


RESEARCH ARTICLE OPEN ACCESS

Identification of Regions of Interest in Neuroimaging Data With Irregular Boundary Based on Semiparametric Transformation Models and Interval-Censored Outcomes

Chun Yin Lee¹ | Haolun Shi² | Da Ma³ | Mirza Faisal Beg⁴ | Jiguo Cao² 

¹Department of Mathematics, Statistics and Insurance, Hang Seng University of Hong Kong, Hong Kong, China | ²Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, Canada | ³School of Medicine, Wake Forest University, Winston-Salem, USA | ⁴School of Engineering, Simon Fraser University, Burnaby, Canada

Correspondence: Jiguo Cao (jiguo_cao@sfu.ca)

Received: 19 January 2025 | **Revised:** 11 September 2025 | **Accepted:** 22 October 2025

Funding: This work was supported by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2023-04057, RGPIN-2021-02963).

Keywords: Alzheimer's disease | EM algorithm | image analysis | survival analysis | triangulation

ABSTRACT

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that leads to memory loss, cognitive decline, and behavioral changes, without a known cure. Neuroimages are often collected alongside the covariates at baseline to forecast the prognosis of the patients. Identifying regions of interest within the neuroimages associated with disease progression is thus of significant clinical importance. One major complication in such analysis is that the domain of the brain area in neuroimages is irregular. Another complication is that the time to AD is interval-censored, as the event can only be observed between two revisit time points. To address these complications, we propose to model the imaging predictors via bivariate splines over triangulation and incorporate the imaging predictors in a flexible class of semiparametric transformation models. The regions of interest can then be identified by maximizing a penalized likelihood. A computationally efficient expectation–maximization algorithm is devised for parameter estimation. An extensive simulation study is conducted to evaluate the finite-sample performance of the proposed method. An illustration with the AD Neuroimaging Initiative dataset is provided.

1 | Introduction

Alzheimer's disease (AD) is a condition that affects a large number of individuals globally. A report from Alzheimer's Disease International in 2018 indicated that over 50 million people globally are currently living with dementia, a number projected to triple by 2050. In the meantime, structural magnetic resonance imaging (MRI) has been widely used to gain insights into the underlying processes of AD in individual patients. Identifying the brain regions linked to the disease progression from MRI

scans has become an area of crucial research interest, essential for developing targeted treatments and interventions.

Numerous retrospective studies have investigated patients with AD to explore how different brain regions contribute to its onset and advancement. The hippocampus is widely recognized as the key region associated with early-stage disease development due to its involvement in forming long-term memories. Additionally, given the network of interconnected brain regions, it is likely that the onset and progression of AD may involve regions beyond

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Statistics in Medicine* published by John Wiley & Sons Ltd.

the hippocampus. For example, the entorhinal cortex has been shown to associate with the preclinical stage of AD progression, whose posterior-lateral subfields are most susceptible to early neurofibrillary tangle [1]. Presubiculum and subiculum subfields have been found to exhibit the earliest atrophy and are predictive of memory performance in the progression of the disease [2]. Recent studies have also shown from a gradient pattern of pathological markers that the amygdala serves as a hub for propagating AD pathology through its connections with other brain regions [3]. Other key regions related to the disease progression include, for example, ventricle, basal forebrain, cingulate cortex, and white matter tracts [4–7].

Among the brain regions considered in various retrospective studies, to reliably identify the most crucial one that is *statistically* associated with the progression of AD, we propose a data-driven approach for the survival analysis of the large MRI datasets from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. Such a method can help corroborate or challenge existing hypotheses about the role of the various brain regions in the disease progression and may find patterns that may not be apparent through traditional retrospective research methods.

Our objective is to develop a reliable model for detecting regions of interest within brain MRI scans that are associated with the onset of AD. There exist a few statistical challenges to this objective. First, in the ADNI study, the regularly followed-up patients were clinically assessed only at each revisit time point to ascertain the onset or progression of the disease. Therefore, the exact time for the conversion of AD was unknown but only observed to fall between two revisit time points, hence interval-censored [8–10]. The standard estimation procedure, such as the partial likelihood, cannot be directly applied to interval-censored data. Second, since the brain area within MRI scans is confined within an irregular region with noise elsewhere, using existing methods for rectangular (regular) boundaries on such images would result in boundary leakage issues [11]. A direct consequence of the boundary leakage is that the background information would be incorrectly included in the analysis due to the dominance of sharp transitions between the image and background, as demonstrated in several examples in Ramsay [12].

To address these challenges, we propose a novel model for identifying key areas in brain MRI scans that are statistically associated with the progression of AD. Our method specifically addresses the difficulties related to the irregular contours found in MRI images and the interval-censored data based on a flexible class of semiparametric transformation models. We propose to approximate the image with irregular boundaries via bivariate polynomial splines constructed on triangulation grids as proposed in Lai and Schumaker [13] and Lai and Wang [14]. By utilizing piecewise polynomial functions on a triangulated surface, the brain scans can be smoothly incorporated into a survival model. This allows for the precise location of the area of interest within a regression framework. Moreover, to incorporate the interval-censored data, we develop a Poisson data augmentation approach, which is shown to be computationally efficient and powered by the Expectation Maximization (EM) algorithm. Finally, as some but not all regions of the brain are of interest, we propose to incorporate group penalty terms on the imaging predictors, selectively emphasizing areas implicated in disease onset.

The regions of interest can then be identified by maximizing a penalized likelihood. The R codes for implementing our method in this paper are publicly available on GitHub: <https://github.com/lcyjames/Roiico>.

Incorporating high-dimensional images for survival models has received considerable attention recently, with a majority of the research work focused on functional principal component analysis (FPCA). Zipunnikov et al. [15] established a linkage between the FPCA and singular value decomposition, which allows for computationally efficient estimation of the single-level FPCA for vectorized images. To retain the spatial structure of the image, Reiss and Ogden [16] studied the functional generalized linear models regressing scalar outcomes on two-dimensional image predictors, approximated via the radial B-spline function. Zhou et al. [17] later extended the generalized linear regression model to incorporate tensor-structured predictors. They assumed a rank- R decomposition for the tensor such that the number of parameters to be estimated is substantially reduced. Wang et al. [18] proposed a minimum penalized total variation approach, based on the scalar-on-image regression model, that yields piecewise-smooth regression coefficients. Jiang et al. [19] studied the supervised FPCA to analyze two-dimensional imaging data and performed risk prediction by incorporating the estimated FPC scores into the Cox model for right-censored outcomes. Bayesian nonparametric methods have also been developed to estimate scalar-on-image regression models; for instance, see Goldsmith et al. [20], Kang et al. [21], and Feng et al. [22]. However, the aforementioned work typically assumes that the image is bounded within a rectangular domain, and the outcome variable is often noncensored or right-censored. Mattos et al. [23] introduced an estimation approach for the tuning parameter in a semiparametric linear mixed effects model, which incorporates a roughness penalty on the unknown smooth function. A scant amount of work has paid attention to the association between image predictors and interval-censored outcome variables.

The rest of the paper is structured as follows. Section 2 first introduces the notations, the models for image predictor and survival outcome, and the likelihood. A flexible class of semiparametric transformation models incorporating imaging predictors is proposed for the survival outcomes. A sieve maximum likelihood approach for interval-censored data with group penalties is proposed for identifying the regions of interest. In Section 3, we discuss the computation algorithms of the penalized estimator. Extensive simulation studies are conducted in Section 4, investigating the finite-sample performance of the proposed methods. We illustrate the methods based on the ADNI dataset in Section 5. Some concluding remarks are provided in Section 6.

