

A Bayesian 2D functional linear model for gray-level co-occurrence matrices in texture analysis of lower grade gliomas

Thierry Chekouo^{a,*}, Shariq Mohammed^{b,c,*}, Arvind Rao^{b,c,d,e,*}

^a Department of Mathematics and Statistics, University of Calgary, Canada

^b Department of Biostatistics, University of Michigan, USA

^c Department of Computational Medicine & Bioinformatics, University of Michigan, USA

^d Department of Biomedical Engineering, University of Michigan, USA

^e Department of Radiation Oncology, University of Michigan, USA

ARTICLE INFO

Keywords:

Bayesian variable selection
2D functional data
GLCM
Imaging sequences
Lower grade glioma
Texture analysis

ABSTRACT

In cancer radiomics, textural features evaluated from image intensity-derived gray-level co-occurrence matrices (GLCMs) have been studied to evaluate gray-level spatial dependence within the regions of interest in the brain. Most of these analysis work with summary statistics (or texture-based features) constructed using the GLCM entries, and potentially overlook other structural properties in the GLCM. In our proposed Bayesian framework, we treat each GLCM as a realization of a two-dimensional stochastic functional process observed with error at discrete time points. The latent process is then combined with the outcome model to evaluate the prediction performance. We use simulation studies to assess the performance of our method and apply it to data collected from individuals with lower grade gliomas. We found our approach to outperform competing methods that use only summary statistics to predict isocitrate dehydrogenase (IDH) mutation status.

1. Introduction

In oncology, characterization of the tumor phenotype is a challenge owing to the heterogeneous tumor micro-environments and diverse clinical pathways. Cancer radiomics has emerged as an area to address such challenges. It deals with quantifying the intrinsic patterns in gray-level intensity patterns, texture, enhancement, morphology and shape of the tumor tissues. Recent advances include different types of quantitative assessments that are facilitated through dimension reduction by constructing features from delineated areas (or regions of interest). These features could potentially guide the understanding of the underlying composition of tissues which are suspected to have malignant pathology. Consequently, these image-derived features are expected to capture the underlying patterns indistinguishable to the human eye and facilitate down-stream analyses.

There has been increased interest in understanding the associations between clinical features, molecular pathology and textural features based on the imaging data (Zulpe and Pawar, 2012; Kumar et al., 2015; Narang et al., 2017). Specifically, textural features evaluated using gray-level co-occurrence matrices (GLCMs) have been used to study gray-level spatial dependence within the delineated regions of interest.

GLCMs are matrices defined over an image domain with entries as counts describing the spatial distribution of co-occurring gray-scale values (Haralick et al., 1973). Most of the analyses involving GLCMs work with the texture-based features (or summary statistics) rather than evaluating the GLCM in its entirety. In this paper, we propose a Bayesian probabilistic modeling framework to identify the associations between a phenotype and the GLCMs constructed based on Magnetic Resonance Imaging (MRI) scans. We specifically focus on lower grade gliomas (LGG), which is a specific type of brain cancer. The proposed framework facilitates modeling the GLCM within a two-dimensional functional space to capture the spatial dependencies in the gray-levels, thereby extending traditional analysis based on summary statistics or textural features. This modeling can be utilized under a classification or regression setting corresponding to any relevant clinical phenotype.

1.1. Lower grade gliomas

Lower grade glioma is a neoplastic disease that develops in the glial cells of the brain. In general, gliomas are graded from I to IV according to the World Health Organization criteria. Tumors grades II and III are usually classified as LGG, as opposed to the more common higher grade

* Corresponding authors at: Department of Mathematics and Statistics, University of Calgary, Canada (T. Chekouo). Department of Biostatistics, University of Michigan, USA (S. Mohammed and A. Rao).

E-mail addresses: thierry.chekouotekou@ucalgary.ca (T. Chekouo), shariqm@umich.edu (S. Mohammed), ukarvind@umich.edu (A. Rao).

<https://doi.org/10.1016/j.nicl.2020.102437>

Received 10 June 2020; Received in revised form 21 August 2020; Accepted 14 September 2020

Available online 18 September 2020

2213-1582/ © 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

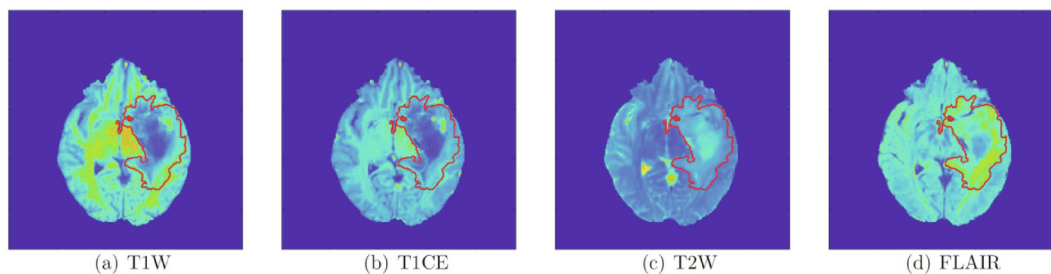


Fig. 1. Axial slice of a brain MRI for a LGG subject shown for four imaging sequences T1W, T1CE, T2W and FLAIR. The segmented tumor region is shown with the (red) boundary overlaid on the images. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tumors, also known as glioblastoma multiforme. We primarily work with The Cancer Genome Atlas (TCGA) (Tomczak et al., 2015) cohort for LGG data which involves 108 cases with the tumor histology as astrocytoma, oligodendroglioma, or oligoastrocytoma. This histologic data and the MRI acquisition protocols can be found in the publicly available data files of patients (Pedano et al., 2016). Patients in this cohort had undergone routine MRI prior to surgery and treatment. These MRI images are acquired in four different sequences: Pre-contrast T1-weighted (T1W), post-gadolinium T1 contrast enhanced (T1CE), T2-weighted (T2W), and T2 fluid-attenuated inversion recovery (FLAIR). We show all four MRI sequences corresponding to an axial slice for an LGG patient in Figs. 1(a-d). The patient image data, the tumor segmentation labels and the corresponding clinical data were extracted from The Cancer Imaging Archive (TCIA) (Clark et al., 2013; Kuthuru et al., 2018) and TCGA. The tumor area is shown as a delineated region on the image for all four sequences in Figs. 1(a-d). However, the whole brain MRIs are three-dimensional data objects with stacked axial layers corresponding to each subject. Voxel intensity values are sensitive to the MRI scanner configuration and are difficult to interpret and compare either between study visits within a single subject, or across different subjects. This emphasizes the necessity of pre-processing in terms of intensity value normalization, which we address by performing a biologically motivated normalization technique using the R package *WhiteStripe* (Shinohara et al., 2014). For consistency across all the subjects and scans, the normalization was implemented under the same default settings in the R package *WhiteStripe*.

1.1.1. Molecular Sub-types in LGG

Several molecular alterations have been shown to be associated with the overall survival of the LGG patients. These include isocitrate dehydrogenase (IDH1/2) mutation status (collectively referred to as IDH mutations), 1p19q chromosomal arm codeletion, MGMT promoter methylation status, and TP53 mutation (Leu et al., 2013; Moritz-Gasser and Herbet, 2013; Pignatti et al., 2002). Multiple molecular sub-types of LGG have been identified (Eckel-Passow et al., 2015; Network, 2015), which predominantly include IDH wild-type, IDH mutant with 1p19q codeleted, and IDH mutant with 1p19q noncodeleted groups. These groupings were used to show that LGGs with IDH wild-type have molecular characteristics and behavior similar to glioblastoma and have been associated with shorter survival. Specifically in the case of glioblastoma, studies also indicate that patients with IDH1 mutation have better prognosis compared to the patients with wild-type (Hartmann et al., 2010). Since the IDH status is such an important aspect in case of gliomas, in this paper we aim to assess the associations between the imaging measurements and the IDH status in the case of LGGs. Prediction of molecular status (e.g. IDH status) in gliomas using radiomic features is an extensively studied problem in the field of radiomics (Chaddad et al., 2015; Hsieh et al., 2017; Li et al., 2017; Chang et al., 2018; Jakola et al., 2018; Li et al., 2018; Rathore et al., 2018; Tian et al., 2018; Han et al., 2019; Liu et al., 2019).

1.2. Gray-level co-occurrence matrices

A predominant approach in image texture analysis to evaluate gray-level spatial dependencies is through the analysis of GLCM. A GLCM is a matrix defined over an image domain with its entries computed as counts, describing the spatial distribution of co-occurring gray-scale values in the image voxels at a given spatial offset and orientation/angle (Haralick et al., 1973). The entry (j, k) in a GLCM is the total number of pixel pairs that have gray-levels j and k in the image in a specified direction and distance. Consider a two-dimensional image (e.g. an axial slice from the MRI), the (j, k) -th entry in the corresponding GLCM represents the number of instances that pixel pairs with gray-levels j and k are neighbors either horizontally, vertically or diagonally. The GLCM entries can be computed by traversing pixel-by-pixel across the whole image or region of interest. That is, for a given pixel (with gray-level j), we consider its neighbors at 0° , 45° , 90° and 135° at a distance one (immediate neighboring pixel). For each of these four neighbors, if its gray-level is k then the (j, k) -th entry of the GLCM is incremented by one, and this leads to a non-symmetric GLCM. However, adding one to both (j, k) -th and (k, j) -th entries creates a symmetric GLCM. Symmetric GLCMs can also be constructed by looking at all eight directions for each pixel. For a three-dimensional image the number of directions increase to thirteen as we have an additional dimension. GLCMs constructed by considering all the directions are symmetric and rotation invariant. GLCMs can also be constructed by looking for pixel pairs at different distances/scales and not necessarily constrained to immediate voxel neighbors.

