Voxel-wise Classification of Prostate Cancer Using Multi-parametric MRI Data

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Jin Jin

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Joseph Koopmeiners and
Dr. Lin Zhang, Advisers

June 2019

ACKNOWLEDGEMENTS

First, I would like to thank my advisors, Drs. Joseph Koopmeiners and Lin Zhang, for the research ideas, time and guidance you have provided throughout the years, the great help during my job application process, and the DOD grant you provided to support my dissertation research. I started working with Joe and Lin when I was having difficulty confirming my interest in statistical research. The guidance and support I received from them have greatly helped me with my productivity and motivation for research, and also had a profound impact on my decision to pursue a career in academia.

I would like to thank Dr. Gregory Metzger and Ethan Leng from the Center for Magnetic Resonance Research at the University of Minnesota, who I have been collaborating with for the DOD grant. Working with them in the last four years has been a wonderful experience. Many thanks for the help they provided for the mpMRI data generation, and the collaboration opportunities they offered. Besides working with my advisors and collaborators, I am also honored to have worked with three other faculty members in our Division: Dr. Wei Pan, with whom I worked on a graphical model project for estimating the differential networks of gene expression levels, Dr. Cavan Reilly, with whom I worked on a microbiome project for taxonomical analysis of the microbial communities, and Dr. Mark Fiecas, with whom I have been working on longitudinal functional principal component analysis of cortical thickness data for the Alzheimer's Disease. I would like to thank them for giving me opportunities to explore my interest in various research fields.

I am super lucky to be in a program where all faculty members are extremely nice and have been offering all kinds of help to students. Dr. James Hodges has helped me a lot with my writing skills, especially when I was preparing for the research statements for the application of Doctoral Dissertation Fellowship and academic jobs. I am deeply influenced

by Dr. Haitao Chu's enthusiasm and efficiency for academic research. Dr. Xianghua Luo, who is the instructor of one of my TA courses, was nice to help me write a recommendation letter for my job applications. We have had conversations not just about work, but also about beliefs and attitudes toward life, which have inspired me in a lot of ways.

My Ph.D. life would not be complete without family and friends. Sincere gratitude to my parents, Zhihua Jin and Xiaoqing Gao, for their education that made me the person I am today, and for their support that made my pursuing for Ph.D. so much easier. I am also very lucky to meet so many good friends during graduate school. Thank you Dr. Jaron Arbet for the friendship and support, especially for our discussion on Genetics topics, which inspired my interest in working on some related topics for post-doctoral research. I would also like to thank Chuyu Deng, Yuan Zhang, Bin Guo, Roland Brown, Brian Hart, Xiaoyue Ma, Drs. Jason Xu, Chen Gao, Tianmeng Lyu, Qinshu Lian, Jincheng Zhou, Chong Wu, and many other friends/classmates that have made the graduate school experience wonderful to me.

Finally, I would like to express my special thanks to Drs. Laisheng Wei, Yaning Yang, Yu Chen, Boqi Liao, Taizhong Hu from the University of Science and Technology of China School of Management, and Drs. Xinan Ma, Falai Chen, Simin Li from the University of Science and Technology of China School of Mathematical Sciences. Your instructions have helped me build solid mathematical and statistical foundations, not only for my research during the Ph.D. program, but also for my potential contributions in the future.

DEDICATION

This dissertation is dedicated to my wonderful parents, Zhihua Jin and Xiaoqin Gao.

ABSTRACT

As a continuously developing tool for the diagnosis and prognosis of prostate cancer, multi-parametric magnetic resonance imaging (mpMRI) has been widely used in a variety of prostate cancer-related topics. While current research has shown the great potential of mpMRI in detecting prostate cancer, further investigation is needed for modeling some specific features of mpMRI, including the anatomic difference between different regions of a prostate, the spatial correlation between voxels within each prostate image, and the difference in the distribution of the observed mpMRI parameters between patients.

This dissertation focuses on novel statistical methods for the voxel-wise classification of prostate cancer using mpMRI data. Systematic modeling frameworks will be proposed to improve cancer classification by incorporating the aforementioned features of mpMRI. Three topics are discussed in depth: (1) development of a general Bayesian modeling framework that can incorporate the various mpMRI features; (2) how to model the mpMRI features in the proposed Bayesian framework, preferably in a computationally efficient manner; (3) development of an alternative approach to accounting for the mpMRI features, which uses a multi-resolution modeling technique to account for the regional heterogeneity, and is flexible to be extended to more complex classification problems for prostate cancer.

The solutions are presented in the following order. In Chapter 2, we propose a Bayesian hierarchical modeling framework that allows complex distributional assumptions for the various data components. Based on the modeling framework, two approaches will be proposed for modeling the heterogeneity between regions of the prostate, which can be combined with a spatial Gaussian kernel smoother to account for residual spatial correlation and reduce random noise in the data. In Chapter 3, we add additional layers in the hierarchical model to model the spatial correlation structure and between-patient heterogeneity.

Modeling the spatial correlation structure is computationally challenging and even infeasible for our mpMRI data set, due to the large number of voxels within each image. Three scalable spatial modeling approaches are then proposed for the correlation between voxels. In Chapter 4, we develop an alternative, machine learning-based method to account for the mpMRI features: a super learner with an ensemble learning technique is utilized to combine base learners trained in multi-resolution sub-regions. Specific algorithms will be introduced for both the classification of binary cancer status, and a more complex problem: classification of an ordinal outcome that indicates the clinical significance of prostate cancer. Method performance will be illustrated by simulation studies and applications to in vivo data that motivated the method's development.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

Prostate cancer is the second most common cancer and the second leading cause of cancer death among American men. According to the American Cancer Society, there will be about 174,650 new cases and 31,620 deaths from prostate cancer in 2019 in the United States (American Cancer Society, 2019).

Multi-parametric magnetic resonance imaging (mpMRI), which is a combination of the traditional anatomic and functional MRI methods, continues to evolve as a valuable tool for the detection and characterization of prostate cancer (Kurhanewicz et al., 2008; Dickinson et al., 2011). In recent years, mpMRI has begun to play an important role in evaluating the extent of prostate cancer and determining treatment strategy (e.g. biopsy, active surveillance, radical prostatectomy) (Hegde et al., 2013). Traditionally, mpMRI examinations were used to manually delineate the cancerous regions within the prostate. Such methods tend to depend highly on radiologists' and urologists' expertise, and require high efficiency due to time constraints. In the literature, various state-of-the-art techniques have been developed for manual examination of prostate cancer using mpMRI. Although mpMRI has shown potential, the implementation in clinical practice is hindered due to substantial variability in the standard radiological assessment (Garcia-Reyes et al., 2015). This variability

persists even with the development of the Prostate Imaging Reporting and Data System (PI-RADS) standard. PI-RADS relies on a qualitative assessment of the mpMRI data, and does not provide sufficient guidance on how to combine the different sources of information obtained from an mpMRI study (Barentsz et al., 2012; Rosenkrantz et al., 2013).

Conventional manual diagnosis using mpMRI has been criticized due to the variability in radiological assessment and human error. In contrast, quantitative predictive models provide user-independent predictions, and address the limitations of direct radiological interpretation. In the literature, various Computer Aided Detection (CAD) algorithms have been developed and evaluated for the diagnosis of prostate cancer using mpMRI. These methods can be broadly classified into voxel-wise detection methods and region-wise detection methods. For voxel-wise classification, the mpMRI features are extracted at the voxel level, and each voxel is classified as either cancer or non-cancer. In contrast, for region-wise classification, mpMRI features are extracted and summarized for pre-specified regions of interests (ROIs), and the cancer status of each ROI is evaluated. Lemaître et al. (2015) gave a highly detailed literature review on the classification methods used in existing CAD systems. Methods include but are not limited to linear and nonlinear model-based classifiers (e.g. Langer et al., 2009; Viswanath et al., 2012; Peng et al., 2013), clustering methods (e.g. Tiwari et al., 2007; Niaf et al., 2012), kernel methods (e.g. Kelm et al., 2007; Ozer et al., 2010; Lopes et al., 2011), naive Bayes (e.g. Mazzetti et al., 2011; Giannini et al., 2013), neural network (e.g. Viswanath et al., 2011; Parfait et al., 2012; Matulewicz et al., 2014), ensemble learning approaches (e.g. Tiwari et al., 2009, 2013; Litjens et al., 2014), and graphical model classifiers (e.g. Liu et al., 2009; Artan et al., 2010). CAD algorithms utilizing textural features and radiomics-based models have also been discussed (Fehr et al., 2015; Cameron et al., 2016; Khalvati et al., 2015). Other related work include Lv et al. (2009); Puech et al. (2009); Shah et al. (2012); Vos et al. (2012), among others.

Currently existing CAD algorithms have shown promising performance. However, some unique features of mpMRI data have not been explored, including heterogeneity in

both the distribution of mpMRI parameters and the voxel-wise cancer prevalence between different regions of the prostate, the spatial correlation between voxels, and the between-patient heterogeneity in the distribution of mpMRI parameters. These features have been previously noted in the literature (e.g. Metzger et al. 2016). However, there is little work on systematic, fully-automated modeling of the specific mpMRI features for improved cancer classification. In this dissertation, we propose to fill these gaps in the literature by developing novel classifiers that take into account the complex data structures of mpMRI.

1.2 Voxel-wise MpMRI Data

This dissertation was motivated by an mpMRI data set obtained from patients that were diagnosed with prostate cancer, the collection procedure for which was described in detail by Metzger et al. (2016). Briefly, the data were collected on a clinical 3T scanner. T2-weighted, diffusion-weighted, and contrast-enhanced images (T2WI, DWI and DCE-MRI, respectively) were acquired in accordance with Prostate Imaging – Reporting and Data System (PI-RADS) v2 guidelines (Weinreb et al., 2016). Maps of the quantitative MRI parameters were then calculated from these data. T2 maps were calculated from T2W turbo spin echo data sets acquired at multiple echo times using methods that were previously developed and validated (Liney et al., 1996; Gibbs et al., 2001). Apparent diffusion coefficient (ADC) maps were calculated from the DWI data acquired using multiple diffusion-encoding b-values. Maps of DCE-MRI parameters, including the area under the gadolinium concentration time curve at 90 seconds (AUGC90), the forward volume transfer constant (K^{trans}), and reflux rate constant (k_{ep}), were generated using a previously developed modified Tofts model (Tofts, 1997). Maps of these quantitative parameters were then manually co-registered. Patients that were imaged subsequently underwent radical prostatectomy, and the ex vivo prostate specimens were collected and processed after surgery, as described in Metzger et al. (2016). The histopathology slides were annotated for cancer by

trained pathologists, then co-registered with the quantitative MR maps using a previously-described registration method (Kalavagunta et al., 2015).

In the original mpMRI data set, 46 images of prostate slices were collected from 34 patients, with 2098 to 5756 voxels per image. Given that only a small number of patients have images of multiple prostate slices, we randomly select one slice from each of these patients to avoid having to model the correlation between slices from the same patient. Based on certain anatomic features, each prostate can be segmented into two main regions, the peripheral zone (PZ), which refers to the portion of the prostate gland that surrounds the distal urethra, and central gland (CG), which is the center of the prostate that contains multiple smaller zones, including the transition, central, anterior fibromuscular, and peri-urethral zones (Myers, 2000; Wasserman et al., 2015). The whole gland (WG) refers to the combination of the PZ and CG. Figure 1.1 illustrates an example of the prostate image showing the two regions of the prostate and areas of cancer and noncancer.

Previously, Metzger et al. (2016) developed an mpMRI model that generated a voxel-wise composite biomarker score (CBS) for the classification of prostate cancer. The CBS model linearly combined multiple mpMRI parameters and showed significantly improved prostate cancer classification compared to using any single mpMRI parameter. An important finding of their work is that, while combining multiple parameters improves classification, there is substantial heterogeneity in the distribution of mpMRI parameters both within and between the PZ and CG and, in addition, large variability exists in cancer prevalence throughout the whole prostate gland, with a much higher prevalence in PZ than in CG. This motivated the region-specific classifiers in Metzger et al. (2016), which were shown to outperform the WG model in terms of classification accuracy. However, manually guided region segmentation is required to obtain the voxel-wise region indicators, which requires human interaction with the data and may be an additional source of human error.