2 | Model and Likelihood

Let $X(s)$ be a random two-dimensional stochastic process modeling the spatially correlated pixels of an image, where location $s \equiv (s_1, s_2) \in \Omega \subset \mathcal{R}^2$, and Ω is a bounded two-dimensional domain of possibly irregular shape. Without loss of generality, we assume that $X(\cdot)$ has mean 0. Suppose that there are n realizations of the random process, where each realization pertains to a two-dimensional surface. Furthermore, we observe N_i sets

of noisy samples from the i th surface, $i = 1, \dots, n$. The observed data point Y_{ij} is given by

$$Y_{ij} = X_i(s_{ij}) + \epsilon_{ij},$$

for $i = 1, \dots, n$ and $j = 1, \dots, N_i$ where $E(\epsilon_{ij}) = 0$, $E(\epsilon_{ij}, \epsilon_{ij'}) = I(j = j')\sigma^2$, ϵ_{ij} is independent of s_{ij} , and $I(\cdot)$ is the indicator function. Note that $X_i(\cdot)$ is intrinsically an infinite-dimensional function measured on a two-dimensional space with an irregular boundary, which refers to a circular region along the axial plane in the study of neuroimaging data. We approximate $X_i(\cdot)$ via the bivariate polynomial splines constructed on the partitions over a two-dimensional triangulated domain [13, 14]. Compared with other polygon shapes, triangulation offers the advantage that any polygon domain of arbitrary shape can be partitioned into finitely many triangles, thus providing greater flexibility. Let $\Delta = \{\tau_1, \dots, \tau_M\}$ be the collection of triangles that forms a triangulation of $\Omega = \cup_{l=1}^M \tau_l$, where each τ_l is the convex hull of three points that are not colinear, and any nonempty intersection between τ_l and $\tau_{l'}$ ($l \neq l'$) is either a common vertex or a common edge. The spline space of a degree d and smoothness r over triangulation Δ can be defined as $\mathcal{T}_d^r(\Delta) = \{u \in C^r(\Omega) : u|_{\tau} \in \mathbb{P}_d, \tau \in \Delta\}$ where $r \geq 0$, $C^r(\Omega)$ denote the collection of all r -th continuously differentiable functions over Ω , and \mathbb{P}_d denotes the space of all polynomials with degree not greater than d .

Specifically, we adopt the Bernstein polynomials, with each polynomial tailored to a specific triangle in Ω . Suppose that $\tau \in \Delta$ is a nondegenerate triangle (with a nonzero area) with vertices $\mathbf{v}_1, \mathbf{v}_2$, and \mathbf{v}_3 . Then, an arbitrary point $s \in \Omega$ can be uniquely represented by $s = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + b_3\mathbf{v}_3$, where (b_1, b_2, b_3) are known as the barycentric coordinates of the point s relative to triangle τ , and $b_1 + b_2 + b_3 = 1$. We define a set of Bernstein polynomial basis functions of degree d for s relative to τ by $\mathbf{B}^{\tau,d}(s)^T = \{B_{ijk}^{\tau,d}(s), i + j + k = d\}$, with dimension $K_d = \binom{d+2}{2}$, where $B_{ijk}^{\tau,d}(s) = d!(i!j!k!)^{-1}b_1^i b_2^j b_3^k$. Therefore, the polynomial piece of spline u restricted to τ is given by $u|_{\tau} = \alpha^{\tau T} \mathbf{B}^{\tau,d}$, where $\alpha^{\tau T} = \{\alpha_{ijk}^{\tau}, i + j + k = d\}$ is a K_d -dimensional vector of coefficients of the basis functions. In the sequel, we assume that d is common to all triangles, and we drop the subscript d in K and superscript d in \mathbf{B} for notational simplicity. The Bernstein polynomial based approximation of $X_i(s)$ is given by

$$\tilde{X}_i(s) = \sum_{l=1}^M \alpha_i^{\tau_l T} \mathbf{B}^{\tau_l}(s). \quad (1)$$

The estimates of $\alpha_i = (\alpha_i^{\tau_1 T}, \dots, \alpha_i^{\tau_M T})^T$ can be obtained by minimizing the least squares criterion with the thin-plate spline penalty [24, 25]:

$$\sum_{i=1}^n \sum_{j=1}^{N_i} \{Y_{ij} - \tilde{X}_i(s_{ij})\}^2 + c_i \sum_{\tau \in \Delta} \int_{\tau} \sum_{i'+j'=2}^2 \binom{2}{i'} (\nabla_{s_1}^{i'} \nabla_{s_2}^{j'} \tilde{X}_i)^2 ds_1 ds_2, \quad (2)$$

where $c_i > 0$ pertains to the roughness penalty parameter, and $\nabla_{s_1}^p$ pertains to the p th-order derivative in the direction s_1 at the point $s = (s_1, s_2)$. In addition, we require $\mathbf{H}\alpha = \mathbf{0}$ to meet the smoothness condition of the splines, where \mathbf{H} is a matrix that enforces smoothness across shared edges of triangles; some

examples on the construction of \mathbf{H} can be found in Zhou and Pan [26].

In addition to the image data, let T be the failure time, and \mathbf{Z} be a vector of demographic covariates. We propose a flexible class of semiparametric transformation models for T with the conditional cumulative hazard function

$$\Lambda(t|X_i, \mathbf{Z}_i) = G \left[\Lambda(t) \exp \left\{ \beta^T \mathbf{Z}_i + \int_{s \in \Omega} \gamma(s) X_i(s) ds \right\}; \rho \right], \quad (3)$$

where G is a prespecified strictly increasing transformation function indexed by parameter ρ , Λ is an unspecified increasing function with $\Lambda(0) = 0$, β is a vector of regression parameters, and $\gamma(s)$ is an imaging parameter which characterizes the effects of imaging data on the hazard function. The transformation model in Equation (3) reduces to the Cox proportional hazards (PH) model when $G(x; \rho) = x$, and the proportional odds (PO) model when $G(x; \rho) = \log(1 + x)$, respectively. Hence, it is reminiscent of the logarithmic or the Box-Cox transformation functions. In this paper, we assume the logarithmic transformation function $G(x; \rho) = \log(1 + \rho x)/\rho$ with $\rho \geq 0$, where $\rho = 0$ and $\rho = 1$ correspond to the Cox PH and PO models, respectively. Equation (3) equipped with the logarithmic transformation function is also known as the generalized odds rate model [27] in the literature. For mathematical tractability, we approximate $X_i(s)$ in Equation (3) by $\tilde{X}_i(s)$ in Equation (1), and suppose that $\gamma(s)$ can also be expanded on the set of bivariate Bernstein polynomial basis functions, that is $\gamma(s) = \sum_{l=1}^M \gamma^{\tau_l T} \mathbf{B}^{\tau_l}(s)$, where γ^{τ_l} is a K -dimensional vector of coefficients of the basis functions restricted to the l th triangle, $l = 1, \dots, M$. Then, we can rewrite Equation (3) as

$$\begin{aligned} \Lambda(t|\tilde{X}_i, \mathbf{Z}_i) &= G \left[\Lambda(t) \exp \left\{ \beta^T \mathbf{Z}_i + \int_{s \in \Omega} \left(\sum_{l=1}^M \gamma^{\tau_l T} \mathbf{B}^{\tau_l}(s) \right) \left(\sum_{l'=1}^M \alpha_i^{\tau_{l'} T} \mathbf{B}^{\tau_{l'}}(s) \right) ds \right\}; \rho \right] \\ &= G \left[\Lambda(t) \exp \left\{ \beta^T \mathbf{Z}_i + \sum_{l=1}^M \int_{s \in \Omega} \gamma^{\tau_l T} \mathbf{B}^{\tau_l}(s) \alpha_i^{\tau_l T} \mathbf{B}^{\tau_l}(s) ds \right\}; \rho \right] \\ &= G \left[\Lambda(t) \exp \left\{ \beta^T \mathbf{Z}_i + \sum_{l=1}^M \gamma^{\tau_l T} \mathbf{W}^{\tau_l} \alpha_i^{\tau_l T} \right\}; \rho \right], \end{aligned} \quad (4)$$

where $\mathbf{W}^{\tau_l} = \langle \mathbf{B}^{\tau_l}, \mathbf{B}^{\tau_l T} \rangle$ is a $K \times K$ dimensional matrix, and $\langle \cdot, \cdot \rangle$ denotes the inner product of functions. The second last line of Equation (4) is obtained based on the fact that $\langle \mathbf{B}^{\tau_l}, \mathbf{B}^{\tau_{l'} T} \rangle = 0$ for any $l' \neq l$. Let $\mathbf{W} = \text{diag}(\mathbf{W}^{\tau_1}, \dots, \mathbf{W}^{\tau_M})$ be a block diagonal and positive definite matrix, and $\xi_i = \mathbf{W} \alpha_i$. Then, Equation (4) becomes

$$\Lambda(t|\tilde{X}_i, \mathbf{Z}_i) = G \left\{ \Lambda(t) e^{\beta^T \mathbf{Z}_i + \gamma^T \xi_i}; \rho \right\}, \quad (5)$$

where $\gamma = (\gamma^{\tau_1 T}, \dots, \gamma^{\tau_M T})^T$.