For each patient and imaging sequence, a GLCM is computed from the three-dimensional tumor volume observed in the corresponding sequence. We consider symmetric as well as non-symmetric GLCMs from T gray-level bins. The GLCM object thus constructed provides a huge dimension reduction mapping from the image space to the data lattice. We choose $T = 8$ which is a common choice in applications involving GLCMs. However, we would like to emphasize that our approach in this paper can be readily employed to various choices of the gray-levels. The GLCMs also facilitate further statistical modeling and analysis in diagnostic oncology, which is otherwise hindered due to irregular shape and sizes of the tumor in the original image space. Current modeling approaches using GLCM focus on deriving textural features, which are summary statistics constructed using the GLCM entries (Hitam et al., 2003; Lee et al., 2015; Ghosh et al., 2015). Many studies have also reported high correlation between the derived textural features which leads to over-fitting (Albregtsen et al., 2008; Zhang et al., 2008; Kassner and Thornhill, 2010). These textural features also potentially overlook other structural properties in the GLCM. Some of the most commonly used textural features based on the GLCM are contrast, correlation, energy, entropy and homogeneity (definitions below). These features are interpreted in their conventional sense as summary statistics and are referred to as the Haralick features.

$$\begin{aligned} \text{Contrast} &= \sum_{j,k} |j - k|^2 p_{jk}, \\ \text{Correlation} &= \sum_{j,k} \frac{(j - \mu_j)(k - \mu_k) p_{jk}}{\sigma_j \sigma_k}, \\ \text{Energy} &= \sum_{j,k} p_{jk}^2, \\ \text{Entropy} &= \sum_{j,k} p_{jk} \log p_{jk}, \\ \text{Homogeneity} &= \sum_{j,k} \frac{p_{jk}}{1 + |j - k|^2}, \end{aligned}$$

where p_{jk} is the (j, k) -th entry in the GLCM, and μ_j and μ_k are the marginal means of row j and column k , respectively. Similarly, σ_j and σ_k are the marginal standard deviations of row j and column k , respectively.

1.3. Statistical framework

In this work, we propose a novel Bayesian approach using a generalized two-dimensional (2D) functional linear models to model the IDH status using the GLCMs. Our framework builds nonlinear association models between scalar responses and functional predictors. To utilize the gray-level spatial structure in the GLCM, we use it as our fundamental data object rather than any texture features derived from them. Li et al., 2019 is one of the very few studies that uses the whole GLCM (instead of the textural features) to classify adrenal lesions through spatial Bayesian modeling. The data object in their study is a structured lattice whereas we consider the GLCM as a two-dimensional functional object. We treat the GLCM counts as 2D functional data for each imaging sequence, and account for both symmetric and non-symmetric GLCMs. Our method is part of the functional prediction regression (FPR) family of models which relate a non-functional response (e.g. IDH status) to functional predictors (e.g. GLCMs). There is an extensive literature on FPR models (Baladandayuthapan et al., 2015; James, 2002; Ramsay and Dalzell, 1991; Morris, 2015; Vannucci et al., 2003; Vannucci et al., 2005; Ferraty and Vieu, 2003; Cardot et al., 1999; Cardot et al., 2003; Ferraty and Vieu, 2002; Muller, 2005). A nice review on functional regression can be found in Morris, 2015. Most of previous approaches to modeling functional regression effects have focused on one-dimensional smoothing, but fewer studies have focused on multidimensional smoothing used mostly as the tensor product of basis of each dimension (Eilers and Marx, 2003; Marx et al., 2011; Marx and Eilers, 2005). Inspired by the work of Chen et al., 2017, the representation of our two-dimensional functional data is based on a tensor product of eigenfunctions of marginal kernels. As this representation has a simple interpretation, we newly adopt it in the context of Bayesian FPR models. Moreover, we also account for measurement error in the GLCMs, and provide regularization using mixture prior distributions.

The manuscript is organized as follows. In Section 2, we illustrate our proposed Bayesian modeling approach. In Section 3, we discuss posterior inference and show how our modeling framework can be used to predict the IDH status. In Section 4, we present results from simulation studies in which we found our approach to outperform competing methods that use summary statistics. In Section 5, we discuss the results of our inference on the data set of TCGA LGG patients. We conclude in Section 6 with a discussion and some directions for future work.

2. A Bayesian 2D functional linear model

In this study, we consider imaging and clinical data from n TCGA LGG patients who had all undergone routine MRI prior to surgery and LGG treatment. For the i th patient, we observe $y_i = 1$ if IDH is mutant and 0

otherwise. Moreover, for each patient i , we observe (log-) GLCM count $x_{ir}(s_j, t_k)$ for imaging sequence $r = 1, \dots, R$ with gray-levels s_j and t_k for every $(j, k) \in \{1, 2, \dots, T\}$, where T is the total number of gray-level intensities. For both simulated and LGG data, we used the Z-score standardization for $x_{ir}(s_j, t_k)$ for each entry (s_j, t_k) and imaging sequence r . In our LGG case study, we consider $T = 8$ gray-levels and $R = 4$ imaging sequences: T1W, T1CE, T2W and FLAIR as described in Section 1.2.

2.1. Theoretical motivation

Let $\mu(s, t) := E(x(s, t))$ be the mean of the process $x(s, t)$ for all $(s, t) \in \mathcal{S}^2$ where $\mathcal{S} \subset \mathbb{R}$. Under fairly general assumptions, Chen et al., 2017 showed that the two dimensional process $x(s, t)$ can be written as

$$x(s, t) \approx \mu(s, t) + \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \xi_{kl} \phi_k(s) \psi_l(t) \equiv z(s, t) \quad (1)$$

where the $\{\phi_k(s)\}$'s, $k \geq 1$ and $\{\psi_l(t)\}$'s, $l \geq 1$ are respectively the eigenfunctions of the marginal kernels $K_1(s, s') = \int_T K[(s, t), (s', t)] dt$ and $K_2(t, t') = \int_T K[(s, t), (s, t')] ds$, $K[(s, t), (s', t')]$ is the covariance function of $x(s, t)$, and ξ_{kl} is the principal component score associated with the pair of eigenfunctions (ϕ_k, ψ_l) such that $\text{cov}(\xi_{kl}, \xi_{k'l'}) = 0$ for $l \neq l'$ and $\text{cov}(\xi_{kl}, \xi_{k'l}) = 0$ for $k \neq k'$. The representation of $x(s, t)$ in Eq. (1) can be seen as an extension of the Karhunen-Loève expansion (Loève, 1963), and is well suited for situations where the two arguments of $x(s, t)$ play symmetric roles. From Eq. (1), we can show that $\{\phi_k(s)\}$'s, $k \geq 1$ and $\{\psi_l(t)\}$'s, $l \geq 1$ are respectively eigenfunctions of $\text{cov}(X_1(s), X_1(s'))$ and $\text{cov}(X_2(t), X_2(t'))$, where $X_1(s) = \int_{\mathcal{S}} x(s, t) dt$ and $X_2(t) = \int_{\mathcal{S}} x(s, t) ds$ are the marginal functions of $x(s, t)$ (see Theorem 6.1 of the Supplementary Materials). Hence, those eigenfunctions can be estimated after performing standard functional principal component analysis on the marginal stochastic processes $X_1(s)$ and $X_2(t)$.

2.2. Model formulation

We treat the non-symmetric GLCM counts from each sequence r as 2D functional data indexed by pairs of gray-levels. We then model the (log-) GLCM counts as follows

$$x_{ir}(s, t) = \mu_r(s, t) + z_{ir}(s, t) + \epsilon_{ir}(s, t), \quad (2)$$

where

$$z_{ir}(s, t) = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \xi_{ikl}^{(r)} \phi_k^{(r)}(s) \psi_l^{(r)}(t), \quad (3)$$

$\epsilon_{ir}(s, t)$ is a mean-zero white-noise process with variance σ_r^2 . Hence we assume that for subject i , the GLCM counts $x_{ir}(s, t)$ are noisy realizations of the true latent process $z_{ir}(s, t)$ at intensity level (s, t) . When GLCMs are symmetric, we then assume that $\phi_k^{(r)}(t) = \psi_l^{(r)}(t)$, and $\xi_{ikl}^{(r)} = \xi_{ilk}^{(r)}$. Hence the indices of the double summation in Eq. (3) would be restricted to belong to the set $\{k, l \in \mathbb{N}, k \leq l\}$ and the off-diagonal terms $\xi_{ikl}^{(r)}$'s ($k < l$) are multiplied by two. We use y_i^* ($i = 1, \dots, n$) to denote a continuous clinical outcome for patient i . To illustrate the main ideas, we describe the simplest case with continuous outcome data and extend the method for binary outcome later. In order to relate imaging with clinical outcome, we assume a functional linear regression model with a scalar response defined as

$$y_i^* = \sum_{r=1}^R \int_{\mathcal{S}^2} z_{ir}(s, t) \beta_r(s, t) ds dt + \mathbf{w}_i^T \mathbf{b} + e_i, \quad e_i \sim_{i.i.d} \mathcal{N}(0, \sigma_e^2). \quad (4)$$

where $\beta_r(s, t)$ is the continuous functional regression coefficient associated with GLCM for imaging sequence r , and \mathbf{w}_i is a Q -dimensional vector that encodes for nonfunctional fixed effects (confounders such as demographic variables: gender, age, etc) with \mathbf{b} as the associated regression coefficients. Here we take the first component of \mathbf{W} to be 1 to

include the intercept term. Using Eq. (3), Eq. (4) can be written as

$$y_i^\star = \sum_{r=1}^R \sum_{k,l} \xi_{ikl}^{(r)} \beta_{kl}^{(r)} + \mathbf{w}_i^T \mathbf{b} + e_i. \quad (5)$$

where

$$\beta_{kl}^{(r)} = \int_{\mathcal{S}} \beta_r(s, t) \phi_k^{(r)}(s) \psi_l^{(r)}(t) ds dt. \quad (6)$$

Under the assumption that $\beta_r(s, t)$ is defined in the tensor Hilbert space $L^2(\mathcal{S}) \otimes L^2(\mathcal{S})$ (Takesaki, 2001; Bratteli and Robinson, 2003), $\beta_r(s, t)$ can be written as a linear combination of the orthonormal system $\{\phi_k^{(r)}(s) \psi_l^{(r)}(t), k, l \geq 1\}$ using the representation

$$\beta_r(s, t) = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \beta_{kl}^{(r)} \phi_k^{(r)}(s) \psi_l^{(r)}(t) \quad (7)$$

The basis functions $\{\phi_k^{(r)}(s), k \geq 1\}$ and $\{\psi_l^{(r)}(t), l \geq 1\}$ are not known, hence we use functional principal component (FPC) analysis to construct their estimators (Ramsay and Silverman, 2005; Hall et al., 2006; Yao et al., 2005) by using respectively the marginal functions $X_1(s)$ and $X_2(t)$. Once the orthonormal basis coefficients or the FPC scores have been estimated, we can reduce (5) by applying truncated approximations in (3), which gives

$$y_i^\star = \sum_{r=1}^R \sum_{k=1}^{K_{r1}} \sum_{l=1}^{K_{r2}} \xi_{ikl}^{(r)} \beta_{kl}^{(r)} + \mathbf{w}_i^T \mathbf{b} + e_i. \quad (8)$$

where K_{r1} and K_{r2} are the truncation parameters for the predictors from the r th imaging sequence. Again, when GLCMs are symmetric, $K_{r1} = K_{r2}$ and $\beta_{kl}^{(r)} = \beta_{lk}^{(r)}$. Hence, the indices of the double summation k and l in Eq. (8) belong now to $\{k, l \in \{1, \dots, K_{r1}\}, k \leq l\}$ and the off-diagonal regression coefficients $\beta_{lk}^{(r)}$ ($k < l$) are also multiplied by two. For both simulated and observed data, we used a fraction of variance explained (FVE) of at least 95% in order to estimate K_{r1} and K_{r2} . Alternatively, we may formulate it as a model selection procedure and choose them by using some model selection criterion, such as the deviance information criterion (DIC).