In addition to the region-level heterogeneity, mpMRI data also show substantial spatial correlation in the voxel-wise cancer status. Between-patient heterogeneity is also observed

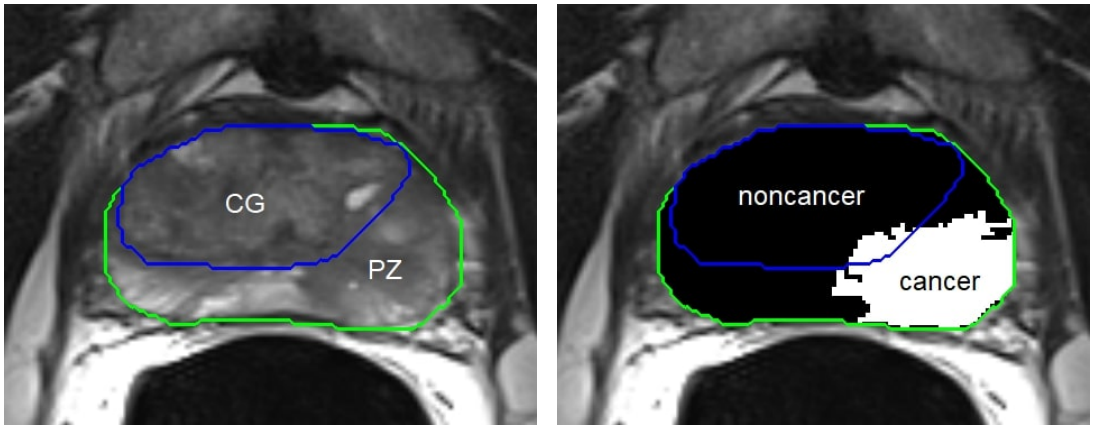

Figure 1.1: An example of the manually guided segmentation of T2-weighted anatomic prostate image in the mpMRI data set. The prostate gland is the area within the green curve. The blue curve demarcates the division between the peripheral zone (PZ) and central gland (CG). Histopathologically identified cancer and noncancer regions are indicated as the white and black areas, respectively, in the second sub-figure.

in the mpMRI parameters, which is possibly due to registration error or variability in patients' physical conditions. These features have not been previously investigated in the context of fully automated, voxel-wise classification of prostate cancer status. Properly modeling the various sources of variability could further improve cancer classification.

1.3 Dissertation Objectives

This dissertation aims to improve the voxel-wise classification of prostate cancer by modeling the aforementioned mpMRI features. The goal was accomplished step-by-step with the following objectives:

(1) propose a general Bayesian hierarchical modeling framework for voxel-wise cancer classification, which provides a natural hierarchical modeling structure for modeling complex mpMRI data structures; based on the general modeling framework, develop a fully-automated classifier that accounts for the region-level variability in mpMRI;

(2) under the proposed Bayesian hierarchical modeling framework, develop computationally efficient modeling approaches for the spatial correlation between voxels, as well as the subject specific effects for patients;

(3) Propose a multi-resolution ensemble learning approach for cancer classification, which addresses some limitations of our proposed Bayesian methods, and can easily be extended to ordered cancer outcomes.

In Chapters 2-4, we will propose solutions to the three objectives in order. Before that, a brief summary of each chapter will be presented in Sections 1.4-1.6.

1.4 Bayesian Hierarchical Modeling Framework and the Modeling of Regional Heterogeneity in MpMRI

Chapter 2 presents a general Bayesian hierarchical modeling framework that allows complex distributional assumptions for both the predictors (mpMRI parameters) and the outcome (cancer status), which is the basis for the developments in Chapters 2 and 3. In addition, we also discuss the modeling of region-level variability in the mpMRI parameters and voxel-wise cancer risk.

Previously, it has been noted that cancer classification could be improved by developing classifiers separately in each region of the prostate (Litjens et al., 2012, 2014; Lemaître et al., 2015). The drawback of a region-specific modeling technique is that it requires region segmentation, which typically would be manually guided, and thus might introduce additional human error.

To overcome these shortcomings, we propose a fully-automated, voxel-wise classifier, which accounts for regional heterogeneity in the mpMRI parameters and cancer risk without the need for manual segmentation. The classifier was built upon a Bayesian hierarchical modeling framework and inspired by the work of Wang et al. (2015). It jointly models the

distribution of voxel-wise mpMRI parameters and cancer status, and discriminant analysis is conducted for cancer classification using the posterior predictive cancer probabilities. Two models are proposed to account for the regional heterogeneity in mpMRI. First, we consider an approach that models the joint distribution of the mpMRI parameters and cancer status as a function of the region indicator. To replace manual region segmentation, automatic segmentation of the prostate is conducted using standardized 2-D coordinates. In the second approach, we directly model the joint distribution as a function of the standardized 2-D coordinate. In addition, we propose to apply a Spatial Gaussian Kernel smoothing technique to account for the spatial variability in the data. Simulation studies and application to the motivating data set indicate that both models improve classification in terms of the area under the Receiver operating characteristic (ROC) curve, i.e. the AUC, and the sensitivity corresponding to 90% specificity (S90), compared to a baseline model, which does not account for the region-level variability.

1.5 Scalable Modeling of Spatial Correlation and Between-Patient Variability

In Chapter 3, we investigate how to model the spatial correlation between voxel-wise cancer status in high-dimensional images, and variability in the distribution of mpMRI parameters between patients, based on the Bayesian hierarchical model framework proposed in Chapter 2.

Modeling the spatial correlation structure in the voxel-wise mpMRI data has not been previously discussed in the context of voxel-wise cancer classification. The major challenge is the extremely high dimensionality of the data. For an image that has n voxels, spatial modeling involves inverting large spatial covariance matrices that typically requires $O(n^3)$ operations and storage of order n^2. Typically, an mpMRI image is comprised of sev-

eral thousand voxels. In the case of our motivating mpMRI data set, simultaneous modeling of the multiple images, each containing 2098 to 5756 voxels, is computationally infeasible. Dimension reduction techniques are required to develop spatial modeling approaches that are applicable to the mpMRI data.

There are two frequently used approaches to modeling large spatial data sets. One is sparse approximation of the spatial covariance or precision matrix, such as covariance tapering assuming compactly supported covariance functions (e.g. Furrer et al., 2006), and approximating the likelihood by the product of lower dimensional conditional densities, Markov random fields, or composite likelihoods (e.g. Vecchia, 1988; Rue and Held, 2005; Eidsvik et al., 2014). Recently, Nearest Neighbor Gaussian Process (NNGP) for high-dimensional spatial modeling was proposed by Datta et al. (2016) for fully process-based inference on the underlying spatial process, which was shown to outperform existing methodologies in terms of inference and scalability. An alternative approach is reduced-rank approximation, which reduces the dimension of the original spatial process by constructing a new process on a lower-dimensional subspace (e.g. Higdon, 2002; Banerjee et al., 2008; Cressie and Johannesson, 2008; Finley et al., 2009).

Based upon the Bayesian hierarchical modeling framework proposed in Chapter 2, we propose several novel classifiers that systematically model the spatial correlation in the voxel-wise cancer status and the between-patient heterogeneity in the mpMRI parameters. The between-patient heterogeneity in the mpMRI parameters is modeled by a subject specific random intercept for the mpMRI parameters. Three computationally efficient approaches are proposed for the modeling of spatial correlation in the voxel-wise cancer status. The first approach utilizes the NNGP model proposed by Datta et al. (2016). The second approach utilizes a knot-based, reduced-rank approximation (Banerjee et al., 2008). The third approach considers a user-defined, fixed correlation structure using the commonly implemented conditional autoregressive (CAR) model. Simulation studies show that the proposed classifiers improve classification both by modeling the between-patient

heterogeneity and by modeling the spatial correlation between voxels. Application to our motivating data set illustrates improvement in classification from modeling the spatial correlation using NNGP and reduced-rank approximation, but not the CAR model. In addition, modeling between-patient heterogeneity did not further improve classification. The NNGP-based classifier is recommended for future application, considering its best overall performance, robust classification results, and relatively fast computation.

1.6 Super Learner Algorithm with a Multi-resolution Modeling Technique for General Cancer Classification Problems

The Bayesian hierarchical models proposed in Chapters 2 and 3 provide computationally feasible approaches to Bayesian classification that account for the regional and spatial heterogeneity in mpMRI. However, they still show limitations and are computationally intensive, in practice. In addition, extension of the proposed Bayesian models to more complex classification/prediction problems can be difficult and may require even longer computing time.

In Chapter 4, we propose a machine learning-based approach for classifying voxel-wise cancer status, which utilizes the super learner algorithm to account for the regional difference in mpMRI. We first select a base learner (a statistical classification model or a machine learning classification algorithm). Considering the difference both between and within the PZ and CG, we propose to segment the prostate images into multi-resolution sub-regions, train the base learner separately in each sub-region of different resolutions, then conduct ensemble learning to combine the multi-resolution base learners. To account for the spatial correlation between voxels and reduce random noise in the data, we apply a spatial Gaussian kernel smoother to the summary statistics for the classification of the

multi-resolution base learners, which will be treated as the covariates for the stage-two model in the proposed super learner algorithm. The spatial smoothing approach is used to reduce computational burden over formally modeling the spatial correlation structure. Both simulation studies and implementation to the motivating data set show that the proposed algorithm improves classification through both the ensemble learning step, and the spatial smoothing step.

An advantage of our proposed ensemble learning approach is that it can be easily extended to more complex voxel-wise classification problems. As an illustration, we also explore classification of an ordered categorical outcome, the clinical significance level of prostate cancer, using the proposed algorithm. Simulation studies and application to the motivating data set show that the classification is improved by both the proposed multi-resolution modeling technique and spatial smoothing. A limitation of our approach is that distinguishing intermediate categories (clinically insignificant cancer v.s. non-cancer or clinically significant cancer) can be difficult due to the low prevalence of clinically insignificant disease in our data. To deal with this issue, we propose to add a weight on each voxel that is inversely proportional to the prevalence of the corresponding cancer category, to up-weight the less prevalent categories. This weighted likelihood approach has been shown to enable our proposed algorithm to identify some clinically insignificant cancer voxels, but with a price of poorer classification for the noncancer category.

Chapter 2

Bayesian Hierarchical Modeling Framework for the Voxel-wise Classification of Prostate Cancer Accounting for Regional Heterogeneity in mpMRI

2.1 Introduction

In the last several decades, mpMRI has been continuously developed and enhanced as a useful tool for the diagnosis and prognosis of prostate cancer (Kurhanewicz et al., 2008; Dickinson et al., 2011). Although mpMRI has shown great potential for prostate cancer examination, its application in clinical practice is limited due to substantial variability of the standard radiological assessment, even with the development of the PI-RADS standard (Barentsz et al., 2012; Rosenkrantz et al., 2013; Garcia-Reyes et al., 2015). In contrast, predictive models provide experience independent identification and possibly classification of the disease. In addition, statistical models based on the mpMRI data provide a formalized framework for combining the mpMRI data in a single predictor, a feature that is only accommodated in a limited form in the most recent Pi-RADS standards, and thus address some limitations of direct radiological interpretation. In the literature, a wide variety of

Computer Aided Detection (CAD) algorithms for the voxel-wise prostate cancer classification utilizing mpMRI parameters have been proposed, including linear and nonlinear regressions, clustering methods, support vector machine, neural network, ensemble learning, Naive Bayes, graphical models, textural feature models, etc. (Kelm et al., 2007; Tiwari et al., 2007; Langer et al., 2009; Liu et al., 2009; Lv et al., 2009; Puech et al., 2009; Tiwari et al., 2009; Artan et al., 2010; Ozer et al., 2010; Lopes et al., 2011; Mazzetti et al., 2011; Viswanath et al., 2011; Niaf et al., 2012; Parfait et al., 2012; Shah et al., 2012; Viswanath et al., 2012; Vos et al., 2012; Giannini et al., 2013; Peng et al., 2013; Tiwari et al., 2013; Litjens et al., 2014; Matulewicz et al., 2014; Fehr et al., 2015; Khalvati et al., 2015; Cameron et al., 2016). However, the existing methodologies tend to assume independency between voxels for scalability and treat the mpMRI parameters as fixed covariates, and rarely make use of other informative features of mpMRI.

Based on the anatomic structures, the whole gland (WG) of a prostate can be segmented into two main regions: the peripheral zone (PZ), which refers to the sub-capsular portion of the posterior aspect of the prostate gland that surrounds the distal urethra, and central gland (CG), which is comprised of multiple structures including the transition, central, anterior fibromuscular, and periurethral zones (Myers, 2000; Wasserman et al., 2015).

Previously, Metzger et al. (2016) developed an mpMRI model that generated a voxel-wise composite biomarker score (CBS) for the voxel-wise classification of prostate cancer. The CBS model linearly combines multiple mpMRI parameters and showed significantly improved prostate cancer classification compared to using any single parameter. An important finding is that, while combining multiple parameters improves classification, there is substantial heterogeneity in the distribution of mpMRI parameters both between and within the PZ and CG and, in addition, substantial variability exists in cancer prevalence throughout the WG with a much higher prevalence observed in the PZ. In this sense, the region (i.e. PZ or CG) meets the classic definition of a confounder, in that the region is associated both with the predictors (mpMRI parameters) and the outcome (voxel-wise cancer occurrence).