Suppose T_i is censored by a random interval (L_i, R_i) , $i = 1, \dots, n$. In the framework, we set $L_i = 0$ and $R_i = \infty$ to accommodate left- and right-censored observations. Under the paradigm

of noninformative censoring, and for given ξ_i 's, the observed likelihood of (Λ, β, γ) is proportional to

$$\prod_{i=1}^n \left[e^{-G\{\Lambda(L_i)e^{\beta^T Z_i + \gamma^T \xi_i}; \rho\}} - e^{-G\{\Lambda(R_i)e^{\beta^T Z_i + \gamma^T \xi_i}; \rho\}} \right].$$

Since Λ is modeled nonparametrically, we propose a sieve maximum likelihood approach for parameter estimation. We approximate $\Lambda(t)$ via $\tilde{\Lambda}(t) = \sum_{j=1}^J \omega_j I_j(t)$, where I_1, \dots, I_J are the I-spline basis functions [28] constructed over $[0, t_{\max}]$, and $\omega = (\omega_1, \dots, \omega_J)$ are nonnegative coefficients ensuring that $\tilde{\Lambda}$ is a monotone function. In practice, t_{\max} is set to be the largest finite value of the combined vector of L_i 's and R_i 's. The observed likelihood of $\theta \equiv (\omega, \beta, \gamma)$ under sieve approximation is given by

$$L(\theta) = \prod_{i=1}^n \left[e^{-G\{\sum_{j=1}^J \omega_j I_j(L_i)e^{\beta^T Z_i + \gamma^T \xi_i}; \rho\}} - e^{-G\{\sum_{j=1}^J \omega_j I_j(R_i)e^{\beta^T Z_i + \gamma^T \xi_i}; \rho\}} \right]. \quad (6)$$

In general, not all regions of the image are of clinical interest, and we wish to identify some useful regions for disease prognosis. We adopt the group lasso [29, 30] to penalize the parameters associated with each of the M triangles. Specifically, we propose to maximize the following objective function:

$$\log L(\theta) - \lambda \sum_{l=1}^M s(K_l) \|\gamma^{\tau_l}\|_2, \quad (7)$$

where $\|\cdot\|_2$ denotes the Euclidean norm, $K_1 = \dots = K_M = K$, $s(\cdot)$ is a function used to rescale the penalty corresponding to the dimensionality of γ^{τ_l} , and $\lambda \geq 0$ is a tuning parameter that controls the size of the penalty. As is customary, we set $s(K_l) = \sqrt{K}$ pertaining to a normalizing constant for $l = 1, \dots, M$, and we normalize each element in ξ_i before carrying out the estimation procedure. The demographic variables Z_i are not penalized in this setting. The group lasso penalty acts as an intermediate between the l_1 - and l_2 -type penalties, selecting variables at the group level. Consequently, if the l th triangle is not penalized, then all values in γ^{τ_l} remain nonzero, retaining all the basis functions restricted to that triangle.

3 | Computational Details

We propose a two-step estimation procedure for computational efficiency. The estimation in Equation (2) can be handled easily via some existing software, such as the BPST package in R. Given the estimates of ξ_i 's, we then approximate θ by maximizing the penalized likelihood in Equation (7) with the detailed procedures as follows.

To simplify the computation, we adopt a class of frailty-induced transformation models [31]. Let $G(x; \rho) = -\log \Phi_\zeta(x; \rho)$ where $\Phi_\zeta(x; \rho) = \int_0^\infty \exp(-\zeta x) f_\zeta(\zeta; \rho) d\zeta$ is the Laplace transform of a positive-valued frailty ζ , and $f_\zeta(\cdot; \rho)$ is the density function of ζ . We obtain the logarithmic transformation $G(x; \rho) = \log(1 + \rho x)/\rho$ when ζ follows a gamma distribution with mean 1 and variance ρ . Furthermore, let Δ_{Li} , Δ_{Ri} , and $\Delta_{Ii} \equiv 1 - \Delta_{Li} - \Delta_{Ri}$ be the indicators for left-, right-, and interval-censored

observations, respectively. Then, Equation (6) can be written as

$$\prod_{i=1}^n \int \left\{ 1 - e^{-\sum_{j=1}^J \zeta_i \omega_j I_j(R_i) e^{\beta^T Z_i + \gamma^T \xi_i}} \right\}^{\Delta_{Li}} \left\{ e^{-\sum_{j=1}^J \zeta_i \omega_j I_j(L_i) e^{\beta^T Z_i + \gamma^T \xi_i}} - e^{-\sum_{j=1}^J \zeta_i \omega_j I_j(R_i) e^{\beta^T Z_i + \gamma^T \xi_i}} \right\}^{\Delta_{Ii}} \times \left\{ e^{-\sum_{j=1}^J \zeta_i \omega_j I_j(L_i) e^{\beta^T Z_i + \gamma^T \xi_i}} \right\}^{\Delta_{Ri}} f_\zeta(\zeta_i; \rho) d\zeta_i, \quad (8)$$

where $f_\zeta(\cdot; \rho)$ is the density of a gamma distribution with both shape and rate parameters equal to ρ^{-1} . The maximization of Equation (8) is challenging due to the lack of analytical expressions for $\omega_1, \dots, \omega_J$. To address this issue and accommodate the latent variable ζ_i , we propose a data augmentation approach and develop an EM algorithm for obtaining the sieve maximum likelihood estimate (MLE). For $i = 1, \dots, n$ and $j = 1, \dots, J$, let A_{ij} and B_{ij} be two latent Poisson variables that are independent given ζ_i , with conditional means $\zeta_i \mu_{ij}$ and $\zeta_i \eta_{ij}$ respectively, where $\mu_{ij} \equiv \omega_j \{\Delta_{Li} I_j(R_i) + \Delta_{Ii} I_j(L_i)\} e^{\beta^T Z_i + \gamma^T \xi_i}$ and $\eta_{ij} \equiv \omega_j \{\Delta_{Ii} \{I_j(R_i) - I_j(L_i)\} + \Delta_{Ri} I_j(L_i)\} e^{\beta^T Z_i + \gamma^T \xi_i}$. Furthermore, let $A_i \equiv \sum_{j=1}^J A_{ij}$, $B_i \equiv \sum_{j=1}^J B_{ij}$, and

$$\tilde{\theta}_i^T = \begin{cases} A_i > 0, B_i = 0, & \text{if } \Delta_{Li} = 1; \\ A_i = 0, B_i > 0, & \text{if } \Delta_{Ii} = 1; \\ A_i = 0, B_i = 0, & \text{otherwise.} \end{cases}$$

One can see that the observed likelihood in Equation (8) is equivalent to that obtained from the data $(\mathcal{O}_1, \dots, \mathcal{O}_n)$ where $\mathcal{O}_i \equiv (L_i, R_i, Z_i, \xi_i, \tilde{\theta}_i^T)$, treating A_{ij} , B_{ij} , and ζ_i as missing data. Therefore, the maximizer of Equation (8) can be computed based on the EM algorithm with the augmented data. It suffices to write down the complete data log-likelihood

$$\ell_C(\theta) = \sum_{i=1}^n \sum_{j=1}^J (-\mu_{ij} \zeta_i + A_{ij} \log \mu_{ij} - \eta_{ij} \zeta_i + B_{ij} \log \eta_{ij}),$$

omitting the terms which do not depend on θ .

Let $\theta^{(d)}$ denote the current estimates of θ in the d th iteration ($d = 0, 1, \dots$) of the EM algorithm. In the E-step, we evaluate $\hat{E}(A_{ij})$, $\hat{E}(B_{ij})$, and $\hat{E}(\zeta_i)$ where $\hat{E}(\cdot)$ denotes the conditional expectation given $(\mathcal{O}_1, \dots, \mathcal{O}_n)$ evaluated at $\theta^{(d)}$. The analytical expressions of these conditional expectations are relegated to the Appendix A, where we show that only conditional expectations of functions of ζ_i are required in the E-step. In particular, those integrals have closed-form expressions, thus, the computationally intensive numerical approximation is not required in the proposed method.

In the M-step, we obtain $\theta^{(d+1)}$ by maximizing the penalized Q -function $Q(\theta, \theta^{(d)}) \equiv \hat{E}(\ell_C(\theta)) - \lambda \sum_{l=1}^M s(K_l) \|\gamma^{\tau_l}\|_2$. Taking the partial derivative of $Q(\theta, \theta^{(d)})$ with respect to ω and setting it to zero yields a closed-form solution for ω , in terms of β and γ , as

$$\omega_j^*(\beta, \gamma) = \frac{\sum_{i=1}^n \Delta_{Li} \hat{E}(A_{ij}) + \Delta_{Ii} \hat{E}(B_{ij})}{\sum_{i=1}^n e^{\beta^T Z_i + \gamma^T \xi_i} \hat{E}(\zeta_i) \{\Delta_{Ri} I_j(L_i) + (1 - \Delta_{Ri}) I_j(R_i)\}},$$

for $j = 1, \dots, J$.