For the case of binary response (IDH status) we consider a probit model that linearly relates GLCMs to the binary response. That is, we adopt a data augmentation approach (Albert and Chib, 1993) to define our predictive model as in (8) where

$$y_i = \begin{cases} 1 & \text{if } y_i^\star > 0, \\ 0 & \text{otherwise,} \end{cases} \quad (9)$$

and $y_i = 1$ if subject i has a mutant IDH and $y_i = 0$ if subject i has a wild type IDH. In this context, the variance of the latent continuous variable y_i^\star is $\sigma_e^2 = 1$ (Albert and Chib, 1993).

2.3. Prior Distributions

Variable selection is an important problem in functional regression model and we aim to identify both imaging sequences and their corresponding components that contribute the most to the prediction of the response. As commonly done in Bayesian variable selection, we define the indicator vector $\gamma^{(r)} = (\gamma_{kl}^{(r)})_{kl}$ where $\gamma_{kl}^{(r)} = 1$ if component (k, l) for imaging sequence r is selected and 0 otherwise. We then define adaptively regularized representations of the effect functions $\beta_r(s, t)$ by placing a mixture prior on $\beta_{kl}^{(r)}$, the coefficient effect of latent scores $\xi_{ikl}^{(r)} = (\xi_{ikl}^{(r)})_i$

$$\pi(\beta^{(r)}) = \prod_{k,l} [(1 - \gamma_{kl}^{(r)}) \delta_0(\beta_{kl}^{(r)}) + \gamma_{kl}^{(r)} \text{Normal}(\beta_{kl}^{(r)}; 0, \sigma_e^2 h_r)] \quad (10)$$

for every $r \in \{1, \dots, R\}$. In Eq. (10), δ_0 is the Dirac delta function concentrated at 0, and $\text{Normal}(x; \mu, \sigma)$ denotes the probability density function of a normal distribution with mean μ and variance σ^2 . It follows from prior (10) that $\beta_{kl}^{(r)} = 0$ if $\gamma_{kl}^{(r)} = 0$ and $\beta_{kl}^{(r)} \neq 0$ if $\gamma_{kl}^{(r)} = 1$. Let us define $A_{kl} = \{r | \gamma_{kl}^{(r)} = 1\}$ and assume that

$\text{Cov}(\beta_{kl}^{(r)}, \beta_{k'l'}^{(r)}) = \sigma_e^2 \sigma_{rr'} = \sigma_e^2 \sigma_{r'r} > 0$ for every $r \in A_{kl}$, $r' \in A_{k'l'}$ ($r \neq r'$), $l, l' \in \{1, \dots, K_{r1}\}$ and $k, k' \in \{1, \dots, K_{r2}\}$. The shrinkage hyper-parameter h_r in Eq. (10) is defined as $h_r = h + \sum_{r' \neq r} \sum_{k,l} \gamma_{kl}^{(r')} \sigma_{rr'}$ ($h > 0$, fixed) such as the covariance matrix of $\beta = \{\beta_{kl}^{(r)}, k, l, r \in A_{kl}\}$ is diagonally dominant (and symmetric), hence, it is guaranteed to be positive definite.

Instead of defining a mixture prior for the coefficient effects $\beta_{kl}^{(r)}$'s (see Eq. 10), we can simply assume that $\beta_{kl}^{(r)}$ follows $\text{Normal}(0, \sigma_e^2 h_r)$ i.e. we do not impose sparsity on the set of components. This approach can be valid when the number of components is less than the number of subjects but the computational cost of its MCMC algorithm is higher than the sparsity approach. We made a comparison of the two approaches (sparsity and no sparsity) in terms of computational cost and component selection in the Supplementary Materials (see Section 7). We have also added the option to run the non-sparsity approach to our MCMC algorithm.

Under our model formulation, $\sigma_{rr'}$ controls the overall correlation between the regression coefficient functions $\beta_r(s, t)$ and $\beta_{r'}(s, t)$ if both imaging sequences r and r' are associated with \mathbf{y} . Otherwise, there is no correlation. This assumption is motivated by the fact that the imaging sequences provide information about the same region and also seem to be positively correlated, hence we would expect their effects to be correlated as well. Moreover, in the case where one imaging sequence is associated and the other are not, the method can learn from the data and impose a partial correlation, thereby enabling the effects of other imaging sequences. Fig. 1 of the Supplementary Materials presents empirical correlations between sequences of the observed GLCMs. It shows that most of the empirical correlations are positive. For example, we observe stronger pairwise correlations between FLAIR and T2W, possibly as the tumor region is indicated by a mass-like hyperintense signal in both sequences (Bruno et al., 2019).

We assume independent binomial prior distributions with probability p_r for the binary parameter $\gamma_{kl}^{(r)}$. We also impose gamma prior distributions on the covariance $\sigma_{rr'}$, $r, r' \in \cup_{k,l} A_{kl}$ with shape parameter a and rate parameter b . If the outcome is binary, it is assumed that the variance of the error term e_i (or the latent variable y_i^\star), σ_e^2 is set to 1 (Albert and Chib, 1993). Otherwise, if it's continuous, we assume a conjugate inverse gamma prior for σ_e^2 with parameters a_e and b_e .

The latent scores $\xi_{ikl}^{(r)}$'s are assumed to be normally distributed $N(0, \lambda_{kl}^{(r)})$, as is commonly assumed in functional principal component analysis (Crainiceanu and Goldsmith, 2010), where $\lambda_{kl}^{(r)}$ follows an inverse gamma distribution with parameters a_l and b_l , and represents the eigenvalues for each imaging sequence r . We finally impose a conjugate inverse gamma prior for σ_r^2 with parameters a_r and b_r , and independent normal $N(0, h_0)$ for components of the coefficient vector $\mathbf{b} = (b_1, \dots, b_Q)$ for nonfunctional features.

3. Posterior inference

For posterior inference, our primary interest is in the estimation of the selection indicator $\gamma_{kl}^{(r)}$ and regression effects of the latent scores $\xi_{ikl}^{(r)}$'s. We implement a Markov Chain Monte Carlo (MCMC) algorithm for posterior inference that employs partially collapsed Gibbs sampling (van and Park, 2008) to sample $\gamma_{kl}^{(r)}$ and the regression coefficients $\beta_{kl}^{(r)}$. In fact, for each sequence r , to sample the binary matrix $(\gamma_{kl}^{(r)})_{kl}$ given $(\gamma_{kl}^{(r')})_{kl}$ ($r' \neq r$), we integrate out first $\beta_{kl}^{(r)}$. Then, we sample the regression coefficients $\beta_{kl}^{(r)}$ from their full conditional distributions. In Section 8 of the Supplementary Materials, we present the full conditionals to update all parameters. We outline the steps of our MCMC algorithm below.

1. Sample $\gamma^{(r)}$ from its full conditionals as defined in Eq. (5) of the Supplementary Materials.
2. Sample $\beta^{(r)}$ from its full conditionals as defined in Eq. (6) of the Supplementary Materials.
3. Sample $\sigma_{rr'}$ from its posterior full conditional using a Metropolis-

- Hasting step as described in Section 8 (Supplementary Materials).
4. Sample the parameters $\xi_{ikl}^{(r)}$, $\lambda_{kl}^{(r)}$, σ_e^2 , σ_r^2 , from their full conditional distributions as defined respectively in Eqs. (7)–(10) of the Supplementary Materials.
 5. When the response is binary, sample y_i^* as a truncated normal distribution as described in the Supplementary Materials (see Section 8).
 6. Sample \mathbf{b} from its full conditionals as defined in Eq. (11) of the Supplementary Materials.

3.1. Prediction

We use an M -fold cross-validation where the n samples are partitioned into M subsets. We follow a similar procedure adopted in Chekouo et al., 2017 for survival time outcome. We set $M = 10$ for our LGG case study, and for each sample $i = 1, \dots, n$, we denote with $\kappa(i)$ the partition that contains i (validation set). Let $\mathbf{y}_{\kappa(i)}$ be the binary outcome vector in the partition $\kappa(i)$, and $\mathbf{y}_{-\kappa(i)}$ be the outcome from the remaining partitions (training set). The cross-validation density for the i -th individual is estimated via model averaging; following Gelfand and Dey, 1994; Lamnisos et al., 2012 and Chekouo et al., 2017, we use $\pi(\Gamma|\mathbf{y})$ as the importance density of the target distribution $\pi(\Gamma|\mathbf{y}_{-\kappa(i)})$, where $\Gamma = \{\gamma_{kl}^{(r)}, r, k, l\}$. The predictive probability can then be written as

$$\begin{aligned} \text{prob}(y_i = 1|\mathbf{y}_{-\kappa(i)}) &= \sum_{\Gamma} \text{prob}(y_i = 1|\mathbf{y}_{-\kappa(i)}, \Gamma)\pi(\Gamma|\mathbf{y}_{-\kappa(i)}) \\ &\approx \sum_{g=1}^G w_g \text{prob}(y_i = 1|\mathbf{y}_{-\kappa(i)}, \Gamma(g)) / \sum_{g=1}^G w_g, \end{aligned} \quad (11)$$

where $w_g = \frac{\pi(\Gamma(g)|\mathbf{y}_{-\kappa(i)})}{\pi(\Gamma(g)|\mathbf{y})} \propto [p(\mathbf{y}_{\kappa(i)}|\mathbf{y}_{-\kappa(i)}, \Gamma(g))]^{-1}$, and G is the number of distinct models $\Gamma(g)$'s obtained from the MCMC samples. We approximate the densities

$$\text{prob}(y_i = 1|\mathbf{y}_{-\kappa(i)}, \Gamma(g)) \approx \Phi(\hat{\xi}_i \hat{\beta}_{\Gamma(g)}) \text{ and} \quad (12)$$

$$p(\mathbf{y}_{\kappa(i)}|\mathbf{y}_{-\kappa(i)}, \Gamma(g)) \approx \prod_{i' \in \kappa(i)} \Phi(\hat{\xi}_{i'} \hat{\beta}_{\Gamma(g)})^{*} > (1 - \Phi(\hat{\xi}_{i'} \hat{\beta}_{\Gamma(g)}))^{1-*} >, \quad (13)$$

where Φ is the cumulative distribution function of the standard normal distribution, $\hat{\xi}_i$ is the posterior mode of ξ_i obtained from Eq. (2), \bar{y}_i^* is the posterior mean determined by averaging sample values of y_i^* from each iteration g , $\hat{\beta}_{\Gamma(g)}$ is the posterior mode of $\pi(\beta|\Gamma(g), \hat{\xi}, \bar{y}_{-\kappa(i)}^*)$, and $* > =$ if $\bar{y}_i^* > 0$ and 0 otherwise. Hence, the inclusion probability for subject $i' \in \kappa(i)$ can be estimated as

$$\begin{aligned} \text{prob}\left(y_{i'} = 1|\mathbf{y}_{-\kappa(i)}\right) &= \sum_{g=1}^G [p(\mathbf{y}_{\kappa(i)}|\mathbf{y}_{-\kappa(i)}, \Gamma(g))]^{-1} \Phi\left(\hat{\xi}_{i'} \hat{\beta}_{\Gamma(g)}\right) \times \\ &1 / \sum_{g=1}^G [p(\mathbf{y}_{\kappa(i)}|\mathbf{y}_{-\kappa(i)}, \Gamma(g))]^{-1}, \end{aligned} \quad (14)$$

4. Simulation study

To demonstrate the performance of our 2D functional linear model on the GLCM data we conduct a simulation study under different scenarios. Note that in our real data analysis GLCMs are generated from the pixel values of the original tumors in the MRI scans of patients with low-grade gliomas. However, for simulation purposes we focus on simulating the GLCMs directly instead of trying to create them from simulated MRI scans.

