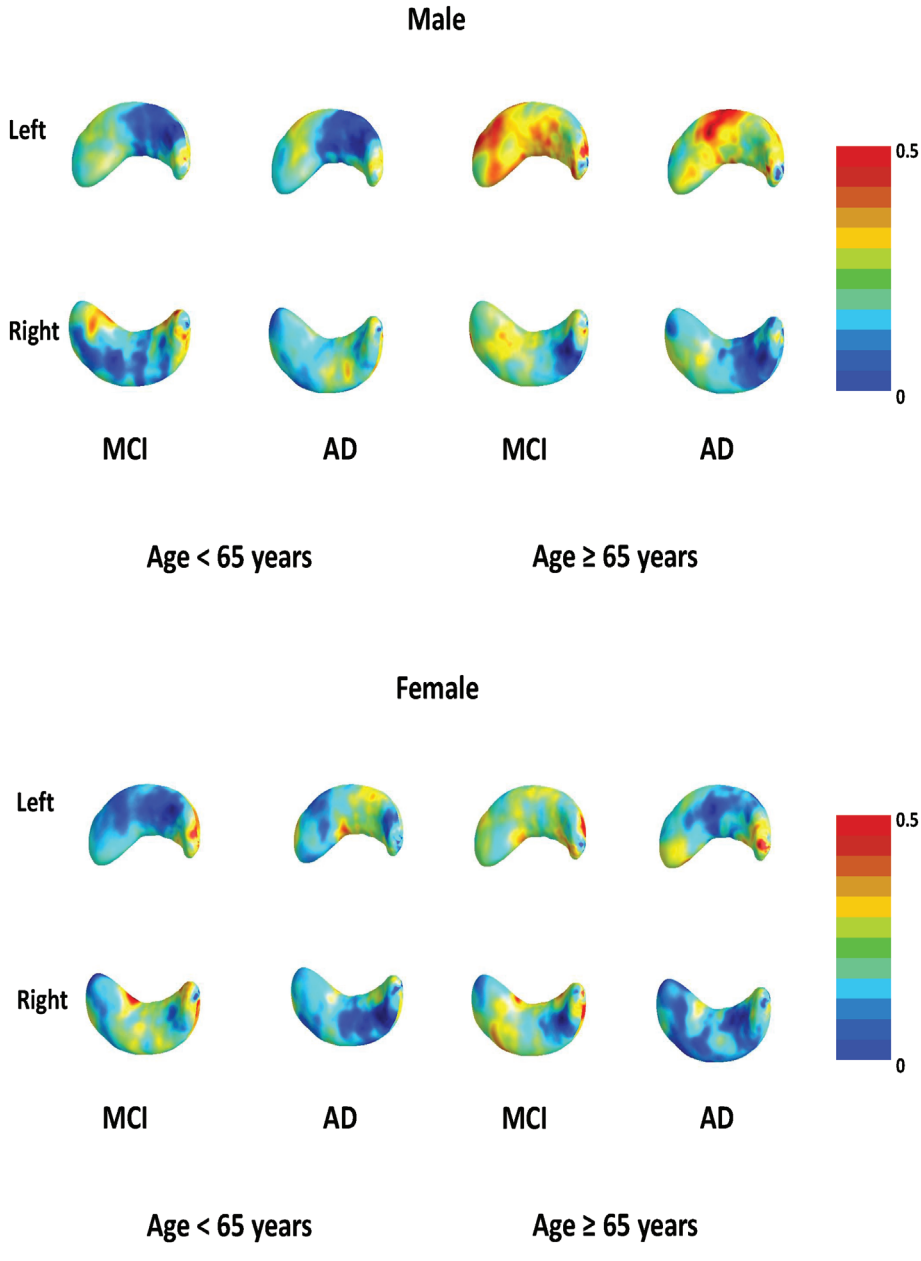

FIGURE 11: ADNI hippocampal surface data analysis: disease map for different subgroups (top: male; bottom: female).

risky subregions detected by SDM are found in the CA1 subfield, some are found on the lateral and medial aspects of the tail (CA1 subfield), and others are found on the dorsolateral aspect of the head (CA1 subfield). It is interesting to note that volumes of similar hippocampal subregions were found to be affected in AD (Frisoni et al., 2008), indicating that the results obtained from SDM agree with those of previous work. Therefore, the detected diseased regions are meaningful and may be treated as potential imaging biomarkers for AD.