By replacing ω_j with $\omega_j^*(\beta, \gamma)$ in $Q(\theta, \theta^{(d)})$, the revised penalized Q-function is

$$\hat{E}(\ell_C(\omega^*, \beta, \gamma)) - \lambda \sum_{l=1}^M s(K_l) \|\gamma^{\tau_l}\|_2, \quad (9)$$

where $\hat{E}(\ell_C(\omega^*, \beta, \gamma))$ is given by

$$\sum_{i=1}^n \sum_{j=1}^J \left(\beta^T Z_i + \gamma^T \xi_i - \log \left[\sum_{k=1}^n e^{\beta^T Z_k + \gamma^T \xi_k} \hat{E}(\zeta_k) \Delta_{Rk} I_j(L_k) \right] \right. \\ \left. + (1 - \Delta_{Rk}) I_j(R_k) \right) \left\{ \Delta_{Li} \hat{E}(A_{ij}) + \Delta_{Li} \hat{E}(B_{ij}) \right\}.$$

Based on Equation (9), $\beta^{(d+1)}$ is obtained via a one-step Newton–Raphson algorithm. Then, we update γ via the block coordinate descent algorithm [30, 32]. Specifically, we fix the values of $\gamma^{\tau_{l'}}$ for any $l' \neq l$, and (ω, β) at $(\omega^{(d)}, \beta^{(d+1)})$, while updating γ^{τ_l} . Based on the direct consequence of the Karush–Kuhn–Tucker conditions, we have $\gamma^{\tau_l^{(d+1)}} = \mathbf{0}$ if $\|\mathcal{G}_l\|_2 \leq \lambda s(K_l)$ where $\mathcal{G}_l = \left\{ \frac{\partial}{\partial \gamma^{\tau_l}} \hat{E}(\ell_C(\omega^*, \beta, \gamma)) \right\}_{\gamma^{\tau_l} = \mathbf{0}}$, and $\gamma^{\tau_l^{(d+1)}}$ is obtained via minimizing the negative penalized-Q function in Equation (9) otherwise, for $l = 1, \dots, M$. Then, we obtain $\omega_j^{(d+1)} = \omega_j^*(\beta^{(d+1)}, \gamma^{(d+1)})$ for $j = 1, \dots, J$. We iterate between the E- and M-steps until convergence, where the maximum absolute difference between two consecutive estimates is smaller than a certain threshold.

Determination of the tuning parameter λ is a crucial task in the estimation procedure. We propose to implement a data-adaptive Akaike information criterion (AIC):

$$\text{AIC}_\lambda = -2 \log L(\hat{\theta}_\lambda) + 2 \text{df}_\lambda, \quad (10)$$

where $\hat{\theta}_\lambda$ is the estimator obtained in Equation (7) under a pre-specified λ , and the degree of freedom df_λ is taken as the total number of nonzero estimates in $\hat{\theta}_\lambda$. In practice, we consider a grid of values for λ and choose the model with a minimum value of AIC_λ . This criterion performs well as indicated in our simulation studies. Other model selection methods, such as the Bayesian information criterion [33] or cross-validation, can also be considered.

4 | Simulation Study

We study the finite-sample performance of the proposed methods to identify the regions of interest. To mimic the practical situation in which images are observed with an irregular boundary, we confine the images to a rectangular grid of size 40×40 , where 1036 of them are situated within a heart-shaped boundary region. We construct the sets of bivariate Bernstein polynomials based on two settings on the fineness of the triangulation, namely $M = 62$ triangles with 47 vertices, and $M = 118$ triangles with 80 vertices. The triangulation grids used in these two settings are depicted in Figure 1. Specifically, the j th pixel of the image surface observed for the i th subject is given by

$$Y_{ij} = \sum_{l=1}^M \alpha_i^{\tau_l T} B^{\tau_l}(s_{ij}) + \epsilon_{ij},$$

for $i = 1, \dots, n; j = 1, \dots, N_i$, and $N_i = 1036$ for all i . Specifically, we set the degree of the Bernstein polynomial $d = 2$ (i.e., $K = 6$) and ϵ_{ij} 's are generated from a normal distribution with mean 0 and variance 0.1². We generate α_i from a zero-mean multivariate normal distribution with covariance matrix Σ where the diagonal elements of Σ are equal to 1, and the within-triangle and between-triangle correlations are equal to 0.4 and 0.1, respectively. Subsequently, we generate the survival time T_i based on the model:

$$\Lambda(t|W, \alpha_i, Z_i) = G \left[\Lambda(t) \exp \left\{ \beta^T Z_i + \sum_{l \in S_i} \gamma^{\tau_l T} W^{\tau_l} \alpha_i^{\tau_l} \right\}; \rho \right],$$

where S_i is a set of indices of the triangles that are assigned to be associated with the survival outcomes. We set the basis coefficient $\gamma^{\tau_l} = (0.1, 0.2, 0.3, 0.5, 0.6, 0.4)$ for all $l \in S_i$, and $\gamma^{\tau_l} = \mathbf{0}$ otherwise. We set $\Lambda(t) = t^2/4$, and the transformation parameter ρ to be 0, 0.5 and 1. The interarrival times follow a uniform distribution with support (0, 1, 0.5). We assume that there is an administrative censoring time, such that observations are right-censored beyond this time point. The average right-censoring proportion varies from 30% to 45% across the scenarios. We set $\beta = (0.5, -0.5)$ and $Z_i = (Z_{i1}, Z_{i2})$ where Z_{i1} follows a Bernoulli distribution with probability 0.5 and Z_{i2} follows the standard normal distribution. We consider $n = 100, 200, 500$, and 1000, and the number of selected triangles $|S_i|$ to be $\lfloor aM \rfloor$, the largest integer value smaller than or equal to aM , where $a = 0.025, 0.05$, and 0.1.

We adopt the I-splines basis functions with degree 3 and 4 interior knots (i.e., $J = 7$) to approximate Λ . The knots were positioned at the empirical quantiles of the endpoints of the observed intervals. For implementing the EM algorithm, we set the initial parameter values as follows: $\omega_j^{(0)} = 0.1$ for $j = 1, \dots, J$, $\beta^{(0)} = \mathbf{0}$, and $\gamma^{(0)} = \mathbf{0}$. The tolerance level for the convergence criterion is set to be 10^{-3} . The optimal value of λ is selected by AIC, as discussed in Section 3.

Table 1 presents the simulation results for two different triangulation grid settings, with 100 replicates in each scenario. The performance of the proposed methods is assessed based on the true positive rate and false positive rate. The true positive rate is defined as the proportion of triangles in S_i being correctly identified, while the false positive rate is the proportion of triangles which does not belong to S_i but are identified. The two rates are computed for each replicate, and the average rates are reported. The empirical results demonstrate that the proposed method performs satisfactorily under the transformation models in general, except when the sample size is as small as $n = 100$ and when the number of triangles in S_i is relatively large, such as when $a = 0.1$, as expected. This indicates that the proposed method can effectively identify all the regions of interest that are truly associated with survival time. Also, the false positive rates remain consistently low across all scenarios, suggesting that the proposed method often accurately excludes regions without any impact during the variable selection process.

To assess the potential impact of an incorrectly specified triangulation grid on the performance of the proposed methods, we conduct a simulation where the grid used for data generation differs from the grid used for identifying regions of interest. Specifically, data are generated using a (true) grid with $M = 62$ triangles, with

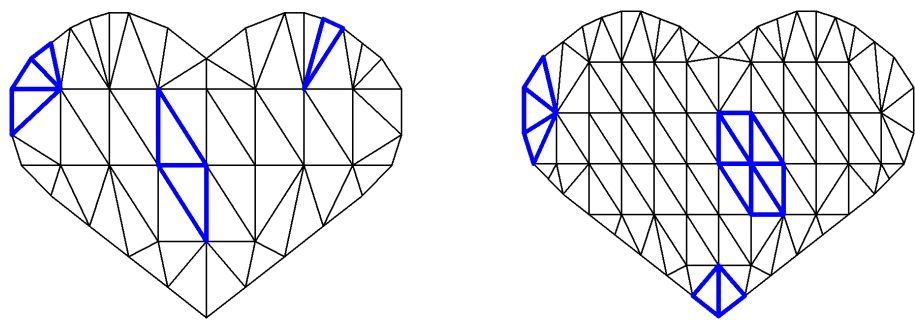


FIGURE 1 | Illustration of the triangulation grids used in the simulation study. The left and right panels displayed grids of 62 and 118 triangles, respectively. The triangles with thickened boundaries illustrate the mechanism by which a certain proportion of the triangles is selected to be associated with the survival outcome.