Each GLCM is a $T \times T$ matrix where the (j, k) -th entry corresponds to the GLCM count generated to represent the neighborhood pattern of gray-level j and gray-level k . For each subject we want to generate four

$T \times T$ GLCMs corresponding to four imaging sequences. Let $G^{(r)} = ((g_{jk}^{(r)}))_{1 \leq j, k \leq T}$ be the GLCM corresponding to the imaging sequence r and $g_{jk}^{(r)}$ be its (j, k) -th entry for sequence r . Let $Q^{(r)} = ((q_{jk}^{(r)}))_{1 \leq j, k \leq T}$ be a matrix which specifies the underlying mean structure of the pattern in the GLCM. The entries of $Q^{(r)}$ are generated in a deterministic way to include different patterns for the GLCM. We determine the value of each $q_{jk}^{(r)}$ by corresponding it to the probability of a certain region under a bivariate normal distribution. Steps (7)-(8) in Algorithm 1 specify the exact computations to generate symmetric GLCMs. Note that Q is constructed to be a symmetric matrix to generate a symmetric GLCM. As alluded to previously, we observe correlation between the GLCMs across imaging sequences, which will be incorporated as correlation between the entries $g_{jk}^{(1)}, \dots, g_{jk}^{(4)}$ for all $j, k = 1, \dots, T$. We incorporate this correlation through the error term $\epsilon_{jk} = \epsilon_{kj} = (\epsilon_{jk}^{(1)}, \dots, \epsilon_{jk}^{(4)})^T$, where $\epsilon_{jk} \sim N_4(0, \sigma_e^2 \Delta_{jk})$ and Δ_{jk} is a pre-specified correlation structure with σ_e^2 being the noise. For the simulations in this section, we consider Δ_{jk} to be the empirical correlations between the (j, k) -th GLCM entries across all four sequences based on all the subjects from our real data. We now compute the logarithm of entries of $G^{(r)}$ as $\log(g_{jk}^{(r)}) = \log(w) + \log(q_{jk}^{(r)}) + \epsilon_{jk}^{(r)}$, where w is an integer sampled randomly from a discrete uniform distribution $U\{v_1, v_2\}$ such that $v_1 < w < v_2$. Here w accommodates for the varying tumor volume (total number of tumor pixels) for each subject. We choose $v_1 = 4 \times 10^4$ and $v_2 = 3 \times 10^6$ based on the real data. The inclusion of additive noise in log-scale could also be interpreted as multiplicative noise for the original GLCM count $g_{jk}^{(r)} = w q_{jk}^{(r)} \exp(\epsilon_{jk}^{(r)})$. The values $g_{jk}^{(r)}$ will be rounded to the nearest integer to generate integer-valued GLCM counts. Algorithm 1 summarizes the procedure to simultaneously generate correlated GLCMs for multiple imaging sequences. Note that first we generate $y_i \in \{0, 1\}$ randomly from a Bernoulli(0.5), which indicates the category label for subject i . Based on this label we assign the mean parameter $\mathbf{m}^{(r)}$ and variance parameter $\Sigma^{(r)}$ to specify the underlying patterns in the GLCM.

Algorithm 1 Simulating symmetric GLCMs for four imaging sequences

- 1: Set $(\mathbf{m}^{(r)}, \Sigma^{(r)}) = (\mathbf{m}_1^{(r)}, \Sigma_1^{(r)})$ or $(\mathbf{m}_0^{(r)}, \Sigma_0^{(r)})$ based on $y = 1$ or 0, respectively.
- 2: Let w be an integer sampled randomly from a discrete uniform distribution $U\{v_1, v_2\}$.
- 3: **for** $j \in 1, \dots, T$ **do**
- 4: **for** $k \in 1, \dots, j$ **do**
- 5: Let $\epsilon_{jk} = \epsilon_{kj} = (\epsilon_{jk}^{(1)}, \dots, \epsilon_{jk}^{(4)})^T$, where $\epsilon_{jk} \sim N_4(0, \sigma_e^2 \Delta_{jk})$.
- 6: **for** $r \in 1, \dots, 4$ **do**
- 7: $S \leftarrow$ square given by $(k-1, T-j), (k, T-j), (k, T-j+1), (k-1, T-j+1)$.
- 8: $q_{jk}^{(r)} = q_{kj}^{(r)} = P(\mathbf{X} \in S)$, where $\mathbf{X} \sim N_2(\mathbf{m}^{(r)}, \Sigma^{(r)})$.
- 9: Compute the GLCM entry as $g_{jk}^{(r)} = w \exp(\log(q_{jk}^{(r)}) + \epsilon_{jk}^{(r)})$.

To create the different simulation scenarios, we fix the number of gray-levels as $T = 8$. To construct the GLCMs for the subjects with $y_i = 1$, we consider the mean parameter as $\mathbf{m}_1^{(r)}$ and covariance matrix as $\Sigma_1^{(r)}$ for all $r = 1, \dots, 4$. Similarly, for the subjects with $y_i = 0$, we consider the mean parameter as $\mathbf{m}_0^{(r)}$ and covariance matrix $\Sigma_0^{(r)}$ for all $r = 1, \dots, 4$. Let $\Sigma = \begin{bmatrix} 1 & & & \\ & -0.1 & & \\ & & 1 & \\ & & & -0.1 \end{bmatrix}$ be a correlation matrix. We assume the same covariance structure to construct GLCMs of all four sequences for the subjects in the same category, that is, $\Sigma_1^{(1)} = \Sigma_1^{(2)} = \Sigma_1^{(3)} = \Sigma_1^{(4)} = c_l \Sigma$ for some $c_l > 0$ and $l = 0, 1$. We broadly divide our simulation scenarios into three cases with respect to different mean patterns in the GLCM between all four imaging sequences. These scenarios are trying to capture various degrees of co-occurrence for different gray-levels. For example, higher values close to the diagonal reflects strong co-occurrence of voxel values indicating spatial homogeneity, whereas, higher values away from the diagonal reflect less co-occurrence indicating spatial heterogeneity. Table 1 shows the specific choices of $\mathbf{m}_1^{(r)}$ and $\mathbf{m}_0^{(r)}$ for all $r = 1, \dots, 4$ for different simulation scenarios we consider. These choices are structured so that two of the four imaging sequences

Table 1

Choice of the mean parameter for the bivariate normal distribution used to generate $Q^{(r)}$ to create simulation cases *A*, *B* and *C*.

Case	$\{m_1^{(1)}, m_1^{(2)}, m_1^{(3)}, m_1^{(4)}\}$	$\{m_0^{(1)}, m_0^{(2)}, m_0^{(3)}, m_0^{(4)}\}$
<i>A</i>	$\{(1, 7), (3, 5), (5, 3), (7, 1)\}$	$\{(1, 7), (2, 6), (4, 4), (7, 1)\}$
<i>B</i>	$\{(1, 7), (1, 4), (1, 1), (5, 2)\}$	$\{(1, 7), (4, 4), (6, 2), (5, 2)\}$
<i>C</i>	$\{(4, 4), (4, 4), (4, 4), (4, 4)\}$	$\{(4, 4), (4, 4), (4, 4), (4, 4)\}$

have similar underlying patterns across both the response categories. That is, only GLCMs with sequences $r = 2$ and $r = 3$ are associated with the response for each scenario (except case *C*). Figs. 2–4 show the generated GLCMs for a subject with $y = 1$ for cases *A*, *B* and *C*, respectively.

For each of the three cases *A*, *B* and *C* we further consider two scenarios through varying covariance structure of the bivariate normals. We consider the first scenario with $c_0 = c_1 = 10$ and the second scenario with $c_0 = 10$ and $c_1 = 8$ for each of the three cases. We also consider two possible values for the noise: (a) $\sigma_\epsilon = 1$ and (b) $\sigma_\epsilon = 3$. In summary, we include the characteristics of GLCMs from MRI data such as (i) varying tumor volume (total number of tumor pixels) between subjects using the parameter w , (ii) dependence between the GLCM entries from the four imaging sequences, and (iii) various spatial heterogeneity/homogeneity patterns using the entries of $Q^{(r)}$. We have also generated similarly non-symmetric GLCMs with the same scenarios. The algorithm to generate them is presented in Section 2 of the Supplementary Materials.

Hyperparameter Settings: We set the hyperparameters as follows. A vague prior is assigned to the coefficient \mathbf{b} of nonfunctional effects by specifying the hyperparameter h_0 as a large value. Specifically, we set $h_0 = 1000$. We also set $h = 1$, the hyperparameter of the regression effects which allows the covariance matrix of β to be positive definite. We set the hyperparameters of the prior distributions of $\sigma_{r'}$ as $a = 0.5$ and $b = 0.5$. We set a non-informative prior on $\lambda_{kl}^{(r)}$ by specifying its hyperparameters as $a_\lambda = b_\lambda = 0.001$. Finally, we set $p_r = 0.5$, the hyperparameter of $\gamma_{kl}^{(r)}$ and the prior probability to select a specific component. A sensitivity analysis with different choices of a , b , and h shows that the posterior inference is quite robust to the specification of these hyperparameters (details in Supplementary Material, Section 4).