Previous authors have noted that the accuracy of voxel-wise classifiers for prostate cancer could be improved by developing region-specific CAD algorithms, as compared to a single, WG CAD algorithm that ignores the differences between regions (Litjens et al., 2012, 2014; Lemaître et al., 2015). In region-specific CAD, classifiers are built separately for different regions of the prostate. However, this requires an additional step of region segmentation, which typically needs manual guidance from clinicians with expertise. Although techniques for automatic segmentation have been investigated (Liu and Yetik, 2011; Toth et al., 2013), few publications have discussed how to generate a fully-automated classifier to model this specific anatomic structure of the prostate without the need for an additional manually guided, region segmentation step.

In this chapter, we propose a fully automated, voxel-wise classifier, which aims to improve the classification of prostate cancer by accounting for the regional heterogeneity in both the distribution of mpMRI parameters and cancer risk without the need for manual segmentation. We will present two approaches to accounting for the regional heterogeneity: a region-specific model, which models the joint distribution of the voxel-wise mpMRI parameters and cancer status as a function of region, and uses the standardized 2-D coordinates to segment the prostate; and a coordinate model, which ignores the region information and directly models the joint distribution of the mpMRI parameters and cancer status as a function of the coordinate. Simulation studies and application to the motivating data set indicate that both approaches improve the area under the ROC curve (AUC) and sensitivity corresponding to 90% specificity (S90) compared to a baseline model that does not account for the regional heterogeneity in mpMRI. In addition, we applied a spatial Gaussian kernel smoother to the voxel-wise posterior predictive cancer probabilities, to account for residual spatial correlation and reduce random noise, which also significantly improved the AUC and S90. The posterior predictive cancer probability is available in closed form, and thus high computational efficiency is achieved by avoiding the need to conduct Markov chain Monte Carlo (MCMC) algorithm to approximate posterior distribution.

The remaining chapter is organized as follows. In Section 2.2, we briefly describe the mpMRI data set and notations for the variables and model parameters. In Section 2.3, we present our analytic approach to developing the voxel-wise classifiers that account for the regional heterogeneity in mpMRI. In Section 2.4, we present simulation studies as well as the application to the motivating data set to evaluate the performance of the proposed classifiers. We conclude with a discussion of the model properties and potential extensions in Section 2.5.

2.2 Data Overview and Notations

We begin by giving an overview of the data set that motivated this dissertation, as well as the general notations that will be used throughout the chapter. Collection procedure of the quantitative mpMRI parameters and histology data used to train and test our classifiers have been described in Metzger et al. (2016). Briefly, as introduced in Section 1.2, the data were collected on a clinical 3T scanner. T2WI, DWI and DCE-MRI data were acquired in accordance with PI-RADS v2 guidelines (Weinreb et al., 2016). Maps of the quantitative MRI parameters were then calculated from these data and manually co-registered. Patients that were imaged subsequently underwent radical prostatectomy, and the ex vivo prostate specimens were collected and processed after surgery. The histopathology slides were annotated for cancer by trained pathologists, then co-registered with the quantitative MR maps using a registration method described in Kalavagunta et al. (2015). The data consists of 46 prostate slices obtained from 34 patients diagnosed with prostate cancer, where there are 2098 to 5756 voxels per slice. For simplicity, we randomly select one slice from each patient that has multiple slices to avoid having to model the correlation between slices from the same patient. Figure 2.1 shows the image of an example slice of prostate.

Let N be the total number of subjects, i.e. prostate slices, and n_i the number of voxels in the i^{th} slice, with $i = 1, \ldots, N$. For the j^{th} voxel in the i^{th} slice, we observe a $d \times 1$

vector of mpMRI parameters, which we denote as $\boldsymbol{y_{ij}} = (y_{ij,1}, ..., y_{ij,d})^T$. The available

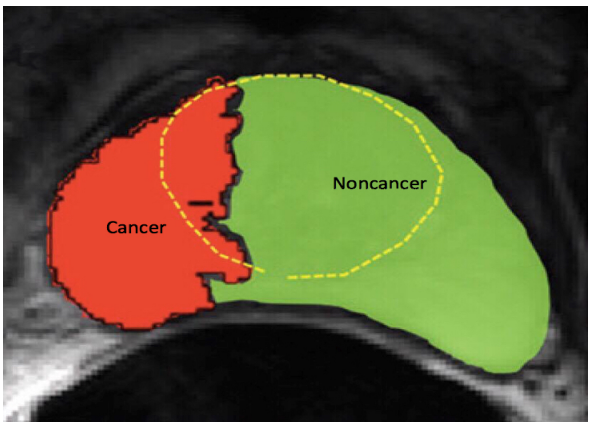

Figure 2.1: An example image showing the prostate capsule, with the yellow dashed curve demarcating the division between peripheral zone (PZ, the area inside the curve) and central gland (CG, the area ouside the curve). Histopathologically identified cancer and noncancer voxels are indicated in the PZ and CG by red and green, respectively.

mpMRI parameters include calculated T2 values, the apparent diffusion coefficient (ADC), and dynamic contrast enhanced MRI (DCE-MRI) parameters, including the area under the gadolinium concentration time curve at 90 seconds (AUGC90), the forward volume transfer constant (K^{trans}), reflux rate constant (k_{ep}) and fractional extravascular extracellular space (V_e). Each voxel also has a binary indicator of its cancer status derived from co-registered pathology data, which we denote as c_{ij}, where $c_{ij} = 1$ indicates cancer and $c_{ij} = 0$ indicates noncancer. This ground truth will be used for the training and evaluation of the cancer classifiers. In addition, each voxel is identified by a set of location information $k_{ij} = \{r_{ij}, \boldsymbol{s_{ij}}\}$, including a binary indicator of region, r_{ij}, with $r_{ij} = 1$ indicating PZ and $r_{ij} = 0$ indicating CG, and a standardized 2-D coordinate, $\boldsymbol{s_{ij}}$, which has been re-scaled to have support $(-1, 1) \times (-1, 1)$, with $(0, 0)$ being the center of each prostate slice.

We denote the set of observed mpMRI parameters of all voxels as $\boldsymbol{Y} = \{\boldsymbol{y_{ij}} | i = 1, \ldots, N, j = 1, \ldots, n_i\}$, and their corresponding cancer status as $\boldsymbol{C} = \{c_{ij} | i = 1, \ldots, N, j = 1, \ldots, n_i\}$. Similarly, we denote the generic parameter set that provides the location infor-

mation for the voxels as $\boldsymbol{K} = \{\boldsymbol{R}, \boldsymbol{S}\}$, with $\boldsymbol{R} = \{r_{ij}|i = 1, \ldots, N, j = 1, \ldots, n_i\}$, and $\boldsymbol{S} = \{\boldsymbol{s_{ij}}|i = 1, \ldots, N, j = 1, \ldots, n_i\}$ denoting the set of region indicators and standardized 2-D coordinates, respectively. Finally, we let $\boldsymbol{Y}^*$, $\boldsymbol{C}^*$ and $\boldsymbol{K}^* = \{\boldsymbol{R}^*, \boldsymbol{S}^*\}$ denote the set of mpMRI parameters, cancer status and location information, respectively, for a new prostate slice whose voxel-wise cancer status are desired.

2.3 Methods

In this section, we discuss the development of prostate cancer classifiers that account for the regional heterogeneity in mpMRI. We first introduce a general Bayesian modeling framework, which models the joint distribution of the voxel-wise mpMRI parameters and cancer risk as a function of the location in a prostate. Cancer classifications for future observations are based on the Bayesian posterior predictive probabilities. Two models will be proposed to incorporate location information: "Mregion", which models the joint distribution of the mpMRI parameters and cancer status by region of the prostate, and "Mcoord", which models the joint distribution by the standardized 2-D coordinate. Both approaches can be combined with post-hoc spatial smoothing to account for residual spatial correlation and random noise. A baseline model will also be presented, which ignores the location information and serves as a baseline for evaluating the gain in classification accuracy from modeling the regional heterogeneity in mpMRI.

2.3.1 General Model Framework

Here we present our general framework for developing voxel-wise classifiers that account for regional heterogeneity in the distribution of the mpMRI parameters and cancer status. We developed our modeling framework by modifying the framework proposed by Wang et al. (2015), which was used for region-wise classification of metastatic liver cancer. Our basic idea is to jointly model the mpMRI parameters, $\boldsymbol{Y}$, and the corresponding cancer

status, $\boldsymbol{C}$, conditional on the location information, $\boldsymbol{K}$, and use the Bayesian posterior predictive cancer probabilities conditional on the observed mpMRI parameters to classify the unknown cancer status for voxels in a new prostate slice. This Bayesian approach in the discriminant analysis setting has been previously pursued (Fraley and Raftery, 2002; Zhou et al., 2013; McLachlan, 2004; Franke et al., 2015; Hamsici and Martinez, 2008).

First, note that the joint model for $\boldsymbol{Y}$ and $\boldsymbol{C}$ conditional on $\boldsymbol{K}$ can be defined hierarchically as $f(\boldsymbol{Y}, \boldsymbol{C}|\boldsymbol{K}) = f(\boldsymbol{Y}|\boldsymbol{C}, \boldsymbol{K})p(\boldsymbol{C}|\boldsymbol{K})$. The two components of the likelihood are defined as follows. First, we assume that the $\boldsymbol{y_{ij}}$'s are independent given c_{ij}'s and k_{ij}'s:

$$\boldsymbol{y_{ij}}|c_{ij}, k_{ij} \overset{ind}{\sim} f(\boldsymbol{y_{ij}}|\boldsymbol{\theta}(c_{ij}, k_{ij})), \tag{2.1}$$

where $\boldsymbol{\theta}(c_{ij}, k_{ij})$ denotes the set of model parameters that are dependent on the cancer status c_{ij} and location information k_{ij}. The cancer status indicators, c_{ij}'s, are assumed to be independent Bernoulli random variables with voxel-wise cancer probability $\rho(k_{ij})$, which is also a function of k_{ij}. In other words, $p(c_{ij}|k_{ij}) = \rho(k_{ij})^{c_{ij}}(1 - \rho(k_{ij}))^{1-c_{ij}}$. The joint likelihood then becomes:

$$f(\boldsymbol{Y}, \boldsymbol{C}|\boldsymbol{K}, \boldsymbol{\theta}(\boldsymbol{C}, \boldsymbol{K}), \rho(\boldsymbol{K})) = \prod_{i=1}^{N}\prod_{j=1}^{n_i}\Big\{f(\boldsymbol{y_{ij}}|\boldsymbol{\theta}(c_{ij}, k_{ij}))p(c_{ij}|k_{ij})\Big\}. \tag{2.2}$$

Model inference is completed under the Bayesian paradigm. Specifically, we assign a prior distribution to $\boldsymbol{\theta}(\boldsymbol{C}, \boldsymbol{K})$, and $\rho(\boldsymbol{K})$ is estimated using an Empirical Bayesian (EB) approach. The cancer status of the voxels in a new prostate slice can be predicted using the posterior predictive cancer probabilities conditional on the observed mpMRI parameters $\boldsymbol{y^*}$, which can be expressed as:

$$p(c^* = 1|\boldsymbol{y^*}, k^*, \boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{K}) = \frac{f(\boldsymbol{y^*}, c^* = 1|k^*, \boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{K})}{\sum_{c=0}^{1} f(\boldsymbol{y^*}, c^* = c|k^*, \boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{K})}, \tag{2.3}$$

where $f(\boldsymbol{y^*}, c^*|k^*, \boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{K})$ is the joint posterior predictive of $\boldsymbol{y^*}$ and c^* conditional on the training data and location information for the new voxel, which is given by:

$$f(\boldsymbol{y^*}, c^*|k^*, \mathbf{Y}, \boldsymbol{C}, \boldsymbol{K}) = \Big\{ \int f(\boldsymbol{y^*}|\boldsymbol{\theta}(c^*, k^*)) f(\boldsymbol{\theta}(c^*, k^*)|\boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{K}) d\boldsymbol{\theta}(c^*, k^*) \Big\} \cdot p(c^*|k^*), \tag{2.4}$$

and $f(\boldsymbol{\theta}(c^*, k^*)|\boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{K})$ is the posterior distribution for $\boldsymbol{\theta}(c^*, k^*)$, which is given by:

$$f(\boldsymbol{\theta}(c^*, k^*)|\boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{K}) = \frac{f(\boldsymbol{Y}|\boldsymbol{\theta}(c^*, k^*), \boldsymbol{C}, \boldsymbol{K}) \pi(\boldsymbol{\theta}(c^*, k^*))}{f(\boldsymbol{Y}|\boldsymbol{C}, \boldsymbol{K})}. \tag{2.5}$$

Here, $\pi(\boldsymbol{\theta}(c^*, k^*))$ denotes the prior distribution for $\boldsymbol{\theta}(c^*, k^*)$, and the denominator, $f(\boldsymbol{Y}|\boldsymbol{C}, \boldsymbol{K}) = \int f(\boldsymbol{Y}|\boldsymbol{\theta}(c^*, k^*), \boldsymbol{C}, \boldsymbol{K}) \pi(\boldsymbol{\theta}(c^*, k^*)) d\boldsymbol{\theta}(c^*, k^*)$, is the marginal likelihood averaged over the prior for $\boldsymbol{\theta}(c^*, k^*)$ (Carlin and Louis, 2008).