TABLE 1 | Empirical true positive (TP) and false positive (FP) for identifying regions of interest with different numbers of triangulation grids M , the proportion of selected triangles a , and sample sizes n for $\rho = 0$ (PH model), $\rho = 0.5$, and $\rho = 1$ (PO model).

		$M = 62$						$M = 118$					
		$a = 0.025$		$a = 0.05$		$a = 0.1$		$a = 0.025$		$a = 0.05$		$a = 0.1$	
ρ	n	TP	FP	TP	FP	TP	FP	TP	FP	TP	FP	TP	FP
0	100	1.000	0.003	0.847	0.007	0.597	0.022	0.995	0.002	0.816	0.014	0.143	0.003
	200	1.000	0.004	1.000	0.012	0.963	0.040	1.000	0.002	0.998	0.016	0.645	0.032
	500	1.000	0.010	1.000	0.009	1.000	0.030	1.000	0.002	1.000	0.013	0.874	0.039
	1000	1.000	0.029	1.000	0.015	1.000	0.035	1.000	0.008	1.000	0.014	0.929	0.040
0.5	100	0.990	0.003	0.850	0.008	0.578	0.018	0.995	0.003	0.730	0.012	0.122	0.004
	200	1.000	0.004	1.000	0.013	0.970	0.037	1.000	0.002	1.000	0.019	0.650	0.030
	500	1.000	0.007	1.000	0.006	1.000	0.028	1.000	0.003	1.000	0.012	0.894	0.044
	1000	1.000	0.019	1.000	0.014	1.000	0.031	1.000	0.004	1.000	0.013	0.940	0.050
1.0	100	0.980	0.002	0.877	0.009	0.575	0.019	0.980	0.002	0.798	0.012	0.097	0.003
	200	1.000	0.003	0.997	0.010	0.960	0.041	1.000	0.002	0.994	0.015	0.564	0.020
	500	1.000	0.006	1.000	0.009	0.998	0.030	1.000	0.003	1.000	0.011	0.878	0.041
	1000	1.000	0.012	1.000	0.008	1.000	0.031	1.000	0.003	1.000	0.009	0.935	0.049

those associated with survival outcomes depicted in the left panel of Figure 1. We then fit the model using a (misspecified) triangulation grid with $M^* = 118$ triangles, and vice versa. All other parameter settings remain unchanged as above. As our main objective is to evaluate the precision of the identification methods, it necessitates a surrogate measurement for determining true and false positive rates under the conditions of misspecified grids. To achieve this, we examine each triangle within the misspecified grid and classify it as associated with the outcome if its centroid falls within the region defined by the true grid. Note that varying triangle proportions a in the true grid yield different numbers of (associated) surrogate triangles in the misspecified grid, leading to different surrogate proportions a^* . Figure 2 illustrates the surrogate triangles, marked with red dashed lines, which are utilized to assess the surrogate performance of the methods, alongside the true region of interest bounded with blue solid lines. The results, along with the values of M , M^* , a , and a^* in the settings, are summarized in Table 2.

When an overly refined grid is erroneously applied to the data, specifically when the true grid size is $M = 62$ but a grid size

of $M^* = 118$ is applied, the true positive rates tend to diminish. This decline is likely attributable to the discordance between the true and misspecified grids, as well as a decrease in the sparsity of the associated triangles, given that the number of surrogate triangles exceeds that of the true triangles. Conversely, when a coarser grid is mistakenly applied, where the true grid size is $M = 118$ but a grid size of $M^* = 62$ is used, the true positive rates remain satisfactory. This result may be due to an increase in the sparsity of the associated triangles. Overall, the false positive rates exhibit a moderate increase, particularly when the adopted grid is coarser than the true grid, compared to scenarios where the triangulation grids are correctly specified. The results demonstrate that the proposed methods still manage to provide reasonable performance, although careful selection of the triangulation grid can enhance precision. In practice, the optimal choice of M is unknown. Adopting a balanced triangulation grid guided by practitioners, that is both clinically meaningful and tailored to the specific context of the study, is recommended to enhance the performance of the identification methods. Some guidelines for the choice of triangulation can be found in Wang et al. [25].

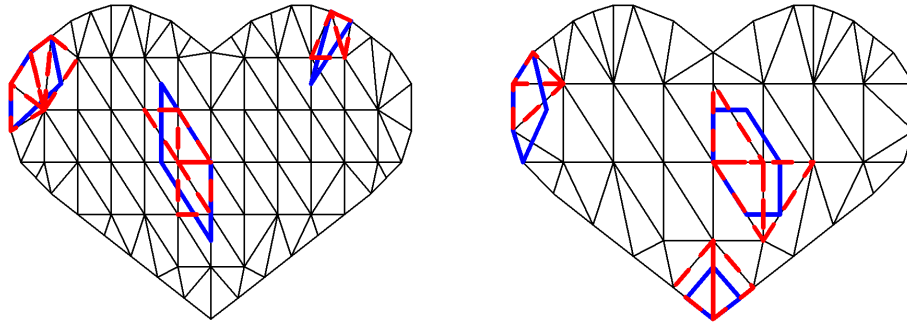


FIGURE 2 | Illustration of the surrogate triangles to facilitate the definition of true and false positive rates under a misspecified triangulation grid in the simulation. In the left (right) panel, the blue solid lines indicate the triangles associated with the survival outcome with triangulation grids $M = 62$ ($M = 118$), and the red dashed lines indicate the surrogate triangles associated with the survival outcome under a misspecified grid where $M^* = 118$ ($M^* = 62$).

TABLE 2 | Empirical true positive (TP) and false positive (FP) for identifying regions of interest under misspecified triangulation grids, the proportion of selected triangles a , and sample sizes n for $\rho = 0$ (PH model), $\rho = 0.5$, and $\rho = 1$ (PO model). Here, “prop” stands for proportion.

True grid		$M = 62$						$M = 118$					
True prop		$a = 0.025$		$a = 0.05$		$a = 0.1$		$a = 0.025$		$a = 0.05$		$a = 0.1$	
Adopted grid		$M^* = 118$						$M^* = 62$					
Surrogate prop		$a^* = 0.017$		$a^* = 0.051$		$a^* = 0.085$		$a^* = 0.032$		$a^* = 0.065$		$a^* = 0.113$	
ρ	n	TP	FP	TP	FP	TP	FP	TP	FP	TP	FP	TP	FP
0	100	0.470	0.006	0.370	0.007	0.241	0.006	0.985	0.008	0.772	0.034	0.324	0.021
	200	0.505	0.009	0.677	0.016	0.590	0.024	1.000	0.007	0.973	0.041	0.713	0.057
	500	0.505	0.010	0.808	0.022	0.770	0.048	1.000	0.020	0.995	0.042	0.856	0.083
	1000	0.500	0.012	0.830	0.023	0.819	0.050	1.000	0.051	1.000	0.050	0.899	0.096
0.5	100	0.375	0.005	0.370	0.006	0.187	0.006	0.985	0.005	0.807	0.031	0.303	0.021
	200	0.550	0.009	0.708	0.018	0.599	0.026	1.000	0.006	0.948	0.036	0.659	0.058
	500	0.500	0.010	0.830	0.023	0.782	0.043	1.000	0.015	1.000	0.039	0.846	0.084
	1000	0.500	0.010	0.847	0.022	0.840	0.055	1.000	0.035	1.000	0.047	0.876	0.095
1.0	100	0.340	0.004	0.308	0.006	0.148	0.004	0.980	0.006	0.760	0.031	0.244	0.017
	200	0.530	0.008	0.723	0.018	0.593	0.024	1.000	0.007	0.948	0.036	0.687	0.060
	500	0.515	0.010	0.832	0.022	0.778	0.040	1.000	0.014	0.998	0.034	0.850	0.092
	1000	0.500	0.010	0.847	0.022	0.823	0.051	1.000	0.029	1.000	0.044	0.877	0.097

5 | Application to AD Study

We demonstrate the proposed methods using the ADNI dataset [34]. This is a longitudinal study in which the cognitive status of the subjects was reassessed approximately every 6 months in the first 2 years, followed by annual evaluations beyond 2 years. The maximum number of visit time points was 13, and the maximum follow-up period was 120 months. At each visit time point, the subjects underwent brain MRI scans and were clinically classified into one of the three stages of AD, namely cognitive normal (CN), mild cognitive impairment (MCI), and AD. We treat the onset of AD as the event of interest, which was interval-censored due to the study design. The main objective of this analysis is to identify the region of interest in the brain MRI scan associated with the survival time of the patients. We focus on a subset of subjects initially diagnosed as CN or MCI

(i.e., AD-free) at baseline. This yields a sample size of $n = 674$ subjects of which 286 have been diagnosed with AD by the end of the study. Apart from the volumetric images at baseline, other covariates, including gender and age at baseline, are available for each subject. Before the statistical analysis, the MRI scans are preprocessed to ensure that they are approximately transformed into the same stereotaxic space, based on the registration procedure in Ma et al. [35]. We proceed to extract the middle slice from both the axial and sagittal axes of the processed image and store the pixel intensities of the slices in the format of a 200×200 matrix.