4.1. Simulation results

To highlight the comparison of using symmetric and non-symmetric GLCM matrices, we defined two Bayesian methods for GLCM matrices: *Bayes2Dsym* is our Bayesian approach for 2D functional data that accounts only for symmetric GLCM, while *Bayes2Dnosym* can account for non-symmetric GLCMs. Both methods were run for 80,000 MCMC iterations with a burn-in period of 10,000 iterations. To assess the convergence of the MCMC algorithm on simulated data, we ran two

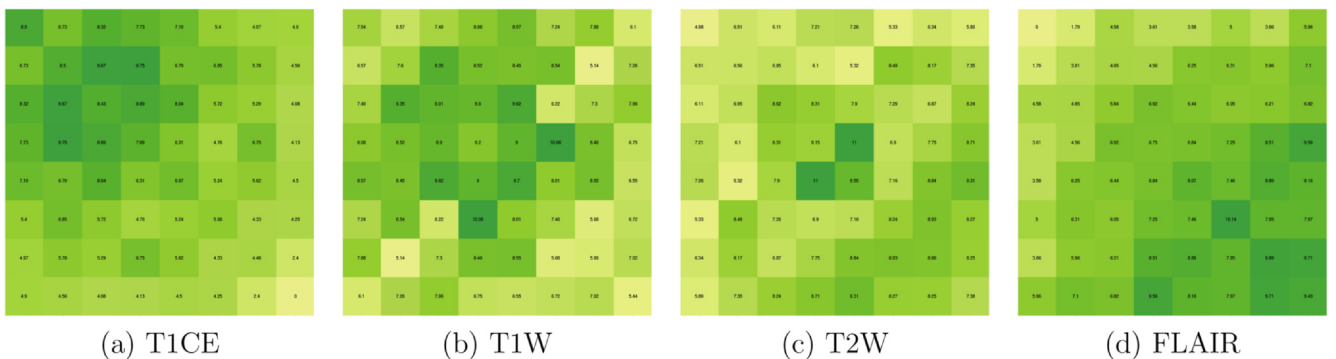


Fig. 2. Case *A*: $\log(\text{Counts})$ of symmetric GLCM for a subject with $y = 1$ where the pattern between imaging sequences is shifted on the diagonal. In this case, we have $\mathbf{m}_1^{(1)} = (1, 7)$, $\mathbf{m}_1^{(2)} = (3, 5)$, $\mathbf{m}_1^{(3)} = (5, 3)$ and $\mathbf{m}_1^{(4)} = (7, 1)$.

MCMC chains for each case with randomly chosen starting points. Correlations between the marginal posterior probabilities of $\gamma_{lk}^{(r)}$ were greater than 0.98 indicating good concordance between the two chains. Additional MCMC diagnostic checks can be found in the Supplementary Materials Section 5.

We further compared our methods with three alternative approaches that use summary statistics obtained from GLCM as covariates (i.e. the 13 Haralick features described in Section 1.2): the penalized L1 lasso binomial regression of Friedman et al., 2010 implemented in the R package *glmnet*, the SCAD penalty for support vector machine (SVM) of Zhang et al., 2005 implemented in the R package *penalizedSVM*, and the Random Forest of Breiman, 2001 implemented in the R package *randomForest*. We applied those methods with a total of $R \times 13 = 52$ covariate features. We computed the AUC to assess the prediction performance, and predictive probabilities were computed as described in Section 3.1 for $M = 10$.

Fig. 5 shows violin plots with density plots on each side, and sample means (marked as a diamond in the middle of plots) of AUCs. It shows the prediction performance for all competing methods across 12 scenarios: *Bayes2Dnosym* always performs best (much higher AUC), and *Bayes2Dsym* is the second best. This actually shows that the use of the raw GLCM (i.e without symmetrizing them) gives better predictive performance in particular when the noise in the GLCMs is large. Moreover, the prediction performance is worst when using summary statistics with the competing methods. However, all methods perform similarly in Case *B* under the scenario of different covariance structure between the two groups of responses (0 and 1), and $\sigma_\epsilon = 1$. We have also compared our methods with the three alternative methods that use the whole symmetric and non-symmetric GLCMs (not summary statistics) as covariates. We still outperform the competing methods (see Figs. 11 and 12 in Section 9 of the Supplementary Materials for details). Finally, we have also run our two methods when $\sigma_{r'} = 0$ i.e. without prior correlation between marginal effects of imaging sequences on the response \mathbf{y} . Results presented in Supplementary Material (Section 3) show that they are not significantly different from those with $\sigma_{r'} \neq 0$ but seem slightly inferior.

Fig. 6 shows plots for marginal posterior probabilities (MPP) of $\gamma_{lk}^{(r)}$, $r = 1, \dots, 4$, $k = 1, 2$ and $l = 1, 2$ for a scenario. They show that all MPPs of components from FLAIR ($r = 4$) and T1CE ($r = 1$) sequences are less than 0.5 but for sequences T1W ($r = 2$) and T2W ($r = 3$), some are larger than 0.5. This is aligned with our simulated data where sequences FLAIR and T1CE are not associated with the binary response as they were generated with the same means and covariances between the two groups of responses (i.e $m_1^{(1)} = m_0^{(1)} = (1, 7)$ and $m_1^{(4)} = m_0^{(4)} = (7, 1)$).

5. Application to TCGA LGG Data

In this section, we summarize results from our methods applied to

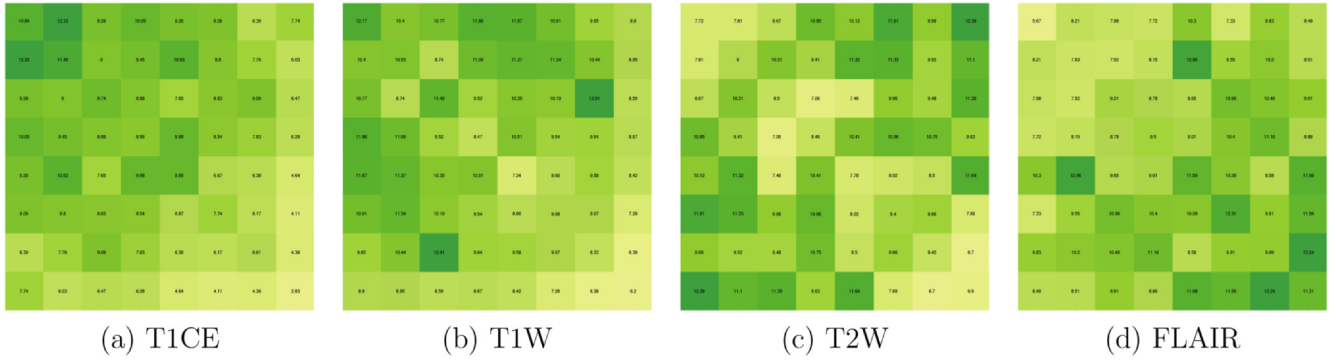


Fig. 3. Case B: log(Counts) of symmetric GLCM for a subject with $y = 1$ where the pattern between imaging sequences is shifted away from the diagonal. In this case, we have $\mathbf{m}_1^{(1)} = (1, 7)$, $\mathbf{m}_1^{(2)} = (1, 4)$, $\mathbf{m}_1^{(3)} = (1, 1)$ and $\mathbf{m}_1^{(4)} = (5, 2)$.



Fig. 4. Case C: log(Counts) of symmetric GLCM for a subject with $y = 1$ where the pattern between imaging sequences is the same. In this case, we have $\mathbf{m}_1^{(1)} = \mathbf{m}_1^{(2)} = \mathbf{m}_1^{(3)} = \mathbf{m}_1^{(4)} = (4, 4)$.

the TCGA LGG dataset described in Section 1.1, which consists of $n = 82$ TCGA LGG patients whose MRI scans were available for all $R = 4$ imaging sequences (T1W, T1CE, T2W and FLAIR). To assess the utility of the whole GLCM in terms of prediction on the LGG data, we fitted the following models: (i) our models Bayes2Dnosym and Bayes2Dsym by using the non-symmetric and symmetric GLCM matrices respectively and (ii) competing methods as described in Section 4.1 using summary statistics as covariates. We adjusted all the models with clinical data such as age and gender.

To assess the convergence of the Bayes2Dnosym and Bayes2Dsym models, we fitted 10 MCMC chains with different starting points. We assessed the convergence of binary variables $\gamma_{kl}^{(r)}$ by evaluating the correlation coefficients between their marginal posterior probabilities. These indicated good concordance between the 10 chains, with all correlations > 0.90 for each imaging sequence (figures in [Supplementary Material, Section 5](#)). Additional MCMC diagnostics checks for continuous parameters (in particular $\sigma_{rr'}$) can be found in [Supplementary Material](#). We run our methods with the same hyperparameter settings as described in Section 4. [Figure 7\(b\)](#) shows violin plots of AUCs for predictive performances of the aforementioned models. It shows that our two methods (Bayes2Dsym and Bayes2Dnosym) perform much better than the competing methods as they have higher AUCs. Moreover, our method Bayes2Dnosym which accounts for non-symmetric GLCMs has the highest predictive performance. Our methods outperform the competing methods, when the competing methods use the whole GLCM data as covariates, showing the importance of modeling the GLCM data as 2D objects. [Fig. 8](#).

[Figs. 9\(a\)–\(d\)](#) show marginal posterior probabilities of component inclusion $\gamma_{kl}^{(r)}$ for each imaging sequence $r = 1, 2, 3, 4$. We can observe that only the imaging sequence T1CE does not have high probabilities which would confirm that sequences T1W, T2W and FLAIR are potential biomarkers for IDH status in LGG disease ([Patel et al., 2017](#)). From the four sequences of the MR imaging in lower grade gliomas, it is

known that the tumor region could be slightly hypointense compared to white matter in T1W, whereas usually there is no enhancement observed in T1CE. However, with the T2W and FLAIR sequences, the tumor regions are usually observed with a mass-like hyperintense signal. This confirms with our findings as the marginal posterior probabilities corresponding to components from T1W, T2W and FLAIR are clearly higher as shown in [Figs. 9\(a\)–9\(d\)](#). This indicates that the differences in signals from these sequences is contributing towards better predictive capability of the model for the IDH status. For diffuse-astrocytomas, the tumor region in T1CE does not usually show enhancement except rare instances of small ill-defined areas ([Bruno et al., 2019](#)). In oligodendrogliomas contrast enhancement is shown in about 50% of the cases but is not a reliable indicator ([Frank, 2019](#)). The lack of enhancement could potentially contribute towards the GLCM not being able to consistently capture the information about the spatial heterogeneity. This could be one of the possible reasons for not seeing significant associations with the GLCMs from T1CE.