2.3.2 Regional Model (Mregion)

Now we propose the first model, "Mregion", which assumes that the joint distribution of the voxel-wise mpMRI parameters and cancer status varies by the binary indicator for region. In this case, the location information, $\boldsymbol{K}$, will refer to the set of voxel-wise regional indicators, $\boldsymbol{R}$, and the two components of the joint likelihood are both region-specific: $f(\boldsymbol{Y}, \boldsymbol{C}|\boldsymbol{R}) = f(\boldsymbol{Y}|\boldsymbol{C}, \boldsymbol{R}) p(\boldsymbol{C}|\boldsymbol{R})$. In addition, we model the probability that a voxel is in either PZ or CG as a function of its standardized 2-D coordinate, to avoid manually guided segmentation for the prostate. The full hierarchical model can then be written as $f(\boldsymbol{Y}, \boldsymbol{C}|\boldsymbol{R}, \boldsymbol{S}) = f(\boldsymbol{Y}|\boldsymbol{C}, \boldsymbol{R}) p(\boldsymbol{C}|\boldsymbol{R}) p(\boldsymbol{R}|\boldsymbol{S})$.

We first discuss the modeling framework assuming $\boldsymbol{R}$ is known, and then discuss the prediction of $\boldsymbol{R}$ using $\boldsymbol{S}$ later in the sub-section. With $\boldsymbol{R}$ known, for $f(\boldsymbol{Y}|\boldsymbol{C}, \boldsymbol{R})$, we assume that the $\boldsymbol{y_{ij}}$'s are independently distributed given the c_{ij}'s and follow a multivariate

normal distribution:

$$\boldsymbol{y_{ij}}|c_{ij}, r_{ij} \overset{ind}{\sim} \mathcal{MVN}(\boldsymbol{\mu_{c_{ij},r_{ij}}}, \boldsymbol{\Sigma_{c_{ij},r_{ij}}}), \tag{2.6}$$

where the mean and covariance parameters, $\boldsymbol{\mu_{c,r}}$ and $\boldsymbol{\Sigma_{c,r}}$, are functions of the cancer status, c, and region indicator, r. The set of model parameters in the multivariate normal distribution is then $\boldsymbol{\theta_{c,r}} = \{\boldsymbol{\mu_{c,r}}, \boldsymbol{\Sigma_{c,r}}|c, r \in \{0, 1\}\}$. Similarly, for the specification of $p(\boldsymbol{C}|\boldsymbol{R})$, the c_{ij}'s are assumed to be independent Bernoulli random variables with probability, ρ_r, which is also region-specific.

The prior distribution for the model parameters $\boldsymbol{\theta_{c,r}}$ is specified as follows. We assume a priori independence of the model parameters, given cancer status and region indicator. Conditional on cancer status $c \in \{0, 1\}$ and region indicator $r \in \{0, 1\}$, we assign conjugate priors, specifically, a flat prior on the mean, $\boldsymbol{\mu_{c,r}}$, and an independent inverse Wishart prior, $\mathcal{W}^{-1}(\boldsymbol{\Sigma_{c,r}}|\delta, \boldsymbol{\Omega_{c,r}})$, on the covariance matrix, $\boldsymbol{\Sigma_{c,r}}$. The hyperparameter δ is set to $m + 1$ to provide maximum entropy, while the hyperparameter $\boldsymbol{\Omega_{c,r}}$ is determined by an EB approach, where $\boldsymbol{\Omega_{c,r}}$ is set equal to the maximizer of the marginal likelihood, $f(\boldsymbol{Y}|\boldsymbol{C}, \boldsymbol{R}, \boldsymbol{\Omega_{c,r}}, \delta = m + 1)$. Rather than placing a prior on ρ_r, $r \in \{0, 1\}$, we instead estimate ρ_r using an EB approach, as well, in which case $\hat{\rho}_r$ equals the sample mean of c_{ij}'s across all prostate slices and voxels within region r. This allows us to obtain a closed-form solution for the posterior distribution, so as to avoid MCMC and reduce computational burden for the proposed method.

Combining the likelihood and prior distribution results in the following posterior distribution for $\boldsymbol{\theta_{c,r}}$:

$$\boldsymbol{\mu_{c,r}}, \boldsymbol{\Sigma_{c,r}}|\boldsymbol{Y^{c,r}} \sim \mathcal{MVN}(\boldsymbol{\mu_{c,r}}|\boldsymbol{\tilde{y}^{c,r}}, \frac{\boldsymbol{\Sigma_{c,r}}}{\sum_{i=1}^{N} n_i^{c,r}}) \times \mathcal{W}^{-1}(\boldsymbol{\Sigma_{c,r}}|\sum_{i=1}^{N} n_i^{c,r} + \delta, \boldsymbol{\tilde{S}^{c,r}} + \boldsymbol{\hat{\Omega}_{c,r}}), \tag{2.7}$$

where $c, r \in \{0, 1\}$, $\boldsymbol{Y}^{\boldsymbol{c,r}}$ denotes the set of mpMRI parameters for the voxels with cancer status c and region indicator r, $n_i^{c,r}$ denotes the number of voxels in the i^{th} prostate slice that have cancer status c and region indicator r, and the two additional terms, $\tilde{\boldsymbol{y}}^{\boldsymbol{c,r}}$ and $\tilde{\boldsymbol{S}}^{\boldsymbol{c,r}}$, are defined as:

$$\tilde{\boldsymbol{y}}^{\boldsymbol{c,r}} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{n_i^{c,r}} \boldsymbol{y}_{\boldsymbol{ij}}^{\boldsymbol{c,r}}}{\sum_{i=1}^{N} n_i^{c,r}}, \tag{2.8}$$

$$\tilde{\boldsymbol{S}}^{\boldsymbol{c,r}} = \sum_{i=1}^{N} \boldsymbol{Y}_{\boldsymbol{i}}^{\boldsymbol{c,r}} (\boldsymbol{Y}_{\boldsymbol{i}}^{\boldsymbol{c,r}})^T - \sum_{i=1}^{N} \sum_{j=1}^{n_i^{c,r}} \bar{\boldsymbol{y}}^{\boldsymbol{c,r}} \bar{\boldsymbol{y}}^{\boldsymbol{c,r}T}. \tag{2.9}$$

For a voxel in a new prostate slice, the posterior predictive distribution of $\boldsymbol{y^*}$ given cancer status c^* and region r^* can then be written as:

$$\begin{aligned} f(\boldsymbol{y^*}|c^*, r^*, \boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{R}) =& (2\pi)^{-\frac{d}{2}} \Big(\frac{\sum_{i=1}^{N} n_i^{c^*,r^*} + 1}{\sum_{i=1}^{N} n_i^{c^*,r^*}} \Big)^{-\frac{d}{2}} \\ & \times \frac{h(\sum_{i=1}^{N} n_i^{c^*,r^*} + \delta, \tilde{\boldsymbol{S}}^{\boldsymbol{c^*,r^*}} + \hat{\boldsymbol{\Omega}}_{\boldsymbol{c^*,r^*}})}{h(\sum_{i=1}^{N} n_i^{c^*,r^*} + \delta + 1, \tilde{\boldsymbol{S}}_{\boldsymbol{*}}^{\boldsymbol{c^*,r^*}} + \hat{\boldsymbol{\Omega}}_{\boldsymbol{c^*,r^*}})}, \end{aligned} \tag{2.10}$$

where $\tilde{\boldsymbol{S}}_{\boldsymbol{*}}^{\boldsymbol{c^*,r^*}}$ is defined similarly as $\tilde{\boldsymbol{S}}^{\boldsymbol{c,r}}$ in (2.9) but with the new voxel included, and $h(\delta, \boldsymbol{\Omega})$ is the normalizing constant in the probability density function of $\mathcal{W}^{-1}(\boldsymbol{\Sigma}|\delta, \boldsymbol{\Omega})$. Combining the prior cancer risk and posterior predictive of the mpMRI parameters gives the posterior predictive distribution of c^* conditional on $\boldsymbol{y^*}$, r^*, $\boldsymbol{Y}$, $\boldsymbol{C}$, and $\boldsymbol{R}$:

$$p(c^* = 1|\boldsymbol{y^*}, r^*, \boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{R}) = \frac{p(c^* = 1|r^*) f(\boldsymbol{y^*}|c^* = 1, r^*, \boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{R})}{\sum_{c=0}^{1} p(c^* = c|r^*) f(\boldsymbol{y^*}|c^* = c, r^*, \boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{R})}, \tag{2.11}$$

where $p(c^* = 1|r^*) = \hat{\rho}_{r^*}$ is the estimated prior cancer risk in region r^*.

The posterior predictive cancer probability in (2.11) is derived under the assumption that region information $\boldsymbol{R}$ is known. For the motivating data set, region segmentation was conducted on the anatomic T2-weighted images using a segmentation software (Segasist, Ontario, Canada) requiring manual guidance. Our objective is to create a fully automated

classifier, in which case this manually guided region segmentation must be avoided. Given the consistency of prostate anatomy across patients, a rough but automatic estimation for a voxel's regional location can be conducted by estimating the probability that a voxel is in the PZ using standardized 2-D coordinates. The prediction can be performed in multiple ways. Here, we adopt an adaptive polynomial spline regression with the logistic likelihood (De Boor et al., 1978). We estimate the vector of coefficients in the spline regression, $\boldsymbol{\beta}_{\text{coord}}$, using iterative weighted least squares assuming the following model for $p(r = 1|\boldsymbol{s})$:

$$\text{logit}(p(r = 1|\boldsymbol{s})) = \boldsymbol{x}_{\text{coord}}^T(\boldsymbol{s})\boldsymbol{\beta}_{\text{coord}}, \tag{2.12}$$

where $\boldsymbol{x}_{\text{coord}}(\boldsymbol{s})$ is a $d_{\boldsymbol{s}} \times 1$ vector of basis functions of the coordinate $\boldsymbol{s}$, which is determined by a model selection procedure (details will be discussed in Section 2.3). Combining (2.11) and (2.12) results in the posterior predictive cancer probability conditional on $\boldsymbol{y}^*$, $\boldsymbol{s}^*$, $\boldsymbol{Y}$, $\boldsymbol{C}$, and $\boldsymbol{R}$:

$$p(c^* = 1|\boldsymbol{y}^*, \boldsymbol{s}^*, \boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{R}) = \sum_{r=0}^{1} p(c^* = 1|\boldsymbol{y}^*, r^* = r, \boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{R}) p(r^* = r|\boldsymbol{s}^*), \tag{2.13}$$

where $p(r^* = 1|\boldsymbol{s}^*) = e^{\boldsymbol{x}_{\text{coord}}^T(\boldsymbol{s}^*)\hat{\boldsymbol{\beta}}_{\text{coord}}}/(1+e^{\boldsymbol{x}_{\text{coord}}^T(\boldsymbol{s}^*)\hat{\boldsymbol{\beta}}_{\text{coord}}})$, and $p(c^* = 1|\boldsymbol{y}^*, r^* = r, \boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{R})$ is defined in (2.11). This final posterior predictive cancer probability is in fact the weighted average of the posterior predictive cancer probabilities in PZ and CG.

2.3.3 Coordinate Model (Mcoord)

The Mregion model assumes that the joint distribution of voxel-wise mpMRI parameters and cancer status varies by region, but is consistent within each region. However, this model requires the region indicators to be predicted by coordinate to avoid manually guided segmentation. An alternative approach for modeling the regional heterogeneity is to simply

ignore the region information, and directly model the joint distribution of mpMRI parameters and cancer status by the standardized 2-D coordinates. We refer to this model as the "Mcoord" model. Here the location information $\boldsymbol{K}$ is summarized by $\boldsymbol{S}$, and the joint likelihood becomes $f(\boldsymbol{Y}, \boldsymbol{C}|\boldsymbol{S}) = f(\boldsymbol{Y}|\boldsymbol{C}, \boldsymbol{S})p(\boldsymbol{C}|\boldsymbol{S})$.