As illustrated in Figure 3, we construct boundaries to envelop the brain regions on the axial and sagittal planes in the MRI scans and partition the regions into $M = 118$ and $M = 103$ triangles, respectively. For $i = 1, \dots, 674$, a total number of $N_i =$

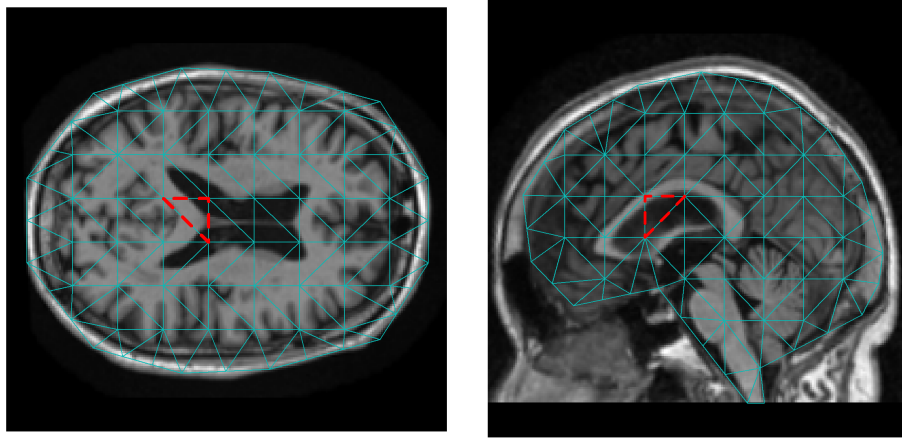


FIGURE 3 | The triangulation over the region of the brain. Left panel: Axial plane with $M = 118$ triangles. Right panel: Sagittal plane with $M = 103$ triangles.

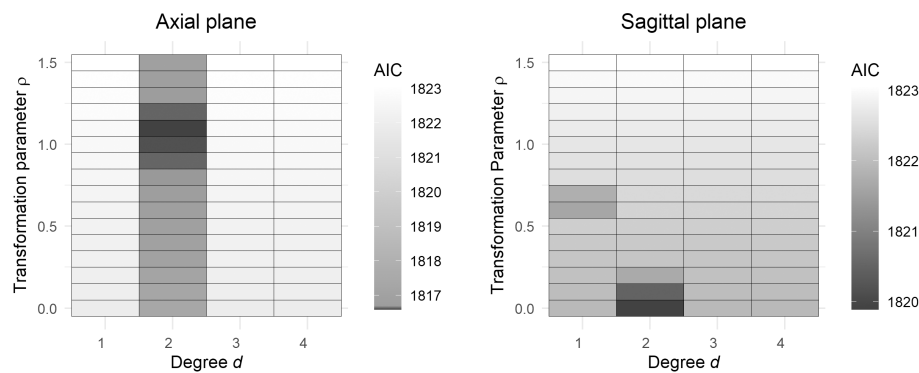


FIGURE 4 | The minimum value of AIC_λ achieved in each given pair of (d, ρ) for the axial and sagittal planes.

21,393 data points fall inside the boundary of the axial plane, constituting Y_{ij} 's, whereas there are $N_i = 16,078$ data points inside the boundary for the sagittal plane. We let gender and age-at-baseline be the unpenalized covariates \mathbf{Z} . We adopt the I-splines basis functions with degree 3 and 4 interior knots. We adopt the same set of initial parameter values for the EM algorithm as in the simulation study. The tolerance level for the convergence of the EM algorithm is set to be 10^{-4} . The degree of the Bernstein polynomial d and the transformation parameter ρ are typically unknown in practice. To achieve a parsimonious fit, we perform a two-dimensional grid search over the parameters (d, ρ) . Specifically, we vary ρ from 0 to 1.5 in increments of 0.1, and we consider d values from the set $\{1, 2, 3, 4\}$. For each pair (d, ρ) , we compute the minimum value of AIC_λ , where AIC_λ is defined in Equation (10) based on a grid of λ values ranging from 10^1 to $10^{2.5}$. The interval $(1, 2.5)$ is divided into 500 uniformly spaced partitions, serving as exponents with base 10. We then identify the optimal model by selecting the pair (d, ρ) that yields the smallest AIC value among all evaluated combinations. In terms of computational burden, the EM algorithm takes approximately 30 min for computing the 500 AIC_λ values for $d = 2$ and a given value of ρ in each of the planes using a standard desktop (2.4GHz, 16GB RAM) without parallel programming.

The results are illustrated in Figure 4. One can observe that the model with $d = 2$ and $\rho = 1.1$ achieves the lowest AIC value for

the axial plane, while the model with $d = 2$ and $\rho = 0$ achieves the lowest AIC value for the sagittal plane. This suggests that the PO and PH models should be applied to the data for axial and sagittal planes, respectively. The selected model identifies only one triangle associated with survival time in both planes. The selected triangles are marked by red dotted lines in the two panels of Figure 3, and they roughly correspond to the lateral ventricle region of the brain. The change in the size of the ventricle is known to be highly associated with impairment in cognitive function and dementia. Enlargement of the ventricle region has been identified as a reliable predictor for AD progression [4, 36].

To assess the model predictive performance of the proposed method, we use a 10-fold internal cross-validation procedure based on the integrated Brier score (IBS). The IBS, introduced by Graf et al. [37], is a prominent and widely adopted tool for gauging prediction error for right-censored data. Tsouprou et al. [38] later extended this method to accommodate interval-censored data. Specifically, it takes the form:

$$IBS = \frac{1}{n} \sum_{i=1}^n \frac{1}{t_{max}} \int_0^{t_{max}} \left\{ I(T_i > t) - \hat{S}_i(t | \tilde{X}_i, \mathbf{Z}_i) \right\}^2 dt,$$

where $\hat{S}_i(t | \tilde{X}_i, \mathbf{Z}_i)$ is the estimated survival function of the i th subject evaluated at time t . Clearly, $I(T_i > t) = 1$ for $t \leq L_i$ and $I(T_i > t) = 0$ for $t > R_i$. When $L_i < t \leq R_i$, $I(T_i > t)$ is unknown

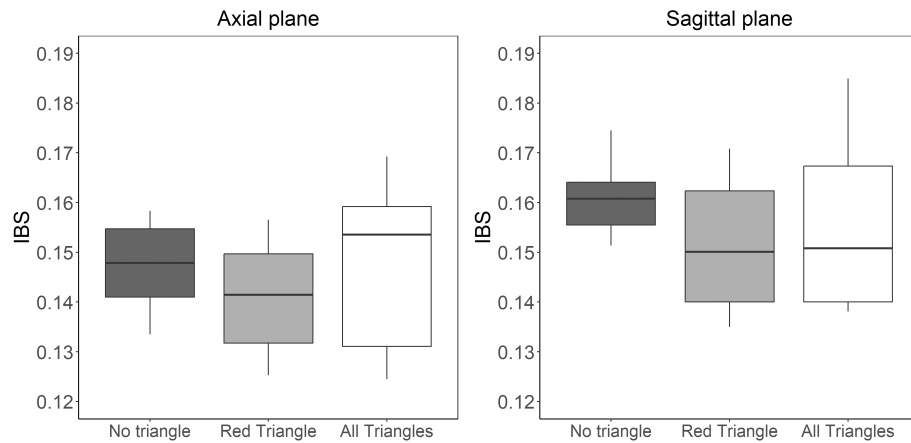


FIGURE 5 | The box plot for the IBS obtained in the 10-fold internal cross-validation based on the models (i) without any triangles; (ii) with the selected red triangle; and (iii) all triangles, over the axial and sagittal planes.

and is estimated by

$$\hat{I}(T_i > t) = \frac{\hat{S}_i(t|\tilde{X}_i, Z_i) - \hat{S}_i(R_i|\tilde{X}_i, Z_i)}{\hat{S}_i(L_i|\tilde{X}_i, Z_i) - \hat{S}_i(R_i|\tilde{X}_i, Z_i)}.$$

It is remarked that IBS ranges between 0 and 1, and a smaller value of this metric reflects a higher model predictive accuracy. In all models used for comparison, we always include the unpenalized covariates Z_i . For the axial plane, we fit the following three PO models (i) without any triangles, setting $\gamma = 0$ in Equation (5); (ii) with the selected triangle highlighted in red in Figure 3, based on a second-stage model, by allowing only γ^{r_i} to be nonzero for that specific triangle in Equation (4); and (iii) with all triangles, permitting all elements in γ to be nonzero in Equation (5). We perform an analogous model comparison for the sagittal plane based on the PH model. The results of the cross-validation are summarized by box plots in Figure 5. For the axial plane, the average IBS values for models (i), (ii), and (iii) are 0.147, 0.141, and 0.148, respectively. For the sagittal plane, the average IBS values for the three models are 0.161, 0.152, and 0.156, respectively. These results underscore the superior predictive performance of the proposed model with the selected triangle in red over the two planes, as evidenced by its minimal average IBS.