[Figs. 10\(a\)–\(d\)](#) show estimated functional effects for the non-symmetric case, where we see that the effect sizes corresponding to the GLCM for T1CE are lower in magnitude in comparison to the other three imaging sequences. In case of T1W since the tumor region is usually observed to be hypointense compared to the white matter, we expect to see more voxels with intensities in the middle of the gray-scale. This fact is captured from our analysis as we see that the functional effect sizes are high in magnitude for the cells corresponding to middle region on the gray-scale. However, in T2 and FLAIR sequences, the tumor region is mostly observed to be hyperintense and we see more voxels with intensities on the higher end of the gray-scale. This aspect is clearly captured as we see higher magnitudes for the functional effect sizes corresponding to the higher end of the gray-scale. Similarly, in the symmetric case we see from [Figs. 11\(a\)–\(d\)](#) that the functional effect sizes for T1CE are too small in magnitude compared to the other three imaging sequences. For T1W images we see that the

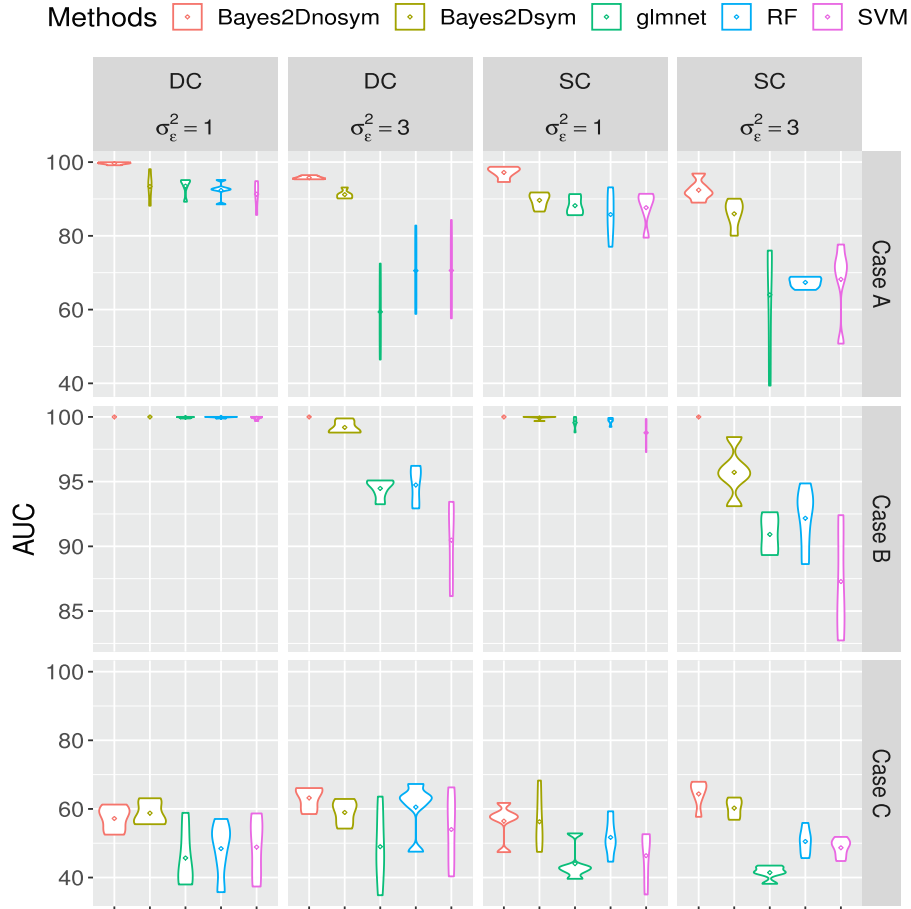


Fig. 5. Simulation results using whole symmetric and non-symmetric GLCMs with our methods (Bayes2Dsym and Bayes2Dnosym), but competing methods (RF, glmnet and SVM) are fitted using (symmetric) GLCM-derived Haralick summary features. AUC were computed (in %) to evaluate the prediction performance. SC stands for same covariance i.e. $c_0 = c_1 = 10$, and DC stands for different covariances i.e. $c_0 = 10$ and $c_1 = 8$.

effect sizes are higher for regions in the middle of the gray-scale consistent with the non-symmetric case. For T2W and FLAIR sequences we see higher magnitudes on the higher-end of the gray-scale as the tumor region is observed to be hyperintense in the image which is also consistent with the non-symmetric case.

6. Conclusion

In this manuscript, we have proposed a Bayesian predictive model to accurately predict the IDH status for LGG patients based on GLCM matrices obtained from multiple imaging sequences. Our Bayesian approach has several innovative characteristics: (i) it employs novel image-dependent variable selection priors, leveraging correlations

between imaging sequences through priors imposed on regression coefficients, (ii) it explicitly accounts for the structure of the whole GLCM matrices rather than restricting its use to some derived summary statistics, (iii) it accounts for both symmetric and non-symmetric GLCM matrices, and (iv) it treats any sequence r of GLCM counts as 2D functional data, indexed by pairwise gray-level intensities. We elaborate on these features below. The performance of our method was evaluated on simulated data and on a data set collected from TCGA LGG patients. Both studies show that using the whole GLCM matrices enhances prediction performances. While GLCM-derived features are usually computed on symmetric matrices as symmetric GLCMs are rotation and direction invariant, our method with non-symmetric GLCMs performs slightly better than symmetric GLCMs in terms of prediction

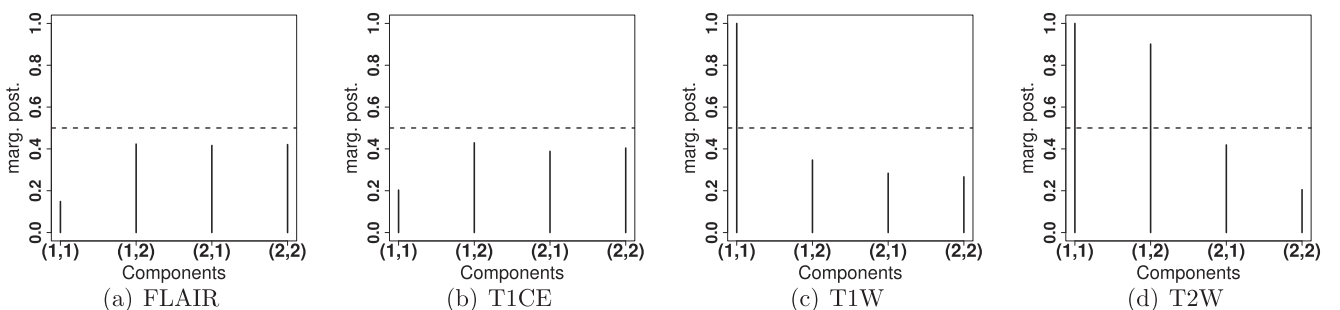


Fig. 6. Marginal posterior probabilities of $\gamma_k^{(r)}$ on a simulated data scenario: Case A, $c_0 = c_1 = 10$ and $\sigma_\epsilon^2 = 3$.

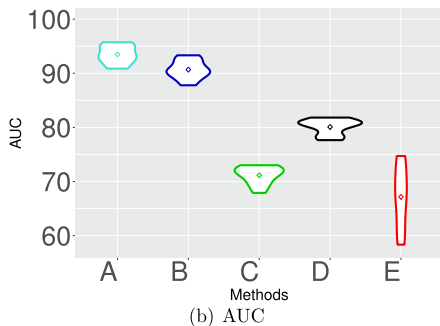
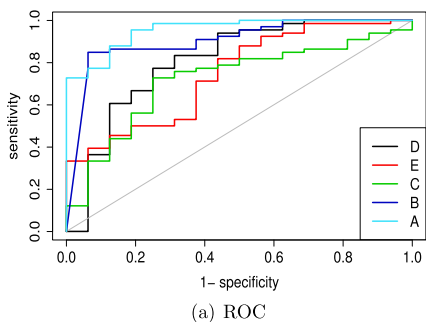


Fig. 7. Comparison of predictive performance on the LGG data over 10 partitions of the sample data. Symmetric and non-symmetric GLCMs are used with our methods Bayes2Dsym and Bayes2Dnosym respectively, but competing methods (RF, glmnet and SVM) are fitted using GLCM-derived haralick features (computed from symmetric GLCMs). (A) Bayes2Dnosym; (B) Bayes2Dsym; (C) glmnet; (D) RF and (E) SVM.

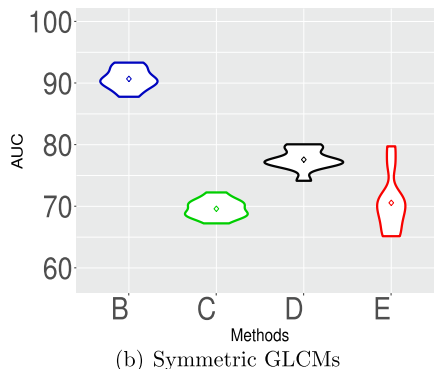
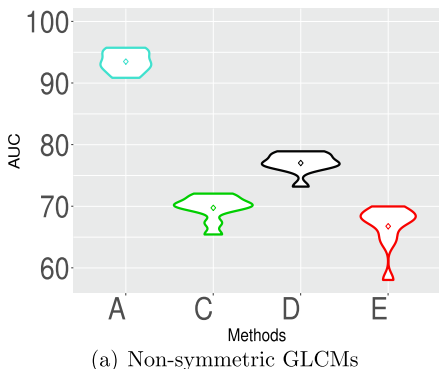


Fig. 8. Comparison of predictive performance on the LGG data over 10 partitions of the sample data. competing methods used the whole GLCM data as covariates. (A) Bayes2Dnosym; (B) Bayes2Dsym; (C) glmnet; (D) RF and (E) SVM.

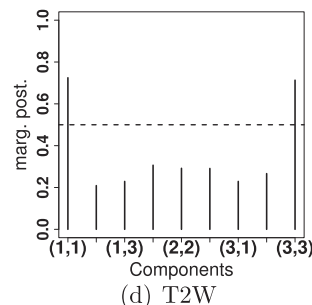
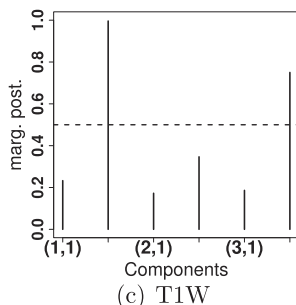
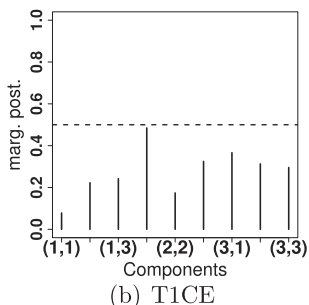
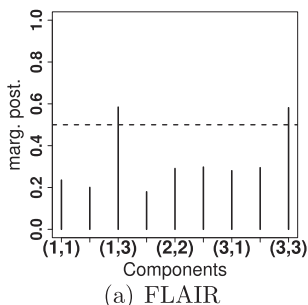


Fig. 9. Marginal posterior probabilities of $\gamma_k^{(r)}$ obtained from the method Bayes2Dnosym.