We again assume that $\boldsymbol{y_{ij}}$'s are independent and follow a multivariate normal distribution with mean and covariance varying by cancer status. We still assume that the mean parameter varies by the location in a prostate, but, unlike the specification in Mregion, the mean is allowed to vary both between and within regions, as a function of the standardized 2-D coordinate. A multivariate adaptive polynomial spline regression is used to describe the coordinate-based mean (De Boor et al., 1978; Kooperberg et al., 1997):

$$\begin{aligned} &\boldsymbol{y_{ij}}|c_{ij}, \boldsymbol{s_{ij}} \overset{ind}{\sim} \mathcal{MVN}(\boldsymbol{\mu_{c_{ij}}}(\boldsymbol{s_{ij}}), \boldsymbol{\Sigma_{c_{ij}}}), \\ &\boldsymbol{\mu}^T_{\boldsymbol{c_{ij}}}(\boldsymbol{s_{ij}}) = \boldsymbol{x}^{\boldsymbol{T}}_{\boldsymbol{c_{ij}}}(\boldsymbol{s_{ij}})\boldsymbol{\beta_{c_{ij}}}, \end{aligned} \tag{2.14}$$

where $\boldsymbol{x_{c_{ij}}}(\boldsymbol{s_{ij}})$ denotes the $d_{c_{ij}} \times 1$ vector of basis functions of the coordinate $\boldsymbol{s_{ij}}$, $\boldsymbol{\beta_{c_{ij}}}$ denotes the $d_{c_{ij}} \times d$ matrix of the corresponding coefficients, with $d_{c_{ij}}$ being the number of basis functions used for cancer status $c_{ij} \in \{0, 1\}$. Note that the mean parameters, $\boldsymbol{\mu_c}(\boldsymbol{s_{ij}})$, $c \in \{0, 1\}$, are no longer constant within each region, and the set of model parameters become $\boldsymbol{\theta_c} = (\boldsymbol{\beta_c}, \boldsymbol{\Sigma_c}|c \in \{0, 1\})$. The cancer status indicators, c_{ij}'s, are still assumed to be independent Bernoulli random variables, but we now allow cancer risk, $\rho(\boldsymbol{s_{ij}})$, to vary by coordinate $\boldsymbol{s_{ij}}$.

We assume a priori independence of $\boldsymbol{\theta_c}$ given c, and specify the prior distribution as follows. We assume a flat prior on the spline regression coefficient, $\boldsymbol{\beta_c}$, and an inverse Wishart prior on $\boldsymbol{\Sigma_c}$. As specified in Section 2.3.2, the hyperparameter δ is set to $m + 1$ to provide maximum entropy, and $\boldsymbol{\Omega_c}$ is determined by an EB approach, where it is set equal to the maximizer of the marginal likelihood $f(\boldsymbol{Y}|\boldsymbol{C}, \boldsymbol{S}, \boldsymbol{\Omega_c}, \delta = m + 1)$. Similar to the strategy used for Mregion model, we estimate the voxel-wise cancer risk using an

EB approach to avoid having to implement MCMC but, in this case, we use an adaptive polynomial spline regression, $\text{logit}(\rho(\boldsymbol{s_{ij}})) = \boldsymbol{x_p}^T(\boldsymbol{s_{ij}})\boldsymbol{\beta_p}$, where $\boldsymbol{x_p}(\boldsymbol{s_{ij}})$ is the $d_p \times 1$ vector of basis functions of $\boldsymbol{s_{ij}}$, and estimate the vector of regression coefficients, $\boldsymbol{\beta_p}$, using iterative weighted least squares.

Combining the likelihood and prior results in the following posterior for $\boldsymbol{\theta_c}$:

$$\boldsymbol{\beta_c}^T, \boldsymbol{\Sigma_c}|\boldsymbol{Y^c} \sim \mathcal{MN}_{d,d_c}(\boldsymbol{\beta_c}^T|\tilde{\boldsymbol{\beta}_{\boldsymbol{c}}}^T, \boldsymbol{\Sigma_c}, \sum_{i=1}^{N}\sum_{j=1}^{n_i^c} \boldsymbol{x_{ij}^c}\boldsymbol{x_{ij}^c}^T) \times \mathcal{W}^{-1}(\boldsymbol{\Sigma_c}|\sum_{i=1}^{N} n_i^c + \delta, \tilde{\boldsymbol{S}}^{\boldsymbol{c}} + \hat{\boldsymbol{\Omega}}_{\boldsymbol{c}}), \tag{2.15}$$

where $\boldsymbol{Y^c}$ denotes the set of mpMRI parmeters for the voxels with cancer status c, $\mathcal{MN}_{d,d_c}$ denotes the matrix normal distribution of a $d \times d_c$ random matrix, and n_i^c denotes the number of voxels in the i^{th} slice that have cancer status c. Here, $\boldsymbol{x_{ij}^c}$ refers to $\boldsymbol{x_c}(\boldsymbol{s_{ij}})$ in (2.14), which is the vector of basis functions in the spline regression. The two additional terms, $\tilde{\boldsymbol{\beta}}_{\boldsymbol{c}}$ and $\tilde{\boldsymbol{S}}^{\boldsymbol{c}}$, are defined as:

$$\tilde{\boldsymbol{\beta}}_{\boldsymbol{c}} = \Big[\sum_{i=1}^{N}\sum_{j=1}^{n_i^c} \boldsymbol{x_{ij}^c}\boldsymbol{x_{ij}^c}^T\Big]^{-1}\Big[\sum_{i=1}^{N}\sum_{j=1}^{n_i^c} \boldsymbol{x_{ij}^c}\boldsymbol{y_{ij}^c}^T\Big], \tag{2.16}$$

$$\tilde{\boldsymbol{S}}^{\boldsymbol{c}} = \sum_{i=1}^{N}\sum_{j=1}^{n_i^c} \boldsymbol{x_{ij}^c}\boldsymbol{x_{ij}^c}^T - \sum_{i=1}^{N}\sum_{j=1}^{n_i^c} \boldsymbol{y_{ij}^c}\boldsymbol{x_{ij}^c}^T\tilde{\boldsymbol{\beta}}_{\boldsymbol{c}}. \tag{2.17}$$

The posterior predictive distribution of the mpMRI parameters for a new voxel, $\boldsymbol{y^*}$, given cancer status c^* and coordinate $\boldsymbol{s^*}$, can then be written as:

$$\begin{aligned} f(\boldsymbol{y^*}|c^*, \boldsymbol{s^*}, \boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{S}) =& (2\pi)^{-\frac{d}{2}}\Big|\Big[\sum_{i=1}^{N}\sum_{j=1}^{n_i^{c^*}} \boldsymbol{x_{ij}^{c^*}}\boldsymbol{x_{ij}^{c^*}}^T\Big]^{-1}\boldsymbol{x_*^{c^*}}\boldsymbol{x_*^{c^*}}^T + \boldsymbol{I}\Big|^{-\frac{d}{2}} \\ & \times \frac{h(\sum_{i=1}^{N} n_i^{c*} + \delta, \tilde{\boldsymbol{S}}^{\boldsymbol{c^*}} + \hat{\boldsymbol{\Omega}}_{\boldsymbol{c^*}})}{h(\sum_{i=1}^{N} n_i^{c^*} + \delta + 1, \tilde{\boldsymbol{S}}_{\boldsymbol{*}}^{\boldsymbol{c^*}} + \hat{\boldsymbol{\Omega}}_{\boldsymbol{c^*}})}, \end{aligned} \tag{2.18}$$

where $\boldsymbol{x_*^{c^*}}$ denotes the vector of basis functions for the new voxel given its cancer status c^*,

and $\tilde{\boldsymbol{S}}_*^{c^*}$ is defined similarly as in (2.17), but with the new voxel included. By combining the estimated prior cancer probability and the posterior predictive of the mpMRI parameters, the posterior predictive cancer probability can be derived for a new voxel:

$$p(c^*=1|\boldsymbol{y}^*,\boldsymbol{s}^*,\boldsymbol{Y},\boldsymbol{C},\boldsymbol{S}) = \frac{p(c^*=1|\boldsymbol{s}^*)f(\boldsymbol{y}^*|c^*=1,\boldsymbol{s}^*,\boldsymbol{Y},\boldsymbol{C},\boldsymbol{S})}{\sum_{c=0}^{1} p(c^*=c|\boldsymbol{s}^*)f(\boldsymbol{y}^*|c^*=c,\boldsymbol{s}^*,\boldsymbol{Y},\boldsymbol{C},\boldsymbol{S})}, \quad (2.19)$$

where $p(c^*=1|\boldsymbol{s}^*) = e^{\boldsymbol{x}_{\boldsymbol{p}}^T(\boldsymbol{s}^*)\hat{\boldsymbol{\beta}}_{\boldsymbol{p}}}/(1+e^{\boldsymbol{x}_{\boldsymbol{p}}^T(\boldsymbol{s}^*)\hat{\boldsymbol{\beta}}_{\boldsymbol{p}}})$ is the estimated prior cancer probability at coordinate $\boldsymbol{s}^*$.

To evaluate the performance of the Mregion and Mcoord models, we also considered an Mbase model, which ignores the location information $\boldsymbol{K}$. Details of the Mbase model can be found in Appendix A.1. The resulting posterior distribution, posterior predictive distribution, and voxel-wise cancer classifier are equivalent to (2.7), (2.10) and (2.11), respectively, except that the parameters in Mbase do not depend on the region information $\boldsymbol{R}$. Comparing Mregion and Mcoord with Mbase helps us evaluate the benefit of modeling the regional heterogeneity in the mpMRI data. Table 2.1 summarizes and compares the

Table 2.1: Comparison of the data components required for model training (columns 2-5) and for prospective classification of cancer status for the voxels in a new prostate slice (columns 6-8) between Mbase, Mregion and Mcoord.

	Model Training				Prospective Classification of Cancer		
	$\boldsymbol{Y}$	$\boldsymbol{C}$	$\boldsymbol{R}$	$\boldsymbol{S}$	$\boldsymbol{Y}^*$	$\boldsymbol{R}^*$	$\boldsymbol{S}^*$
Mbase	✓	✓			✓		
Mregion	✓	✓	✓	✓	✓		✓
Mcoord	✓	✓		✓	✓		✓

Note: * indicates a new prostate slice.

data components required for the model training and classification of the Mbase, Mregion and Mcoord models. For the Mbase model, which does not take into account the regional heterogeneity, voxel-wise location information is not required for either modeling training or classification. In contrast, the Mregion and Mcoord models both require standardized

2-D coordinates to train the classifier and to obtain predictions for future voxels. The primary distinction between the Mregion model and the Mcoord model is that, Mregion requires manually guided region segmentation for the training data to train the classifier, while Mcoord models the joint distribution of mpMRI parameters and cancer status as a function of the coordinates, directly, and does not. Finally, we note that neither Mregion nor Mcoord requires manually guided region segmentation for future data, which is one of our primary objectives.

2.3.4 Classification for the Voxels with Missing mpMRI Parameters

A common problem of the mpMRI data is that some mpMRI parameters may have missing values due to failed MRI parametric mapping, most frequently being a result of uncorrected motion between or within data sets. This is a practical problem that complicates the application of regression-based classifiers: implementing regression-based classifiers in the presence of missing data requires that the missing mpMRI parameters are imputed. In the Bayesian framework, however, we can simply integrate out the missing parameters from the joint posterior predictive distribution of the mpMRI parameters and cancer status:

$$f(\boldsymbol{y}^*_{\boldsymbol{obs}}, c^*|k^*, \boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{K}) = \int f(\boldsymbol{y}^*_{\boldsymbol{obs}}, \boldsymbol{y}^*_{\boldsymbol{mis}}, c^*|k^*, \boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{K}) d\boldsymbol{y}^*_{\boldsymbol{mis}}, \tag{2.20}$$

where $\boldsymbol{y}^* = (\boldsymbol{y}^*_{\boldsymbol{obs}}, \boldsymbol{y}^*_{\boldsymbol{mis}})$, with $\boldsymbol{y}^*_{\boldsymbol{obs}}$ and $\boldsymbol{y}^*_{\boldsymbol{mis}}$ denoting the observed and missing mpMRI parameters, respectively. In this case, voxel-wise classification is conducted by replacing $f(\boldsymbol{y}^*|c^*, k^*, \boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{K})$ with $f(\boldsymbol{y}^*_{\boldsymbol{obs}}|c^*, k^*, \boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{K})$ in (2.11) and (2.19). In other words, if a voxel is missing one or more mpMRI parameters, its cancer status can still be classified using only the observed mpMRI parameters. If a voxel is missing all mpMRI parameters, then its posterior predictive cancer probability is simply estimated as a function of the location in a prostate. This approach assumes missing completely at random (MCAR) (Little and Rubin, 2002), which is a strong and likely untrue assumption, but we will show

that the approach still leads to reasonable classification results on our motivating data set.