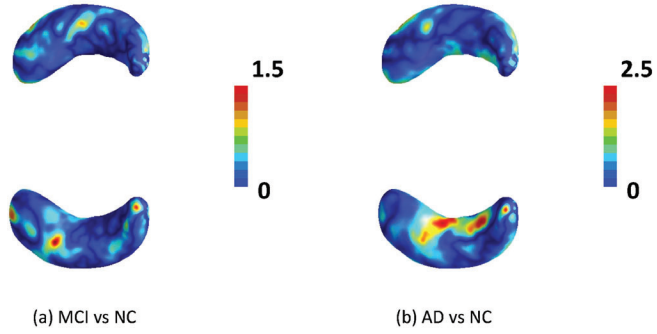

FIGURE 12: ADNI hippocampal surface data analysis: FDR-adjusted $-\log_{10}(p)$ value maps for two group comparisons: (a) MCI versus NC and (b) AD versus NC.

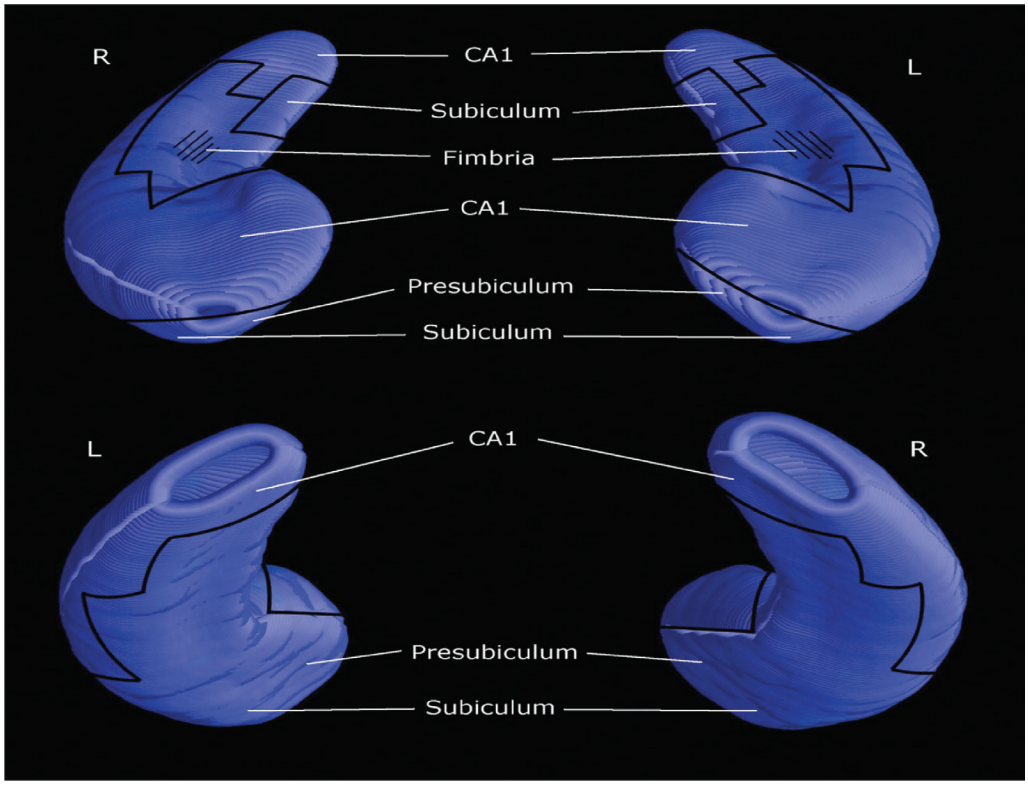

FIGURE 13: ADNI hippocampal surface data analysis: cytoarchitectonic subregions mapped on blank MR-based models at 3T of the hippocampal formation (Duvernoy, 2005; Frisoni et al., 2006; Frisoni et al., 2008).

4. CONCLUSION AND DISCUSSION

In this article, we have developed a novel SDM framework to address the challenges stemming from disease heterogeneity at both the group and individual scales. Our SDM included two components: (i) an image-on-scalar regression model that integrates HMRFM and MVCM and (ii) a disease regression model SZIPM. Specifically, the functional data analysis tools in MVCM can help SDM capture spatial smoothness and correlation in local feature maps across

subjects; the HMRFM can detect individual diseased regions; and the DRM can capture the disease pattern within specific subgroups. Both simulation studies and real data analysis have revealed that SDM can efficiently delineate imaging heterogeneity at both global and local scales.