6 | Conclusions and Discussion

Brain MRI is pivotal in prognostic assessments for AD, yet integrating it into statistical analyses presents complex challenges. This study introduces an innovative methodology for pinpointing regions of interest within brain MRI scans that are indicative of AD onset. Specifically, the proposed method tackles the challenges posed by irregular boundaries in the MRI scans and the interval-censored nature of the data based on a flexible class of transformation models. By employing piecewise polynomial functions applied over triangulation, the brain images can be seamlessly integrated into the survival model and the region of interest can be subsequently identified within a regression-based framework, enhancing the precision of AD prognosis. The efficacy of the proposed methodology is demonstrated using the ADNI dataset. A specific brain region has been identified as significantly associated with AD onset. In both axial and sagittal

planes, the model with the selected region shows a smaller IBS (i.e., higher predictive accuracy) compared to the models without the imaging predictors and with all the imaging predictors. This finding offers valuable insights for clinical practitioners, especially in informing disease progression and risk stratification of patients.

The proposed method can be extended in the following ways. First, we utilize the brain MRI scans at baseline for risk prediction. However, the ADNI study has amassed longitudinal MRI scans with irregular revisiting time intervals, which presents an opportunity to enrich the predictive capability of the model. An advanced triangulation technique could be developed to capture features not only spatially across image pixels but also temporally across successive scans, potentially necessitating a three-dimensional stochastic modeling framework for sparse functional data. Second, the proposed approach focuses on modeling two-dimensional imaging data with irregular boundaries as predictors for survival outcomes. To further utilize the inherently three-dimensional imaging data with irregular boundaries, the current method could be extended using, for example, the trivariate spline smoothing techniques recently proposed by Wang et al. [39]. Third, the proposed approach adopts a two-stage estimation procedure. The initial stage involves approximating the image using Bernstein polynomials, followed by incorporating the approximated image into a survival model to perform group variable selection. While this method is computationally expedient, it is susceptible to the propagation of approximation errors from the first stage into the subsequent survival analysis. An alternative strategy could involve a joint modeling approach, wherein latent frailty terms are concurrently estimated, linking the imaging data directly with the latency component of the disease progression. Such methodological advancements would likely demand significantly increased computational resources.

Acknowledgments

Dr. Lee's research is substantially supported by the General Research Fund grant (Project No. 15301124) from the Research Grants Council of the Hong Kong Special Administrative Region, China. Dr. Shi's research is supported by the Discovery grant (RGPIN-2021-02963) from the Natural Sciences and Engineering Research Council of Canada (NSERC). Dr.

Cao's research is supported by the Discovery grant (RGPIN-2023-04057) from NSERC and the Canada Research Chair program.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The R codes associated with this paper are openly available on GitHub: <https://github.com/lcyjames/Roiico>.

References

1. J. Llamas-Rodríguez, J. Oltmer, D. N. Greve, et al., "Entorhinal Subfield Vulnerability to Neurofibrillary Tangles in Aging and the Preclinical Stage of Alzheimer's Disease," *Journal of Alzheimer's Disease* 87 (2022): 1379–1399.
2. G. A. Carlesimo, F. Piras, M. D. Orfei, M. Iorio, C. Caltaigirone, and G. Spalletta, "Atrophy of Presubiculum and Subiculum Is the Earliest Hippocampal Anatomical Marker of Alzheimer's Disease," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 1 (2015): 24–32.
3. M. Gonzalez-Rodriguez, S. Villar-Conde, V. Astillero-Lopez, et al., "Human Amygdala Involvement in Alzheimer's Disease Revealed by Stereological and Dia-PASEF Analysis," *Brain Pathology* 33 (2023): e13180, <https://doi.org/10.1111/bpa.13180>.
4. S. M. Nestor, R. Rupsingh, M. Borrie, et al., "Ventricular Enlargement as a Possible Measure of Alzheimer's Disease Progression Validated Using the Alzheimer's Disease Neuroimaging Initiative Database," *Brain* 131 (2008): 2443–2454, <https://doi.org/10.1093/brain/awn146>.
5. R. Leech and D. J. Sharp, "The Role of the Posterior Cingulate Cortex in Cognition and Disease," *Brain* 137 (2013): 12–32.
6. F. Rémy, N. Vayssière, L. Saint-Aubert, E. Barbeau, and J. Pariente, "White Matter Disruption at the Prodromal Stage of Alzheimer's Disease: Relationships With Hippocampal Atrophy and Episodic Memory Performance," *NeuroImage* 7 (2015): 482–492.
7. T. W. Schmitz and S. R. Nathan, "Basal Forebrain Degeneration Precedes and Predicts the Cortical Spread of Alzheimer's Pathology," *Nature Communications* 7 (2016): 13249.
8. J. Huang and J. A. Wellner, "Asymptotic Normality of the NPMLE of Linear Functionals for Interval Censored Data, Case 1," *Statistica Neerlandica* 49, no. 2 (1995): 153–163.
9. J. P. Klein, H. C. Van Houwelingen, J. G. Ibrahim, and T. H. Scheike, *Handbook of Survival Analysis* (CRC Press, 2016).
10. Z. Zhang and J. Sun, "Interval censoring," *Statistical Methods in Medical Research* 19, no. 1 (2010): 53–70.
11. S. N. Wood, M. V. Bravington, and S. L. Hedley, "Soap Film Smoothing," *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 70, no. 5 (2008): 931–955.
12. T. Ramsay, "Spline Smoothing Over Difficult Regions," *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 64, no. 2 (2002): 307–319.
13. M. J. Lai and L. L. Schumaker, *Spline Functions on Triangulations* (Cambridge University Press, 2007).
14. M. J. Lai and L. Wang, "Bivariate Penalized Splines for Regression," *Statistica Sinica* 23 (2013): 1399–1417.
15. V. Zippunikov, B. Caffo, D. M. Yousem, C. Davatzikos, B. S. Schwartz, and C. Crainiceanu, "Functional Principal Component Model for High-Dimensional Brain Imaging," *NeuroImage* 58, no. 3 (2011): 772–784.
16. P. T. Reiss and R. T. Ogden, "Functional Generalized Linear Models With Images as Predictors," *Biometrics* 66, no. 1 (2010): 61–69.
17. H. Zhou, L. Li, and H. Zhu, "Tensor Regression With Applications in Neuroimaging Data Analysis," *Journal of the American Statistical Association* 108, no. 502 (2013): 540–552.
18. X. Wang, H. Zhu, and A. D. N. Initiative, "Generalized Scalar-On-Image Regression Models via Total Variation," *Journal of the American Statistical Association* 112, no. 519 (2017): 1156–1168.
19. S. Jiang, J. Cao, B. Rosner, and G. A. Colditz, "Supervised Two-Dimensional Functional Principal Component Analysis With Time-To-Event Outcomes and Mammogram Imaging Data," *Biometrics* 79, no. 2 (2023): 1359–1369.
20. J. Goldsmith, L. Huang, and C. M. Crainiceanu, "Smooth Scalar-On-Image Regression via Spatial Bayesian Variable Selection," *Journal of Computational and Graphical Statistics* 23, no. 1 (2014): 46–64.
21. J. Kang, B. J. Reich, and A. M. Staicu, "Scalar-On-Image Regression via the Soft-Thresholded Gaussian Process," *Biometrika* 105, no. 1 (2018): 165–184.
22. X. Feng, T. Li, X. Song, and H. Zhu, "Bayesian Scalar on Image Regression With Nonignorable Nonresponse," *Journal of the American Statistical Association* 115, no. 532 (2020): 1574–1597.
23. T. B. Mattos, L. A. Matos, and V. H. Lachos, "A Semiparametric Mixed-Effects Model for Censored Longitudinal Data," *Statistical Methods in Medical Research* 30, no. 12 (2018): 2582–2603.
24. P. J. Green and B. W. Silverman, *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach* (CRC Press, 1993).
25. L. Wang, G. Wang, M. J. Lai, and L. Gao, "Efficient Estimation of Partially Linear Models for Data on Complicated Domains by Bivariate Penalized Splines Over Triangulations," *Statistica Sinica* 30, no. 1 (2020): 347–369.
26. L. Zhou and H. Pan, "Smoothing Noisy Data for Irregular Regions Using Penalized Bivariate Splines on Triangulations," *Computational Statistics* 29 (2014): 263–281.
27. M. Mao and J. L. Wang, "Semiparametric Efficient Estimation for a Class of Generalized Proportional Odds Cure Models," *Journal of the American Statistical Association* 105, no. 489 (2010): 302–311.
28. J. O. Ramsay, "Monotone Regression Splines in Action," *Statistical Science* 3, no. 4 (1988): 425–441.
29. M. Yuan and Y. Lin, "Model Selection and Estimation in Regression With Grouped Variables," *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 68, no. 1 (2006): 49–67.
30. L. Meier, S. Van De Geer, and P. Bühlmann, "The Group Lasso for Logistic Regression," *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 70, no. 1 (2008): 53–71.
31. D. Zeng, L. Mao, and D. Y. Lin, "Maximum Likelihood Estimation for Semiparametric Transformation Models With Interval-Censored Data," *Biometrika* 103, no. 2 (2016): 253–271.
32. P. Tseng, "Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization," *Journal of Optimization Theory and Applications* 109 (2001): 475–494.
33. J. Huang, L. Liu, Y. Liu, and X. Zhao, "Group Selection in the Cox Model With a Diverging Number of Covariates," *Statistica Sinica* 24 (2014): 1787–1810.
34. C. R. Jack, M. A. Bernstein, N. C. Fox, et al., "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI Methods," *Journal of Magnetic Resonance Imaging* 27, no. 4 (2008): 685–691.
35. D. Ma, K. Popuri, M. Bhalla, et al., "Quantitative Assessment of Field Strength, Total Intracranial Volume, Sex, and Age Effects on the Goodness of Harmonization for Volumetric Analysis on the ADNI Database," *Human Brain Mapping* 40, no. 5 (2019): 1507–1527.