2.3.5 Spatial Gaussian Kernel Smoother and Msmooth

Previously, we assumed independency between voxels for scalability. However, there may exist residual spatial correlation between voxels, and taking into account the spatial correlation could potentially further improve the classification. Ideally, we would formally model the spatial correlation structure between voxels. However, modeling the spatial correlation structure in our setting is nontrivial due to the high-dimensionality of the data (RUE and Tjelmeland, 2002; Stein et al., 2004). Instead, we propose to apply a spatial Gaussian-Kernal smoothing technique to ensure computational efficiency. Here we use the Nadaraya-Watson estimator with a Gaussian kernel (Nadaraya, 1964; Watson, 1964). The smoothed posterior predictive cancer probability at coordinate $\boldsymbol{s_{*j}}$ can then be written as:

$$\tilde{\hat{p}}(\boldsymbol{s_{*j}}) = \frac{\sum_{k=1,k\neq j}^{n_*} K_\sigma(||(\boldsymbol{s_{*j}}) - (\boldsymbol{s_{*k}})||_2)\hat{p}(\boldsymbol{s_{*k}})}{\sum_{k=1,k\neq j}^{n_*} K_\sigma(||(\boldsymbol{s_{*j}}) - (\boldsymbol{s_{*k}})||_2)}, \tag{2.21}$$

where n_* denotes the number of voxels in the new prostate slice, $||\cdot||_2$ denotes L_2 norm, K_σ denotes the gaussian kernel with bandwidth σ, and $\hat{p}(\boldsymbol{s_{*k}})$ is the unsmoothed posterior predictive cancer probability for the voxel at coordinate $\boldsymbol{s_{*k}}$ in the new slice. The bandwidth σ controls the contribution of the neighboring voxels (which is also determined by their distance from $\boldsymbol{s_{*j}}$), and is chosen by Cross-Validation to maximize the average AUC. Note that spatial smoothing of the voxels on the edge of a slice requires edge correction. This is realized by replacing the denominator in (2.21) by the inverse of an edge correction term $e(\boldsymbol{s_{*j}})$, which is the reciprocal of the kernel mass within the new prostate slice:

$$e(\boldsymbol{s_{*j}}) = \Big[\int_{v\in G} k(||v - \boldsymbol{s_{*j}}||_2)dv\Big]^{-1}, \tag{2.22}$$

where G denotes the area within the new prostate slice. This spatial Gaussian kernel smoothing technique was implemented using the R package "spatstat" (Baddeley et al., 2005).

2.4 Numerical Results

2.4.1 Simulation Studies

Simulation Settings

Simulation studies were conducted to evaluate the performance of the proposed classifiers in different data scenarios. In each simulation, we generated 34 training prostate slices and 10 test slices. For each slice, a mask (i.e shape of the slice, including voxel-wise region indicators and standardized 2-D coordinates) was sampled with replacement from the 46 prostate slices in the motivating data set. Within each slice, voxel-wise mpMRI parameters and binary cancer indicators were generated according to the following model:

$$\begin{aligned} \boldsymbol{y_{ij}}|c_{ij}, r_{ij} &\sim \mathcal{MVN}(\boldsymbol{y_{ij}}|\boldsymbol{\mu_{c_{ij},r_{ij}}}, \boldsymbol{\Gamma_{c_{ij},r_{ij}}}), \\ c_{ij} &= I(p_{ij} > 0.5) = I(q_{ij} > 0), \\ q_{ij} &= \Phi^{-1}(p_{ij}) = q_{0,r_{ij}} + w_{ij}, \\ \boldsymbol{w_i} &\sim \mathcal{MVN}(\mathbf{0}, \boldsymbol{C_i}(\phi, \sigma^2)). \end{aligned} \tag{2.23}$$

Specifically, for the j^{th} voxel in the i^{th} slice, cancer probability p_{ij} was simulated by adding a spatial random effect, w_{ij}, to the probit (which we denote as $\Phi^{-1}(\cdot)$) of the region-specific cancer probability $q_{0,r_{ij}}$, $r_{ij} \in \{0, 1\}$. To induce spatial correlation between voxel-wise cancer status, we assumed that $\boldsymbol{w_i} = (w_{i1}, w_{i2}, ..., w_{i,n_i})^T$ followed a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{C_i}(\phi, \sigma^2)$, i.e. w_{ij}'s, $j = 1, ..., n_i$, were spatially correlated. The (j, k)-th entry of $\boldsymbol{C_i}(\phi, \sigma^2)$ (i.e. the spatial correlation between

w_{ij} and w_{ik}) was defined as $\boldsymbol{C_i}(\phi, \sigma^2)_{j,k} = \sigma^2 e^{-\phi||\boldsymbol{s_{ij}} - \boldsymbol{s_{ik}}||_2}$, with σ^2 and ϕ being the spatial variance and range parameter, respectively. The binary cancer status, c_{ij}, was set to 1 if $p_{ij} > 0.5$, and 0 if $p_{ij} \leqslant 0.5$. Finally, we allowed the distribution of $\boldsymbol{y_{ij}}$ to vary between the PZ and CG, and between cancer and noncancer voxels, by simulating $\boldsymbol{y_{ij}}$ based on the region indicator r_{ij} and cancer indicator c_{ij}.

Simulations were conducted assuming different scales for the regional heterogeneity and different patterns for the spatial correlation structure. To generate regional heterogeneity in the distribution of mpMRI parameters, we set the difference in mean between the PZ and CG equal to either the difference in the sample mean calculated from the motivating data set, half of the difference in the sample mean, or $\mathbf{0}$, separately for the cancer and noncancer voxels. To induce regional heterogeneity in the voxel-wise cancer probabilities, we fixed the voxel-wise cancer probabilities in the CG to $p_{CG} = 0.092$, which was the sample estimate obtained from the motivating data set, and set the difference in cancer probability between PZ and CG to 0.184, 0.092, or 0. In addition, we set the spatial variance to $\sigma^2 = 30$, 15 or 5, and the spatial correlation to $\phi = 5$, 2 or 1, so as to induce different spatial correlation patterns for the simulated data.

We completed 1000 simulations for each data scenario. In each simulation, we trained the classifiers using 34 training prostate slices, and evaluated the performance using 10 test slices. Classification accuracy of the various classifiers were summarized by the average AUC across the 1000 simulations. In each data scenario, the kernel bandwidth for spatial Gaussian kernel smoothing was treated as a tuning parameter and selected by Leave-one-out Cross-Validation (LOOCV) as the maximizer for average AUC. The basis functions in the spline regressions were determined by stepwise model selection procedure to minimize the penalized residual sum of squares, which was conducted using the R package "polspline" (Kooperberg et al., 1997; Kooperberg and Kooperberg, 2019).

Simulation Results

Figure 2.2 presents the AUC for the Mbase, Mregion, Mcoord and Msmooth models given different scales for the regional heterogeneity, spatial variance, σ^2, and spatial correlation range parameter, ϕ. In the figure, the error bars represent the 95% confidence intervals for the AUC, and the points at the center of the bars show the average AUC. We first look at the performance of Mregion and Mcoord given different scales for the regional heterogeneity. In each sub-figure, when there is no regional heterogeneity, Mregion and Mcoord perform similarly as Mbase. When there is regional heterogeneity, Mregion and Mcoord provide higher average AUC than Mbase, and this advantage increases as the scale of regional heterogeneity increases. This agrees with our expectation: Mregion and Mcoord improve cancer classification by accounting for the regional heterogeneity in both the distribution of mpMRI parameters and cancer risk. Next, we investigate the relative performance of Mcoord and Msmooth given different spatial correlation patterns. In the top left figure where there is large spatial correlation range ($\phi = 5$) and large spatial variance ($\sigma^2 = 30$), the improvement in AUC from Mcoord to Msmooth is large but decreases as the spatial correlation range and spatial variance decrease. When the correlation range parameter is relatively small ($\phi = 1$), spatial Gaussian kernel smoothing results in little improvement in AUC, but does not over-smooth to the point of reducing the AUC of the Mbase model. Detailed simulation results are reported in Appendix A.2.

We also evaluated the performance of the spatial Gaussian kernel smoother where there existed more complex spatial correlation patterns. Specifically, we considered two scenarios. In one, the spatial correlation structure varied by prostate slice, where we simulated data sets with spatial correlation range parameter, ϕ, varying by slice. In two, there was non-stationarity in the spatial correlation structure within each slice, where we varied the spatial correlation range parameter, ϕ, by the location in a prostate slice. Detailed simulation settings and results are summarized in Appendix A.2. Overall, model performance

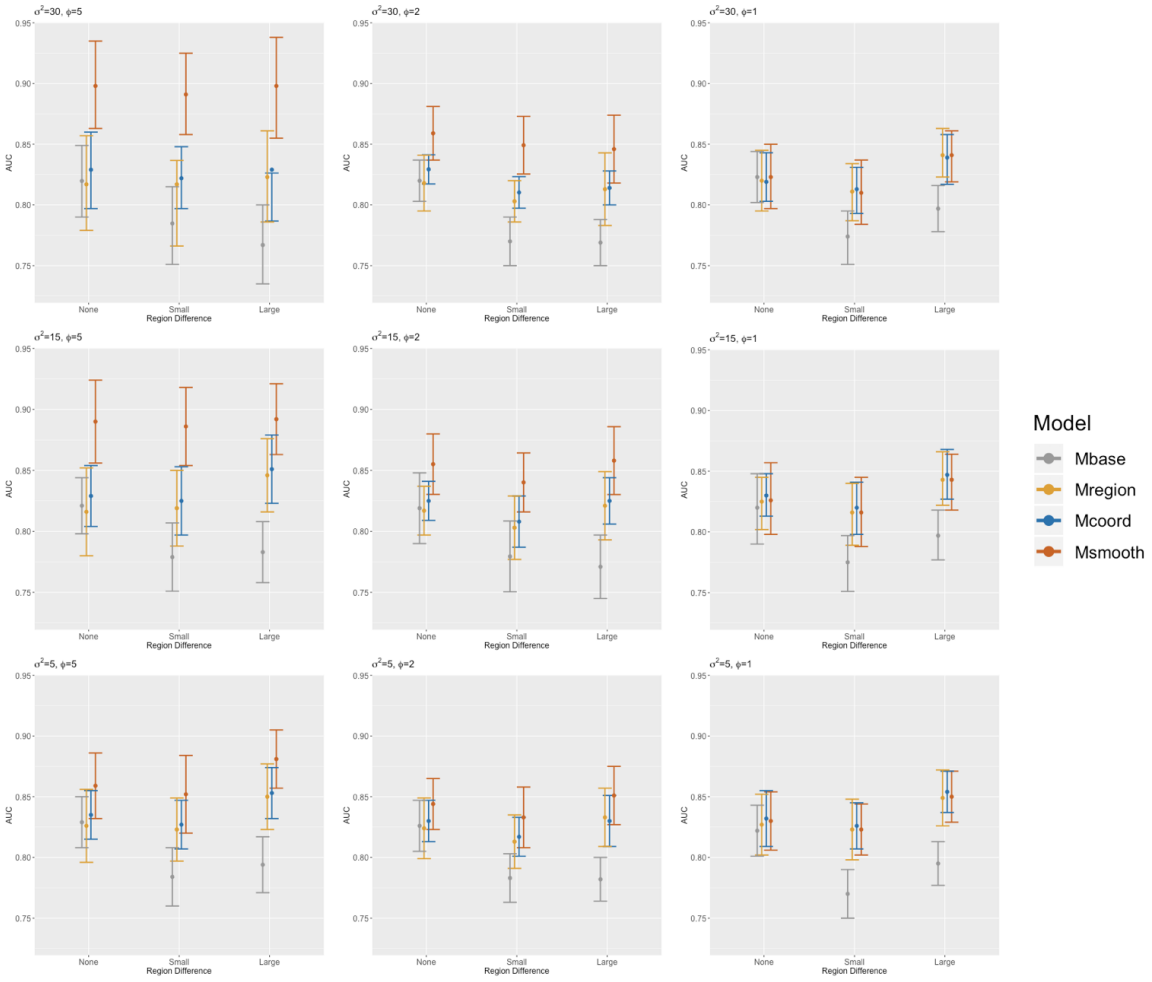

Figure 2.2: Comparison of AUC between Mbase, Mregion, Mcoord and Msmooth, under different settings for the regional heterogeneity, spatial correlation range parameter, ϕ, and spatial variance, σ^2.

observed in these scenarios are similar as observed in the scenarios assuming stationary spatial correlation structure, except that the improvement due to spatial smoothing is lower.