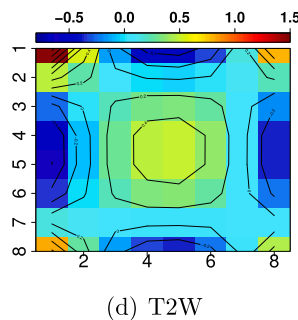
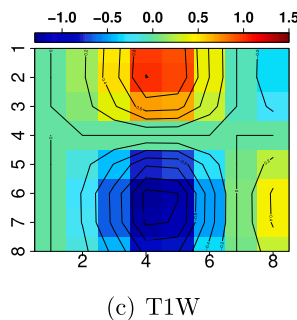
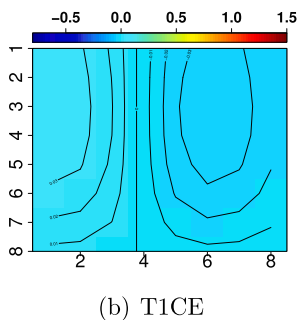
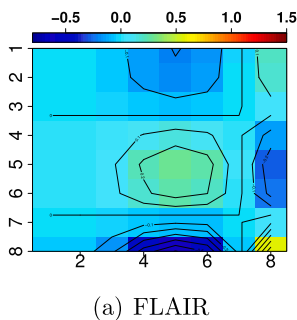


Fig. 10. Functional regression effects of GLCMs for imaging sequence r , $\beta_i(s, t)$, based on non-symmetric matrices.

performance.

Prediction of IDH mutation status (and other molecular status) in gliomas using radiomic features is an extensively studied problem in the field of radiomics. The first step in these methods is feature construction

from one or more MRI sequences, followed by either (a) standard statistical modeling (e.g. logistic regression, random forest) using those features (Chaddad et al., 2015; Hsieh et al., 2017; Jakola et al., 2018; Li et al., 2018; Tian et al., 2018; Han et al., 2019; Liu et al., 2019), or (b)

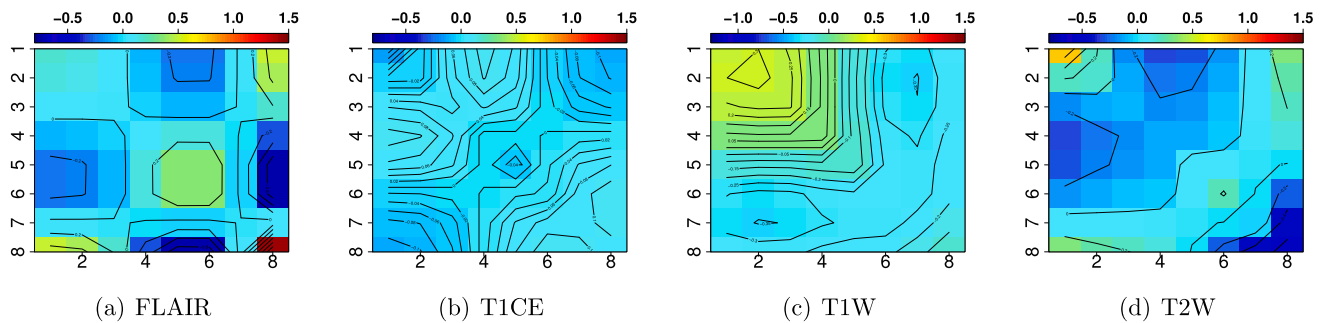


Fig. 11. Functional regression effects of GLCMs for imaging sequence r , $\beta_r(s, t)$, based on symmetric GLCM matrices.

deep learning-based approaches (Li et al., 2017; Chang et al., 2018). Some of the existing literature for IDH prediction in gliomas, using either texture features or morphological features or both, report AUCs greater than 0.9. However, one of the main drawbacks of these approaches is the subjectivity in the choice of what features to include, and how many of them to include. The choice of these features is also reliant upon the pre-processing steps involved during the feature selection. However, our proposed method addresses the issue of subjectivity in the choice of features, and is an end-to-end approach that takes as input the MRI scans from four imaging sequences to compute the corresponding GLCMs. These GLCMs are directly used as data objects in the modeling such that there is no subjectivity in feature construction. We build a classification model for the IDH status by directly using the GLCMs from the four MRI sequences as predictors and treating them as 2D functional data objects. Our approach aims to capture the entire information in the GLCM instead of only using summary statistics constructed from the GLCM as predictors. We developed a rigorous functional data based approach, that respects the underlying geometry of the space of GLCMs, to predict the IDH status. Our approach provides an alternative way that complements the existing approaches which use texture/morphological features to predict the molecular status in LGGs.

Our Bayesian hierarchical model can be generalized and extended in different directions. One can incorporate molecular data features (e.g. RNA-seq and proteomic data) as potential response variables. This may also require the incorporation of the dependence between GLCM and molecular data as similarly implemented in Chekouo et al., 2016. Although our motivating example was LGGs, our framework could readily be applied to gliomas in general, and other types of cancer where GLCMs can be constructed from imaging data.

CRediT authorship contribution statement

Thierry Chekouo: Methodology, Software, Formal analysis, Resources, Data curation, Writing - original draft, Funding acquisition. **Shariq Mohammed:** Conceptualization, Software, Formal analysis, Data curation, Writing - original draft, Funding acquisition. **Arvind Rao:** Conceptualization, Data curation, Formal analysis, Writing - review & editing, Resources, Project administration, Funding acquisition.

Acknowledgements

Shariq Mohammed and Arvind Rao were supported through CCSG P30 CA046592, Institutional Research Grants from The University of Michigan, NCI R37CA214955-01A1, and a Research Scholar Grant from the American Cancer Society (RSG-16-005-01). Shariq Mohammed was partially supported by the Precision Health at The University of Michigan. Thierry Chekouo was partially supported by NSERC Discovery Grants number RGPIN-2019-04810.

Appendix A. Supplementary data

Web Appendices 1–9, referenced in Sections 2, 3, 4 and 5 are available along with this manuscript. We have created an R package called *BayesF2D* and published it on GitHub. Its source R and C codes, along with the user pdf manual are also available in the GitHub repository <https://github.com/chekouo/BayesF2D>.

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.nicl.2020.102437>.

References

- Albert, J.H., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. *J. Am. Stat. Assoc.* 88 (422), 669–679.
- Albregtsen, F., et al. 2008. Statistical texture measures computed from gray level co-occurrence matrices. Image processing laboratory, department of informatics, university of oslo 5.
- Baladandayuthapan, V., Wang, X., Mallick, B.K., Do, K.-A., 2015. *Bayesian Functional Mixed Models for Survival Responses with Application to Prostate Cancer*. Springer International Publishing, Cham, pp. 35–59.
- Bratteli, O., Robinson, D., 2003. Operator Algebras and Quantum Statistical Mechanics: Equilibrium States. *Models in Quantum Statistical Mechanics. Theoretical and Mathematical Physics*. Springer Berlin Heidelberg.
- Breiman, L., 2001. Random forests. *Machine Learning* 45 (1), 5–32.
- Bruno, D.M., Frank, G., et al., 2019. (accessed August, 2020). Diffuse astrocytoma. <https://radiopaedia.org/articles/diffuse-astrocytoma-1?lang=us>.
- Cardot, H., Faivre, R., Goulard, M., 2003. Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data. *J. Appl. Stat.* 30 (10), 1185–1199.
- Cardot, H., Ferraty, F., Sarda, P., 1999. Functional linear model. *Statistics Probability Lett.* 45 (1), 11–22.
- Chaddad, A., Zinn, P.O. Colen, R.R., 2015. Radiomics texture feature extraction for characterizing gbm phenotypes using glcm. In 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 84–87. IEEE.
- Chang, K., Bai, H.X., Zhou, H., Su, C., Bi, W.L., Agbodza, E., Kavouridis, V.K., Senders, J.T., Boaro, A., Beers, A., et al., 2018. Residual convolutional neural network for the determination of idh status in low-and high-grade gliomas from mr imaging. *Clin. Cancer Res.* 24 (5), 1073–1081.
- Chekouo, T., Stingo, F.C., Doekoe, J.D., Do, K.-A., 2017. A Bayesian integrative approach for multi-platform genomic data: A kidney cancer case study. *Biometrics* 73 (2), 615–624.
- Chekouo, T., Stingo, F.C., Guindani, M., Do, K.-A., 2016. A Bayesian predictive model for imaging genetics with application to schizophrenia. *Ann. Appl. Stat.* 10 (3), 1547–1571.
- Chen, K., Delicado, P., Müller, H.-G., 2017. Modelling function-valued stochastic processes, with applications to fertility dynamics. *J. R. Stat. Soc.: Series B (Stat. Methodol.)* 79 (1), 177–196.
- Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al., 2013. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J. Digital Imaging* 26 (6), 1045–1057.
- Crainiceanu, C., Goldsmith, A., 2010. Bayesian functional data analysis using WinBUGS. *J. Stat. Softw., Articles* 32 (11), 1–33.
- Eckel-Passow, J.E., Lachance, D.H., Molinaro, A.M., Walsh, K.M., Decker, P.A., Sicotte, H., Pekmezci, M., Rice, T., Kosel, M.L., Smirnov, I.V., et al., 2015. Glioma groups based on 1p/19q, IDH, and TERT promoter mutations in tumors. *N. Engl. J. Med.* 372 (26), 2499–2508.
- Eilers, P.H.C., Marx, B.D., 2003. Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics Intell. Lab. Syst.* 66 (2), 159–174.
- Ferraty, F., Vieu, P., 2002. The functional nonparametric model and application to spectrometric data. *Comput. Stat.* 17 (4), 545–564.
- Ferraty, F., Vieu, P., 2003. Curves discrimination: a nonparametric functional approach. *Computational Statistics & Data Analysis* 44(1), 161–173. Special Issue in Honour of