36. S. H. Guptha, E. Holroyd, and G. Campbell, "Progressive Lateral Ventricular Enlargement as a Clue to Alzheimer's Disease," *Lancet* 359 (2002): 2040, [https://doi.org/10.1016/S0140-6736\(02\)08806-2](https://doi.org/10.1016/S0140-6736(02)08806-2).

37. E. Graf, C. Schmoor, W. Sauerbrei, and M. Schumacher, "Assessment and Comparison of Prognostic Classification Schemes for Survival Data," *Statistics in Medicine* 18, no. 17–18 (1999): 2529–2545.

38. S. Tsouprou, H. Putter, and M. Fiocco, "Measures of Discrimination and Predictive Accuracy for Interval Censored Survival Data," (PhD thesis, Leiden University, 2015).

39. G. Wang, Y. Wang, A. S. Gao, L. Wang, and A. D. N. Initiative, "Efficient Nonparametric Estimation of 3D Point Cloud Signals Through Distributed Learning," *Journal of Computational and Graphical Statistics* 34 (2024): 1–13.

Appendix A

Conditional Expectations of the Latent Variables

Note that $E(A_{ij}|A_i = 0, \mathcal{O}_i, \theta^{(d)}) = 0$ and $(A_{ij}|A_i > 0, \mathcal{O}_i, \theta^{(d)}) \sim \text{Binomial}(A_i, \mu_{ij}/\sum_{l=1}^J \mu_{il})$. We have

$$\begin{aligned} \hat{E}(A_{ij}) &\equiv E(A_{ij}|\mathcal{O}_i, \theta^{(d)}) = E\{E(A_{ij}|A_i > 0, \mathcal{O}_i, \theta^{(d)})|\mathcal{O}_i, \theta^{(d)}\} \\ &= E\left(\frac{\mu_{ij}A_i}{\sum_{l=1}^J \mu_{il}} \middle| A_i > 0, \mathcal{O}_i, \theta^{(d)}\right) \\ &= \frac{\mu_{ij}}{\sum_{l=1}^J \mu_{il}} E(A_i | A_i > 0, \mathcal{O}_i, \theta^{(d)}) \\ &= \frac{\mu_{ij}}{\sum_{l=1}^J \mu_{il}} E\{E(A_i | A_i > 0, \zeta_i, \mathcal{O}_i, \theta^{(d)})|\mathcal{O}_i, \theta^{(d)}\} \\ &= \frac{\mu_{ij}}{\sum_{l=1}^J \mu_{il}} E\left\{\frac{\zeta_i \sum_{l=1}^J \mu_{il}}{1 - e^{-\zeta_i \sum_{l=1}^J \mu_{il}}} \middle| \mathcal{O}_i, \theta^{(d)}\right\} \\ &= E\left\{\frac{\zeta_i \mu_{ij}}{1 - e^{-\zeta_i \sum_{l=1}^J \mu_{il}}} \middle| \mathcal{O}_i, \theta^{(d)}\right\}, \end{aligned}$$

where the second last line refers to the mean of a zero-truncated Poisson distribution. Similarly, we obtain

$$\hat{E}(B_{ij}) \equiv E(B_{ij}|\mathcal{O}_i, \theta^{(d)}) = E\left\{\frac{\zeta_i \eta_{ij}}{1 - e^{-\zeta_i \sum_{l=1}^J \eta_{il}}} \middle| \mathcal{O}_i, \theta^{(d)}\right\}.$$

Thus, both $\hat{E}(A_{ij})$ and $\hat{E}(B_{ij})$ can be expressed analytically as the conditional expectation of a function of ζ_i . Specifically, the conditional expectation of any function g of ζ_i given the observed data is

$$E\{g(\zeta_i)|\mathcal{O}_i\} = \frac{\int g(\zeta) h_i(\zeta) d\zeta}{\int h_i(\zeta) d\zeta},$$

where

$$\begin{aligned} h_i(\zeta_i) &\equiv \left\{1 - e^{-\sum_{j=1}^J \zeta_i \omega_j I_j(R_i)} e^{\beta^T Z_i + \gamma^T \xi_i}\right\}^{\Delta_{Li}} \left\{e^{-\sum_{j=1}^J \zeta_i \omega_j I_j(L_i)} e^{\beta^T Z_i + \gamma^T \xi_i}\right. \\ &\quad \left.- e^{-\sum_{j=1}^J \zeta_i \omega_j I_j(R_i)} e^{\beta^T Z_i + \gamma^T \xi_i}\right\}^{\Delta_{Ri}} f_{\zeta}(\zeta_i; \rho) \end{aligned}$$

is proportional to the conditional density of ζ_i given \mathcal{O}_i . For notational simplicity below, let $S_{Li} = \tilde{\lambda}(L_i) e^{\beta^T Z_i + \gamma^T \xi_i}$ and $S_{Ri} = \tilde{\lambda}(R_i) e^{\beta^T Z_i + \gamma^T \xi_i}$. Note that $\hat{E}(A_{ij})$ is nonzero only if $\Delta_{Li} = 1$, in which

$$\hat{E}(A_{ij}) = \frac{\mu_{ij}}{e^{-G(S_{Li}; \rho)} - e^{-G(S_{Ri}; \rho)}} = \frac{\mu_{ij}}{1 - e^{-G(S_{Ri}; \rho)}}.$$

In the same vein, $\hat{E}(B_{ij})$ is nonzero only if $\Delta_{Ri} = 1$, in which

$$\hat{E}(B_{ij}) = \frac{\eta_{ij} e^{-G(S_{Li}; \rho)} G'(S_{Li}; \rho)}{e^{-G(S_{Li}; \rho)} - e^{-G(S_{Ri}; \rho)}},$$

where G' denotes the first-order derivative of G . Finally, we can also show that

$$\hat{E}(\zeta_i) = \frac{e^{-G(S_{Li}; \rho)} G'(S_{Li}; \rho) - e^{-G(S_{Ri}; \rho)} G'(S_{Ri}; \rho)}{e^{-G(S_{Li}; \rho)} - e^{-G(S_{Ri}; \rho)}}.$$

Therefore, all the conditional expectations in the E-step have closed-form expressions where numerical integration is not required, yielding a computationally efficient EM algorithm.