2.4.2 Application to in Vivo Data

Now we evaluate model performance on the motivating data set described in Section 2.2. Previously, Metzger et al. (2016) developed two mpMRI classifiers: (1) a WG classifier,

which used $d = 4$ mpMRI parameters, including ADC, K^{trans}, k_{ep} and AUGC90; and (2) a PZ-specific classifier, which utilized T2 in place of K^{trans}. For the application of our proposed classifiers, different sets of mpMRI parameters were considered, and it was found that the combination of ADC, K^{trans}, k_{ep} and AUGC90 gave the highest classification accuracy and, therefore, these four parameters were used as the mpMRI parameters in all subsequent modeling. Among the four mpMRI parameters, ADC values are approximately normally distributed, while K^{trans}, k_{ep} and AUGC90 values are skewed and thus log transformed before the analysis to achieve an approximately normal distribution.

We evaluate classification accuracy of the various classifiers using ROC curve and its summary statistics, including the AUC and S90, which were estimated using LOOCV to account for overfitting due to building and evaluating the model on the same data set (Bose et al., 2003; Lemaître et al., 2015). In our case, LOOCV was completed on the subject/slice level rather than the voxel level, considering the correlation between voxels from the same slice. The number of basis functions in the adaptive polynomial spline regression and the kernel bandwidth of the spatial Gaussian kernel smoother were treated as tuning parameters and selected by LOOCV to maximize the average AUC. The basis functions in the spline regression were determined by stepwise model selection to minimize the penalized residual sum of squares using an R package "polspline" (Kooperberg et al., 1997; Kooperberg and Kooperberg, 2019). 95% Confidence intervals and p-values for model comparison were calculated using a Bootstrapping procedure that resampled prostate slices, rather than voxels. Voxels were not resampled within each slice in order to retain the spatial structure in the data. Model evaluation is based on the classification on 1000 resampled data sets.

Table 2.2 and Figure 2.3 summarize the classification results of the Mbase, Mregion, Mcoord and Msmooth models. Table 2.2 presents the estimated AUC and S90 for the complete cases (i.e. only including the voxels with no missing mpMRI parameters) and for all voxels (i.e. including the voxels with missing mpMRI parameters). Figure 2.3 shows ROC curves for the classification on complete voxels and on all voxels.

We first discuss the classification results on the complete cases, and then discuss the results on all voxels. Overall, the application to the motivating data set supports our findings from the simulation studies. Both the Mregion model and the Mcoord model provided significantly improved AUC ($p = 0.013$ and $p = 0.011$, respectively) and S90 ($p = 0.015$ and $p = 0.012$, respectively), compared with the Mbase model. The average AUC and S90 of Mregion were 0.038 and 0.046 higher than that of Mbase, respectively. We note that Mregion uses an automated procedure to segment the prostate into PZ and CG. The average AUC by LOOCV for this region segmentation step was 0.972, and the Mregion with automated region segmentation performed similarly as Mregion assuming known region information (AUC: 0.767 vs. 0.763; S90: 0.475 vs. 0.476). Mcoord achieved similar classification accuracy, resulting in AUC and S90 that were 0.043 and 0.065 higher than those of Mbase, respectively. Compared with Mregion, Mcoord achieved higher AUC (0.772 vs. 0.767) and S90 (0.494 vs. 0.475), but the differences were not significant (AUC: $p = 0.413$; S90: $p = 0.320$).

Table 2.2: Classification results on the motivating data set. Each point estimate is the average value obtained from LOOCV. Results in the parentheses are the 95% Bootstrap confidence intervals obtained from 1000 Bootstrap samples. Bootstrap p-values for the comparison of AUC and S90 are listed in the third and fifth columns, respectively, with the superscript indicating which two models are compared: 1-4 represent Mbase, Mregion, Mcoord and Msmooth, respectively.

Model	AUC	P-value		S90	P-value	
Complete Cases						
Mbase	$.729(.679, .781)$			$.429(.337, .515)$		
Mregion	$.767(.715, .816)$	$.013^{1,2}$		$.475(.375, .577)$	$.015^{1,2}$	
Mcoord	$.772(.716, .826)$	$.011^{1,3}$	$.413^{2,3}$	$.494(.391, .597)$	$.012^{1,3}$	$.320^{2,3}$
Msmooth	$.822(.756, .880)$	$< .001^{1,4}$	$.001^{3,4}$	$.599(.485, .712)$	$.001^{1,4}$	$< .001^{3,4}$
All Cases						
Mbase	$.721(.669, .774)$			$.419(.332, .506)$		
Mregion	$.760(.709, .808)$	$.036^{1,2}$		$.468(.366, .568)$	$.039^{1,2}$	
Mcoord	$.766(.709, .821)$	$.012^{1,3}$	$.293^{2,3}$	$.486(.387, .584)$	$.013^{1,3}$	$.209^{2,3}$
Msmooth	$.818(.752, .880)$	$.002^{1,4}$	$< .001^{3,4}$	$.596(.482, .709)$	$.003^{1,4}$	$< .001^{3,4}$

The spatial Gaussian kernel smoothing step appeared to further improve cancer classification. Here, Msmooth was obtained by combining Mcoord with the spatial smoother discussed in Section 2.3.5. From Table 2.2, it can be observed that Msmooth gave significantly higher AUC and S90 compared with Mcoord ($p = 0.001$ and $p < 0.001$, respectively). The AUC and S90 for Msmooth were 0.050 and 0.105 higher than those of Mcoord, respectively. This gain in AUC and S90 from the spatial Gaussian kernel smoothing is similar in magnitude to the improvement from modeling the regional heterogeneity (i.e. comparing Mregion and Mcoord to Mbase). Overall, after modeling the regional heterogeneity and conducting spatial smoothing on the voxel-wise posterior predictive cancer probabilities, Msmooth gave AUC and S90 that were 0.093 (12.8%) and 0.170 (39.6%) higher than those of Mbase, respectively ($p < 0.001$ and $p = 0.001$, respectively).

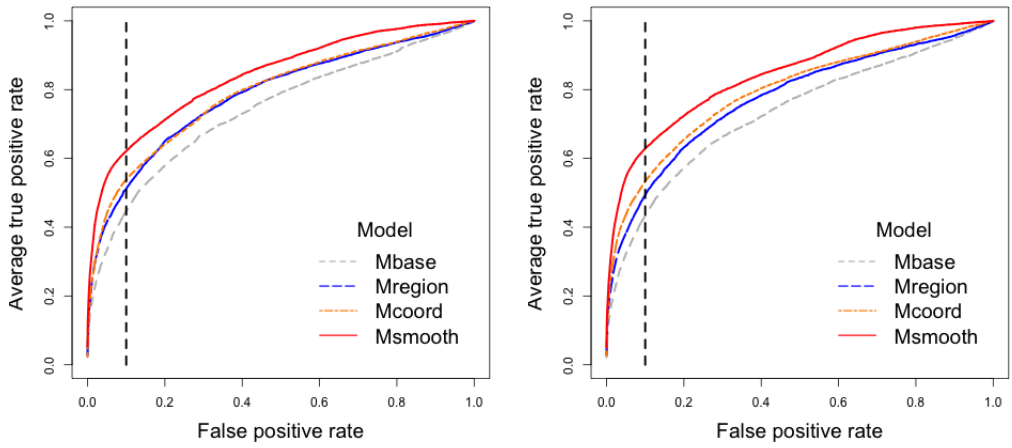

Figure 2.3: Average ROC curves obtained from LOOCV. The sub-figure on the left shows results on the complete cases (i.e. the voxels that have no missing mpMRI parameters). The sub-figure on the right shows results on all voxels. The dashed lines overlaying the ROC curves indicate the S90 values.

We now compare the model performance on all voxels to the performance on the complete cases. In our motivating data set, 8.553% of the voxels have at least one missing mpMRI parameter, most of which have one (3.915%) or three (4.347%) parameters miss-

ing. Even with this substantial amount of missing data, the resulting AUC and S90 of Msmooth were similar to the ones from complete case analysis (AUC: 0.818 v.s. 0.822, $p = 0.326$; S90: 0.596 v.s. 0.599, $p = 0.393$).

Figure 2.4 presents heatmaps of the voxel-wise posterior predictive cancer probabilities for three randomly selected prostate slices in our data. Overall, we observe improved classification moving from the left to the right. The Mbase model generated diffuse areas of high posterior predictive cancer probabilities, which sometimes failed to cover the true tumor areas. The Mregion and Mcoord models, on the contrary, identified larger area of high cancer probability that covered the true tumor areas, and the overestimated cancer probabilities for the remaining areas were reduced. In addition, while the Mregion and Mcoord models generated similar posterior predictive cancer probabilities, the heatmaps of Mcoord have more areas of deep red, indicating high probability of cancer, that overlap with the true cancer areas. Our final model, Msmooth, further improved the classification by reducing the sharp difference between neighboring voxels.

Figure 2.5 presents a binary representation of the cancer classification results, which categorizes the voxels as true positive, false positive, false negative and true negative using the probability cut-off corresponding to 90% specificity. The area of true positives (i.e. the red regions) clearly increases from the left to the right, with a decrease in the area of false negatives (i.e. the blue regions). We also observe a decrease in the area of false positives (i.e. the yellow regions) from the left to the right, and the random false-positive voxels are eliminated by Msmooth. However, the maps do highlight one drawback of spatial smoothing, as it has the potential to over-smooth some areas in a prostate slice and create continuous false positive areas adjacent to the predicted tumor areas. This is likely due to the use of a constant spatial smoothing bandwidth across all prostate slices, or local non-stationarity in the underlying spatial process in some slices. Even so, simulation results in Section 2.4.1 have shown that, in both scenarios, the average classification accuracy can still be improved by the spatial smoothing technique.

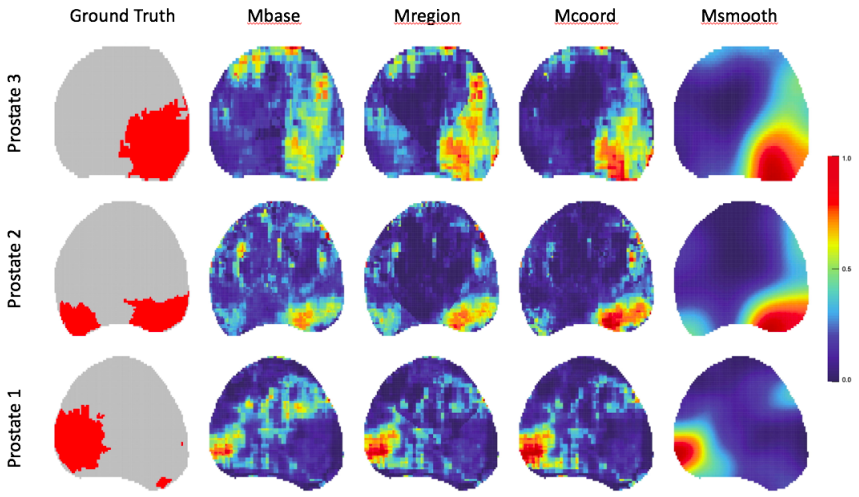

Figure 2.4: Maps of three prostate slices: ground truth (column 1, where red indicates true cancer voxels), heatmaps of the voxel-wise posterior predictive cancer probabilities for Mbase, Mregion, Mcoord and Msmooth, respectively (columns 2-5, where warmer color indicates higher predicted cancer probability, and the color is re-scaled separately for each prostate slice).

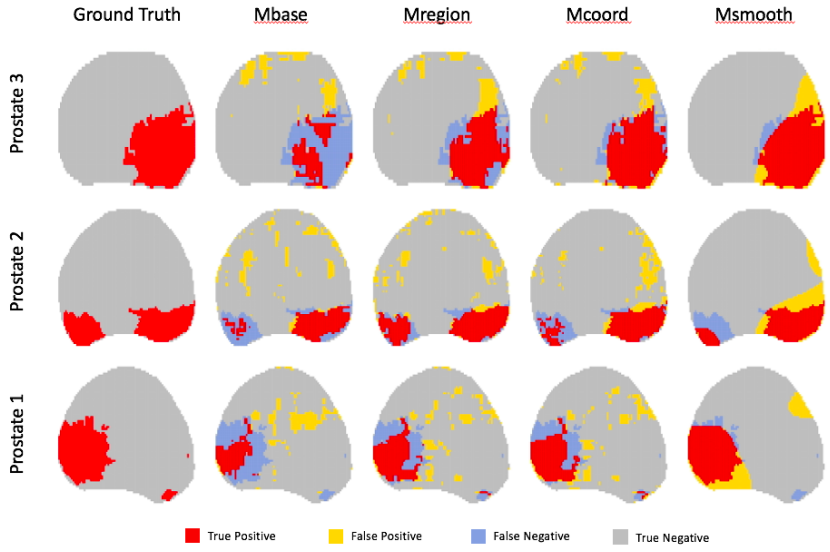

Figure 2.5: Maps of three prostate slices: ground truth (column 1), maps of classification results using the probability cut-off corresponding to 90% specificity, for Mbase, Mregion, Mcoord and Msmooth, respectively (columns 2-5).