- Stan Azen: a Birthday Celebration.
- Frank, G.E.A. (2019 (accessed August, 2020)). Oligodendroglioma. <https://radiopaedia.org/articles/oligodendroglioma?lang=us>.
- Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33 (1), 1–22.
- Gelfand, A.E., Dey, D.K., 1994. Bayesian model choice: asymptotics and exact calculations. *J. R. Stat. Soc. Series B (Method.)* 56 (3), 501–514.
- Ghosh, P., Tamboli, P., Vikram, R., Rao, A., 2015. Imaging-genomic pipeline for identifying gene mutations using three-dimensional intra-tumor heterogeneity features. *J. Med. Imaging 2* (4) 041009.
- Hall, P., Muller, H.-G., Wang, J.-L., 2006. Properties of principal component methods for functional and longitudinal data analysis. *Ann. Stat.* 34, 1493–1517.
- Han, L., Wang, S., Miao, Y., Shen, H., Guo, Y., Xie, L., Shang, Y., Dong, J., Li, X., Wang, W., et al., 2019. Mri texture analysis based on 3d tumor measurement reflects the idh1 mutations in gliomas—a preliminary study. *Eur. J. Radiol.* 112, 169–179.
- Haralick, R.M., Shanmugam, K., Dinstein, I.H., 1973. Textural features for image classification. *IEEE Trans. Syst., Man, Cybern. SMC* 3 (6), 610–621.
- Hartmann, C., Hentschel, B., Wick, W., Capper, D., Felsberg, J., Simon, M., Westphal, M., Schackert, G., Meyermann, R., Pietsch, T., et al., 2010. Patients with IDH1 wild type anaplastic astrocytomas exhibit worse prognosis than IDH1-mutated glioblastomas, and IDH1 mutation status accounts for the unfavorable prognostic effect of higher age: implications for classification of gliomas. *Acta Neuropathologica* 120 (6), 707–718.
- Hitam, M.S., Muslan, M.Y.H., Deris, M.M., Saman, M.Y.M., 2003. Image texture classification using gray level co-occurrence matrix and neural network. In *Proceedings of the 2nd WSEAS International Conference on Electronics, Control and Signal Processing*, pp. 69. World Scientific and Engineering Academy and Society (WSEAS).
- Hsieh, K.L.-C., Chen, C.-Y., Lo, C.-M., 2017. Quantitative glioma grading using transformed gray-scale invariant textures of mri. *Computers Biol. Med.* 83, 102–108.
- Jakola, A.S., Zhang, Y.-H., Skjulsvik, A.J., Solheim, O., Bø, H.K., Berntsen, E.M., Reinertsen, I., Gulati, S., Förander, P., Brismar, T.B., 2018. Quantitative texture analysis in the prediction of idh status in low-grade gliomas. *Clinical Neurol. Neurosurgery* 164, 114–120.
- James, G.M., 2002. Generalized linear models with functional predictors. *J. R. Stat. Soc.: Series B (Stat. Methodol.)* 64 (3), 411–432.
- Kassner, A., Thornhill, R., 2010. Texture analysis: a review of neurologic mr imaging applications. *Am. J. Neuroradiol.* 31 (5), 809–816.
- Kumar, R., Srivastava, R., Srivastava, S., 2015. Detection and classification of cancer from microscopic biopsy images using clinically significant and biologically interpretable features. *J. Med. Eng. 2015* 457906.
- Kuthuru, S., Deaderick, W., Bai, H., Su, C., Vu, T., Monga, V., Rao, A., 2018. A visually interpretable, dictionary-based approach to imaging-genomic modeling, with low-grade glioma as a case study. *Cancer Inform.* 17 1176935118802796.
- Lamnisos, D., Griffin, J.E., Steel, M.F.J., 2012. Cross-validation prior choice in Bayesian probit regression with many covariates. *Stat. Comput.* 22 (2), 359–373.
- Lee, J., Narang, S., Martinez, J., Rao, G., Rao, A., 2015. Spatial habitat features derived from multiparametric magnetic resonance imaging data are associated with molecular subtype and 12-month survival status in glioblastoma multiforme. *PloS one* 10 (9) e0136557.
- Leu, S., von Felten, S., Frank, S., Vassella, E., Vajtai, I., Taylor, E., Schulz, M., Hutter, G., Hench, J., Schucht, P., et al., 2013. IDH/MGMT-driven molecular classification of low-grade glioma is a strong predictor for long-term survival. *Neuro-oncology* 15 (4), 469–479.
- Li, X., Guindani, M., Ng, C.S., Hobbs, B.P., 2019. Spatial Bayesian modeling of GLCM with application to malignant lesion characterization. *J. Appl. Stat.* 46 (2), 230–246.
- Li, Z., Wang, Y., Yu, J., Guo, Y., Cao, W., 2017. Deep learning based radiomics (dlr) and its usage in noninvasive idh1 prediction for low grade glioma. *Sci. Rep.* 7 (1), 1–11.
- Li, Z.-C., Bai, H., Sun, Q., Zhao, Y., Lv, Y., Zhou, J., Liang, C., Chen, Y., Liang, D., Zheng, H., 2018. Multiregional radiomics profiling from multiparametric mri: Identifying an imaging predictor of idh1 mutation status in glioblastoma. *Cancer medicine* 7 (12), 5999–6009.
- Liu, X., Li, Y., Li, S., Fan, X., Sun, Z., Yang, Z., Wang, K., Zhang, Z., Jiang, T., Liu, Y., et al., 2019. Idh mutation-specific radiomic signature in lower-grade gliomas. *Aging (Albany NY)* 11 (2), 673.
- Loève, M., 1963. *Probability Theory*, 3rd ed. Springer, New York.
- Marx, B.D., Eilers, P.H., 2005. Multidimensional penalized signal regression. *Technometrics* 47 (1), 13–22.
- Marx, B.D., Eilers, P.H., Li, B., 2011. Multidimensional single-index signal regression. *Chemometrics and Intelligent Laboratory Systems* 109 (2), 120–130.
- Moritz-Gasser, S., Herbet, G., 2013. Language and other cognitive evaluations. I diffuse low-grade gliomas in adults: natural history, interaction with the brain, and new individualized therapeutic strategies.
- Morris, J.S., 2015. Functional regression. *Ann. Rev. Stat. Appl.* 2 (1), 321–359.
- Muller, H.-G., 2005. Functional modelling and classification of longitudinal data. *Scand. J. Stat.* 32 (2), 223–240.
- Narang, S., Kim, D., Aithala, S., Heimberger, A.B., Ahmed, S., Rao, D., Rao, G., Rao, A., 2017. Tumor image-derived texture features are associated with CD3 T-cell infiltration status in glioblastoma. *Oncotarget* 8 (60) 101244.
- Network, C.G.A.R., 2015. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* 372 (26), 2481–2498.
- Patel, S.H., Poisson, L.M., Brat, D.J., Zhou, Y., Cooper, L., Snuderl, M., Thomas, C., Franceschi, A.M., Griffith, B., Flanders, A.E., Golfinos, J.G., Chi, A.S., Jain, R., 2017. T2-FLAIR Mismatch, an Imaging Biomarker for IDH and 1p/19q Status in Lower-grade Gliomas: A TCGA/TCIA Project. *Clin. Cancer Res.* 23 (20), 6078–6085.
- Pedano, N., Flanders, A.E., Scarpace, L., Mikkelsen, T., Eschbacher, J., Hermes, B., Ostrom, Q., 2016. Radiology data from the cancer genome atlas low grade glioma [tcga-agg] collection. *Cancer Imaging Arch.* 2.
- Pignatti, F., Van Den Bent, M., Curran, D., Debruyne, C., Sylvester, R., Therasse, P., Afra, D., Cornu, P., Bolla, M., Vecht, C., et al., 2002. Prognostic factors for survival in adult patients with cerebral low-grade glioma. *J. Clin. Oncol.* 20 (8), 2076–2084.
- Ramsay, J.O., Dalzell, C.J., 1991. Some Tools for Functional Data Analysis (with discussion). *J. R. Stat. Soc.* 53 (3), 539–572.
- Ramsay, J.O., Silverman, B.W., 2005. *Functional Data Analysis* (2nd ed.). Springer Series in Statistics. Springer.
- Rathore, S., Akbari, H., Rozycki, M., Abdullah, K.G., Nasrallah, M.P., Binder, Z.A., Davuluri, R.V., Lustig, R.A., Dahmane, N., Bilello, M., et al., 2018. Radiomic mri signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond idh1. *Sci. Rep.* 8 (1), 1–12.
- Shinohara, R.T., Sweeney, E.M., Goldsmith, J., Shiee, N., Mateen, F.J., Calabresi, P.A., Jarso, S., Pham, D.L., Reich, D.S., Crainiceanu, C.M., et al., 2014. Statistical normalization techniques for magnetic resonance imaging. *NeuroImage: Clinical* 6, 9–19.
- Takesaki, M., 2001. *Theory of Operator Algebras I. Encyclopaedia of Mathematical Sciences*. Springer, Berlin Heidelberg.
- Tian, Q., Yan, L.-F., Zhang, X., Zhang, X., Hu, Y.-C., Han, Y., Liu, Z.-C., Nan, H.-Y., Sun, Q., Sun, Y.-Z., et al., 2018. Radiomics strategy for glioma grading using texture features from multiparametric mri. *J. Magn. Reson. Imaging* 48 (6), 1518–1528.
- Tomczak, K., Czerwińska, P., Wiznerowicz, M., 2015. The cancer genome atlas (TCGA): an immeasurable source of knowledge. *Contemporary Oncol.* 19 (1A), A68.
- van Dyk, D.A., Park, T., 2008. Partially collapsed gibbs samplers. *J. Am. Stat. Assoc.* 103 (482), 790–796.
- Vannucci, M., Brown, P., Fearn, T., 2003. A decision theoretical approach to wavelet regression on curves with a high number of regressors. *J. Stat. Planning Inference* 112(1), 195–212. Special issue II: Model Selection, Model Diagnostics, Empirical Bayes and Hierarchical Bayes.
- Vannucci, M., Sha, N., Brown, P.J., 2005. Nir and mass spectra classification: Bayesian methods for wavelet-based feature selection. *Chemometrics Intell. Lab. Syst.* 77 (1), 139–148.
- Yao, F., Muller, H.-G., Wang, J.-L., 2005. Functional data analysis for sparse longitudinal data. *J. Am. Stat. Assoc.* 100 (470), 577–590.
- Zhang, H.H., Ahn, J., Lin, X., Park, C., 2005. Gene selection using support vector machines with non-convex penalty. *Bioinformatics* 22 (1), 88–95.
- Zhang, J., Tong, L., Wang, L., Li, N., 2008. Texture analysis of multiple sclerosis: a comparative study. *Magnetic Resonance Imaging* 26 (8), 1160–1166.
- Zulpe, N., Pawar, V., 2012. GLCM textural features for brain tumor classification. *Int. J. Computer Sci. Issues (IJCSI)* 9 (3), 354.