Some important issues need to be addressed in future research. First, in our SDM, the SZIPM is a subsequent analysis based on the estimation procedure in the first component of SDM. Compared to the vertex-wise analysis, our SDM builds the group-level disease map through the detected subject-specific diseased regions instead of hypothesis testing-based methods, which indicates that SDM is insensitive to the potential imaging heterogeneity. However, this framework has a drawback that the latent variables $\{b_i(\boldsymbol{s})\}_{i=1}^n$ are double modelled in SDM, that is, HMRFM and SZIPM are adopted to model the latent variables in each of the two components in SDM, respectively. However, the relationship between these two models has not been well investigated, which will be one of our future research topics, a few efficient joint analysis frameworks will also be investigated.

Second, in order to detect the subject-specific diseased region detection, an HMRFM, the Potts model, is considered in the first component of SDM. Although some other models such as Ising model (Shu, Nan & Koeppe, 2015) can be applied here instead, the Potts model in this article has a simple Gibbs form and shows great performance in both simulation studies and real data analysis in terms of diseased region detection. However, for a more complicated imaging dataset in the future, this simple structure may not be enough. In that case, some weakly supervised learning method, such as multiple instance learning (Zhou, 2018), will be considered to discover the potential individual disease patterns.

ACKNOWLEDGEMENTS

We are grateful for all the constructive comments from the referees. Hongtu Zhu was partially supported by grants from the National Institutes of Health of the United States (Grant ID: MH086633 and MH092335).

APPENDIX

The estimation procedure for Steps 1 and 2 is summarized below.

Algorithm 1. Estimation procedure for Step 1 and Step 2

Input Data: $\{\boldsymbol{y}_i(\boldsymbol{s}_k)\}_{k=1}^m, \boldsymbol{x}_i, z_i, i = 1, \ldots, n$
Initialize random effects (K-means): $\{\tilde{\boldsymbol{b}}_i^{(0)}\}_{i=1}^n$
WLS method for normal controls:
$\{\widehat{\boldsymbol{B}}(s_k), \widehat{\boldsymbol{\Sigma}}_\eta(\boldsymbol{s}_k), \widehat{\boldsymbol{\Omega}}_\epsilon(\boldsymbol{s}_k), \widehat{\boldsymbol{\eta}}_i(\boldsymbol{s}_k)\}_{k=1}^m, i = 1, \ldots, n_0$
Initialize estimation $\{\tilde{\boldsymbol{\eta}}_i^{(0)}(\boldsymbol{s}_k)\}_{k=1}^m$ **and** $\tilde{\tilde{\boldsymbol{B}}}^{(0)}$ **for patients**
Repeat $(r \geq 0)$
- Update $\{\tilde{\boldsymbol{b}}_i^{(r+1)}\}_{i=n_0+1}^n$ via MRF-MAP method
- Update $\tilde{\tau}^{(r+1)}$ via pseudo-likelihood method
- Update $\tilde{\tilde{\boldsymbol{B}}}^{(r+1)}$ and $\{\tilde{\boldsymbol{\eta}}_i^{(r+1)}(\boldsymbol{s}_k)\}_{k=1}^m$ via WLS method

End repeat
Output: detection results $\{\widehat{b}_i(\boldsymbol{s}_k)\}_{k=1}^m$ for each patient

The estimation procedure for Step 3 is summarized below.

Algorithm 2. Estimation Procedure for Step 3

Input Data: detection results $\{q_k = \sum_{i=1}^{n} \widehat{b}_i(\boldsymbol{s}_k)\delta_i\}_{k=1}^{m}$ for subgroup of interest
Initialize estimation $\{\tilde{\xi}_{\lambda}^{(0)}(\boldsymbol{s}_k)\}_{k=1}^{m}$ **and** $\{\tilde{\xi}_{\pi}^{(0)}(\boldsymbol{s}_k)\}_{k=1}^{m}$
EM algorithm (iteration r)
- E-step: calculate $E(\nu_k|q_k, \tilde{\xi}_{\lambda}^{(r)}(\boldsymbol{s}_k), \tilde{\xi}_{\pi}^{(r)}(\boldsymbol{s}_k)), k = 1, \ldots, m$
- M-step: update $\tilde{\xi}_{\lambda}^{(r+1)}(\boldsymbol{s})$ and $\tilde{\xi}_{\pi}^{(r+1)}(\boldsymbol{s})$ via Newton–Raphson method

End iteration
Output: disease mapping in (5) for subgroup of interest

Adjusted Rand index (ARI)

Given a set of n objects and two groupings or partitions (i.e., one is the detection result, and the other is ground truth) of these elements, namely, $X = \{X_1, X_2\}$ and $Y = \{Y_1, Y_2\}$, the number of objects in common between X_i and Y_j is defined as $n_{ij} = X_i \cap Y_j, i = 1, 2, j = 1, 2$. Then, the original ARI using the permutation model is

$$ARI = \frac{\sum_{i=1}^{2}\sum_{j=1}^{2}\binom{n_{ij}}{2} - \left[\sum_{i=1}^{2}\binom{a_i}{2}\right]\left[\sum_{j=1}^{2}\binom{b_j}{2}\right]/\binom{n}{2}}{\frac{1}{2}\left[\sum_{i=1}^{2}\binom{a_i}{2} + \sum_{j=1}^{2}\binom{b_j}{2}\right] - \left[\sum_{i=1}^{2}\binom{a_i}{2}\right]\left[\sum_{j=1}^{2}\binom{b_j}{2}\right]/\binom{n}{2}},$$

where $a_i = n_{i1} + n_{i2}, i = 1, 2$, and $b_j = n_{1j} + n_{2j}, j = 1, 2$.

BIBLIOGRAPHY

Allen, N. E., Sudlow, C., Peakman, T., Collins, R., & UK Biobank. (2014). UK biobank data: Come and get it. *Science Translational Medicine*, 6(224), 224ed4.

Bashir, U., Siddique, M. M., Mclean, E., Goh, V., & Cook, G. J. (2016). Imaging heterogeneity in lung cancer: Techniques, applications, and challenges. *American Journal of Roentgenology*, 207(3), 534–543.

Batsch, N. L. & Mittelman, M. S. (2015). World Alzheimer report 2012. Overcoming the stigma of dementia. Alzheimer's Disease International (ADI), 5.

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B. Methodological*, 48(3), 259–302.

Brooks, F. J. & Grigsby, P. W. (2013). Quantification of heterogeneity observed in medical images. *BMC Medical Imaging*, 13(1), 7.

Chen, Z., Yan, T., Wang, E., Jiang, H., Tang, Y., Yu, X., Zhang, J., & Liu, C. (2020). Detecting abnormal brain regions in schizophrenia using structural MRI via machine learning. *Computational Intelligence and Neuroscience*, 2020, 1–13.

Davnall, F., Yip, C. S., Ljungqvist, G., Selmi, M., Ng, F., Sanghera, B., Ganeshan, B., Miles, K. A., Cook, G. J., & Goh, V. (2012). Assessment of tumor heterogeneity: An emerging imaging tool for clinical practice? *Insights into Imaging*, 3(6), 573–589.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1–22.

Duvernoy, H. M. (2005). *The Human Hippocampus: Functional Anatomy, Vascularization and Serial Sections with MRI*. Springer Science & Business Media, Berlin Heidelberg.

Efron, B. & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*, Vol. 57. CRC Press, New York.

Fan, J., Farmen, M., & Gijbels, I. (1998). Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(3), 591–608.

Fan, J. & Gijbels, I. (1996). *Local Polynomial Modelling and its Applications: Monographs on Statistics and Applied Probability*, Vol. 66. CRC Press, Boca Raton.

Frisoni, G. B., Ganzola, R., Canu, E., Rüb, U., Pizzini, F. B., Alessandrini, F., Zoccatelli, G., Beltramello, A., Caltagirone, C., & Thompson, P. M. (2008). Mapping local hippocampal changes in Alzheimer's disease and normal ageing with MRI at 3 tesla. *Brain*, 131(12), 3266–3276.

Frisoni, G. B., Sabattoli, F., Lee, A. D., Dutton, R. A., Toga, A. W., & Thompson, P. M. (2006). In vivo neuropathology of the hippocampal formation in ad: A radial mapping MR-based study. *Neuroimage*, 32(1), 104–110.

Geman, S. & Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6), 721–741.

Geman, S. & Graffigne, C. (1986). Markov random field image models and their applications to computer vision. In *Proceedings of the International Congress of Mathematicians*. Vol. 1. American Mathematical Society, Providence, RI p. 2.

Goldstein, M. & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PLoS One*, 11(4), e0152173.

Huang, C., Shan, L., Charles, H. C., Wirth, W., Niethammer, M., & Zhu, H. (2015). Diseased region detection of longitudinal knee magnetic resonance imaging data. *IEEE Transactions on Medical Imaging*, 34(9), 1914–1927.

Huang, C., Thompson, P., Wang, Y., Yu, Y., Zhang, J., Kong, D., Colen, R. R., Knickmeyer, R. C., Zhu, H., & Initiative, A. D. N. (2017). Fgwas: Functional genome wide association analysis. *NeuroImage*, 159, 107–121.

Matteoli, S., Diani, M., & Corsini, G. (2010). A tutorial overview of anomaly detection in hyperspectral images. *IEEE Aerospace and Electronic Systems Magazine*, 25(7), 5–28.

Ota, T., Shinotoh, H., Fukushi, K., Nagatsuka, S.-i., Namba, H., Iyo, M., Aotsuka, A., Tanaka, N., Sato, K., Shiraishi, T., et al. (2004). A simple method for the detection of abnormal brain regions in Alzheimer's disease patients using [11C] MP4A: Comparison with [123I] IMP SPECT. *Annals of Nuclear Medicine*, 18(3), 187–193.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier, London.

Qian, W. & Titterington, D. M. (1991). Estimation of parameters in hidden Markov models. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, 337(1647), 407–428.

Ramteke, R. & Monali, Y. K. (2012). Automatic medical image classification and abnormality detection using k-nearest neighbour. *International Journal of Advanced Computer Research*, 2(4), 190–196.

Salmond, C., De Haan, M., Friston, K., Gadian, D., & Vargha-Khadem, F. (2003). Investigating individual differences in brain abnormalities in autism. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1430), 405–413.

Schnack, H. G. (2019). Improving individual predictions: Machine learning approaches for detecting and attacking heterogeneity in schizophrenia (and other psychiatric diseases). *Schizophrenia Research*, 214, 34–42.

Shaker, M., Erdogmus, D., Dy, J., & Bouix, S. (2017). Subject-specific abnormal region detection in traumatic brain injury using sparse model selection on high dimensional diffusion data. *Medical Image Analysis*, 37, 56–65.

Shi, F., Liu, B., Zhou, Y., Yu, C., & Jiang, T. (2009). Hippocampal volume and asymmetry in mild cognitive impairment and Alzheimer's disease: Meta-analyses of MRI studies. *Hippocampus*, 19(11), 1055–1064.

Shi, J., Thompson, P. M., Gutman, B., & Wang, Y. (2013). Surface fluid registration of conformal representation: Application to detect disease burden and genetic influence on hippocampus. *NeuroImage*, 78, 111–134.

Shu, H., Nan, B., & Koeppe, R. (2015). Multiple testing for neuroimaging via hidden Markov random field. *Biometrics*, 71(3), 741–750.

Spagnolo, D. M., Al-Kofahi, Y., Zhu, P., Lezon, T. R., Gough, A., Stern, A. M., Lee, A. V., Ginty, F., Sarachan, B., Taylor, D. L., & Chennubhotla, S. C. (2017). Platform for quantitative evaluation

of spatial intratumoral heterogeneity in multiplexed fluorescence images. *Cancer Research*, 77(21), e71–e74.

Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., & Suetens, P. (2001). Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Transactions on Medical Imaging*, 20(8), 677–688.

Varol, E., Sotiras, A., Davatzikos, C., Initiative, A. D. N., et al. (2017). Hydra: Revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. *Neuroimage*, 145, 346–364.

Wang, Y., Song, Y., Rajagopalan, P., An, T., Liu, K., Chou, Y. Y., Gutman, B., Toga, A. W., & Thompson, P. M. (2011). Surface-based TBM boosts power to detect disease effects on the brain: An $n = 804$ ADNI study. *NeuroImage*, 56, 1993–2010.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Cedarbaum, J., Donohue, M. C., Green, R. C., Harvey, D., Jack, C. R., Jr., Jagust, W., Morris, J. C., Petersen, R. C., Saykin, A. J., Shaw, L., Thompson, P. M., Toga, A. W., Trojanowski, J. Q., & Initiative, A. D. N. (2015). Impact of the Alzheimers disease neuroimaging initiative, 2004 to 2014. *Alzheimer's & Dementia*, 11(7), 865–884.

Yekutieli, D. & Benjamini, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference*, 82(1), 171–196.

Zhang, J. T. & Chen, J. (2007). Statistical inferences for functional data. *The Annals of Statistics*, 35(3), 1052–1079.

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.

Zhou, Z.-H. (2018). A brief introduction to weakly supervised learning. *National Science Review*, 5(1), 44–53.

Zhu, H., Kong, L., Li, R., Styner, M., Gerig, G., Lin, W., & Gilmore, J. H. (2011). Fadtts: Functional analysis of diffusion tensor tract statistics. *NeuroImage*, 56(3), 1412–1425.

Zhu, H., Li, R., & Kong, L. (2012). Multivariate varying coefficient model for functional responses. *The Annals of Statistics*, 40, 2634–2666.

Received 17 September 2019
Accepted 15 September 2020