For all models, we implement the EB approach to obtain a closed-form solution for the voxel-wise posterior predictive cancer probabilities, which potentially comes at the price of lower classification accuracy compared with a fully Bayesian model that better captures uncertainty in the estimates of the model parameters. To ensure minimal loss in the classification accuracy, we also compared Mregion and Mcoord with their fully Bayesian versions. Specifically, for the fully Bayesian version of Mregion, we assumed a Unif$(0, 1)$ prior on the prior cancer risk $p(c^*|r^*)$ in (2.11), and a flat prior on $\boldsymbol{\beta}_{\text{coord}}$ in (2.12). In addition, instead of estimating the hyperparameter $\boldsymbol{\Omega_{c,r}}$ in (2.6) by an EB approach, we considered the commonly used specification: $\boldsymbol{\Omega_{c,r}} = \boldsymbol{I_d}$, where $\boldsymbol{I_d}$ denotes a $d \times d$ identity matrix. For the fully Bayesian version of Mcoord, we assumed a flat prior on $\boldsymbol{\beta_p}$ in Section 2.3.3, the vector of coefficients in the logistic spline regression for predicting prior cancer risk using standardized 2-D coordinates. Similar to Mregion, we also assumed an inverse Wishart prior for $\boldsymbol{\Sigma_c}$, with $\delta = d + 1$ and $\boldsymbol{\Omega_c}$ equal to $\boldsymbol{I_d}$.

We ran the fully Bayesian models in JAGS using the R package "rjags" (Plummer, 2013). Posterior distributions were approximated by 5000 MCMC iterations, after 2000 iterations for initial burn-in. The AUC and S90 were still estimated by LOOCV. For the Mregion model, the average AUC was 0.766 using both the EB approach and fully Bayesian approach, while the average S90 was 0.475 using the EB approach, and 0.474 using the fully Bayesian approach. Similarly, for the Mcoord model, the average AUC was 0.770 using the EB approach, and 0.772 using the fully Bayesian approach, while the average S90 was 0.494 using the EB approach, and 0.497 using the fully Bayesian approach. These results suggest that, at least for the data we have, there is little cost in classification accuracy associated with our suggested EB approach, as compared to the fully Bayesian approach.

2.5 Discussion

We proposed a Bayesian mpMRI classifier that improves the voxel-wise classification of prostate cancer by accounting for the anatomic structure of the prostate: the existence of heterogeneity between the PZ and CG in both the distribution of mpMRI parameters and cancer risk. Additionally, we account for the residual spatial correlation in the voxel-wise cancer risk using a spatial Gaussian kernel smoothing technique, which maintains the accessibility of a closed-form solution for the posterior predictive cancer probabilities. Our simulation results indicate that the proposed classifiers result in improved classification compared with a baseline model that does not account for the regional heterogeneity. The application to the motivating data set shows that our final model, Msmooth, results in a 0.093 (12.8%) increase in AUC, and a 0.170 (39.6%) increase in S90 compared with the baseline model.

An important advantage of the proposed classifiers is the high computational efficiency. The classifiers were developed in the Bayesian paradigm, which typically requires numerical techniques, such as the MCMC, to approximate the posterior distribution, which, for an mpMRI data set with several thousands voxels per image, will be computationally intensive. We are able to avoid sampling from the posterior by estimating nuisance parameters using an EB approach, which allows us to derive a closed-form solution for the voxel-wise posterior predictive cancer probabilities. This leads to a large reduction in computing time, but with minimal loss of classification accuracy as compared to the fully Bayesian approach. For our data, it took approximately 4.76 seconds on average for Msmooth, our final model, to conduct voxel-wise cancer classification for a prostate slice with 3467 voxels. As a comparison, the fully Bayesian version implemented in JAGS using 5000 MCMC iterations after a burn-in stage of 2000 iterations was about 138 times slower, with no improvement in the classification accuracy.

An additional advantage of the proposed classifiers is that, cancer classification for

the voxels with missing mpMRI parameters is available without the need for imputation. Application to the motivating data set where 8.553% of the voxels have missing mpMRI parameters only resulted in a slight decrease in AUC and S90 compared with complete case analysis. Note that direct application of our proposed classifiers to the voxels with missing mpMRI parameters relies on the assumption of MCAR, which is unlikely to be true in reality, but leads to reasonable classification results on our data. Modeling the missing data mechanism can potentially result in higher classification accuracy by allowing us to leverage information in the missing data pattern, and therefore will be one of our future considerations.

The four mpMRI parameters, ADC, K^{trans}, k_{ep} and AUGC90, were selected as the set of candidate parameters that maximized the average AUC. As a result, another mpMRI parameter, T2, which is also widely used for the diagnosis of prostate cancer, was not included in the models. However, several combinations of the mpMRI parameters that included T2 lead to similar classification accuracy, and should not be ruled out as significantly worse than the selected set of parameters. Further work is needed for proposing an optimal model selection procedure for the current modeling framework. Alternately, Bayesian model averaging could be used in the place of model selection. The model weights can be set to vary by the standardized 2-D coordinates, as the optimal combination of mpMRI parameters might potentially depend on the location in a prostate.

Our proposed classifier utilizes a hierarchical modeling framework that allows us to incorporate further structural complexity for the mpMRI parameters and cancer risk. We have already exploited this flexibility to account for the regional heterogeneity, but this framework could be extended to account for other mpMRI features, as well. For example, strong, subject-specific shifts from the overall mean have been observed for the mpMRI parameters. In principle, this could be accommodated in our modeling framework by including subject-specific random effects into the likelihood, but additional work is needed to understand the practical implications of this approach. In this dissertation, we focused our

model development on 2-D prostate slices, due to technical limitations that currently prevent 3-D co-registration of the MRI and pathology data sets. However, our method could be easily extended to 3-D mpMRI data by replacing the 2-D splines with 3-D splines (Ruppert et al., 2009; Spinello et al., 2010; Lee et al., 2014), and the computation time should scale linearly with the number of voxels in the image.

Currently, our classifier utilizes Spatial Gaussian Kernal Smoothing to account for the spatial correlation between voxels, which was shown in Section 2.4.2 to result in significant improvement in the classification. The spatial dependency between voxels can also be accounted for by assigning spatially correlated priors on $p(c^*|\boldsymbol{s^*})$. However, modeling the spatial correlation structure is not trivial for the mpMRI data set. First, specifying spatial correlation structure in the voxel-wise cancer probabilities leads to intractable posterior distributions, which requires sampling techniques (e.g. MCMC) to approximate the posterior, and thus may result in a big loss of computational efficiency. Second, each MCMC iteration requires inverting multiple spatial covariance matrices (one for each prostate slice), each having a dimension of $n_i \times n_i$, where n_i ranges from 2098 to 5756, thus making the model computationally infeasible. However, we do think that modeling the complex spatial correlation structure via stationary or non-stationary spatial processes has the potential to improve our classifier. We leave this task to Chapter 3.

Chapter 3

Voxel-wise Classification of Prostate Cancer Accounting for the Spatial Correlation and Between-Patient Heterogeneity in mpMRI

3.1 Introduction

Previously, it has been observed that there is substantial difference in the distribution of the observed mpMRI parameters and cancer risk both between and within the two regions of the prostate: the PZ and CG. In Chapter 2, Jin et al. (2018) proposed to use a Bayesian hierarchical modeling framework, which has a flexible structure allowing complex distributional assumptions for both the predictors (mpMRI parameters) and the outcome (cancer status). Based on this framework, fully automated classifiers were developed to model the regional heterogeneity in mpMRI, and were shown to improve cancer classification.

In addition to the regional heterogeneity, the mpMRI data also show substantial spatial correlation in the voxel-wise cancer status, and between-patient heterogeneity in the observed mpMRI parameters, where the latter is possibly due to registration error or variability in patients' physical conditions. These features have not been previously investigated in the context of fully automated, voxel-wise prostate cancer classification, but properly

modeling these sources of variability could further improve the performance of the current classifiers.

The major challenge associated with modeling the spatial correlation between voxels is the extreme high dimensionality of the data. Typically, an mpMRI image has a resolution of $50 \times 50 \sim 80 \times 80$, meaning thousands of voxels per image, which must be modeled simultaneously over multiple images for model development. Spatial modeling on each image requires inverting large spatial covariance matrices that typically requires $\sim n^3$ floating point operations and storage of order n^2, with n denoting the number of voxels in the image, which is computationally challenging and, in fact, infeasible for our motivating data set.

There are two general approaches that can be considered for spatial modeling of high-dimensional mpMRI data. One is sparse approximation, which introduces sparsity in the spatial covariance or precision matrix. Methods include covariance tapering assuming compactly supported covariance functions (e.g. Furrer et al., 2006; Kaufman et al., 2008), and approximating the likelihood by the product of lower dimensional conditional densities (e.g. Vecchia, 1988; Stein, 2014), Markov random fields (e.g. Rue and Held, 2005) or composite likelihoods (e.g. Eidsvik et al., 2014). In addition, Datta et al. (2016) proposed a Nearest Neighbor Gaussian Process (NNGP) for fully process-based modeling of large spatial data sets, which approximates the likelihood of a spatial process by the product of conditional densities between nearest neighbors. The other general approach is reduced-rank approximation, which constructs spatial processes on a lower-dimensional subspace. Methods include predictive process models (Banerjee et al., 2008; Finley et al., 2009), low rank splines or basis functions (Cressie and Johannesson, 2008) and kernel convolutions (Higdon, 2002).

In this chapter, we propose Bayesian hierarchical models that systematically model the between-patient heterogeneity in the mpMRI parameters and spatial correlation in the cancer status to improve voxel-wise classification of prostate cancer. Our models were de-

veloped based on the approach of Jin et al. (2018) in Chapter 2, which models the regional heterogeneity in mpMRI for voxel-wise cancer classification, and was developed under a flexible Bayesian hierarchical model framework that allows us to introduce additional model structures to account for the between-patient heterogeneity and spatial correlation in mpMRI. This baseline model will also serve as a reference to assess the performance of our proposed models. The between-patient heterogeneity in the observed mpMRI parameters is incorporated by specifying a subject specific random intercept for the mpMRI parameters, and we propose three computationally efficient approaches to modeling the spatial correlation in the voxel-wise cancer statuses. First, we consider a sparse approximation using the NNGP proposed by Datta et al. (2016). Second, we propose an approach through knot-based, reduced-rank approximation (Banerjee et al., 2008). Finally, we propose an approach using Gaussian Markov random fields with a commonly used and easily implemented conditional autoregressive (CAR) model. Our simulation results illustrate that the proposed models substantially improve classification accuracy by incorporating both between-patient heterogeneity and spatial correlation. Application to our motivating data demonstrates improvement due to modeling the spatial correlation using the NNGP and reduced-rank approximation but not the CAR model, while modeling the between-patient heterogeneity did not further improve classification accuracy. Among our proposed models, the model using NNGP for the modeling of spatial correlation had the best overall performance, with robust classification accuracy and high computational efficiency.

The remaining chapter proceeds as follows. In Section 3.2, we introduce methods development and the MCMC algorithm for model implementation. In Section 3.3, we present simulation studies and application on the motivating data set to illustrate the model performance. We conclude with a discussion in Section 3.4.

3.2 Method

3.2.1 Baseline Model

The classification models proposed in this chapter were built upon the work of Jin et al. (2018) presented in Chapter 2. Previously, Jin et al. (2018) observed heterogeneity in both the distribution of mpMRI parameters and cancer prevalence between the two regions of the prostate, PZ and CG. To incorporate regional heterogeneity, Jin et al. (2018) proposed a "Mregion" model, which models region-specific joint density of the mpMRI parameters, $\boldsymbol{Y}$, and cancer status, $\boldsymbol{C}$, given region information, $\boldsymbol{R}$, via a Bayesian hierarchical model framework. The unknown cancer status for the voxels in a new prostate image, $\boldsymbol{C}^*$, are classified using Bayesian posterior predictive probabilities given $\boldsymbol{Y}^*$ and $\boldsymbol{R}^*$. Specifically, it is assumed that the joint distribution of $\boldsymbol{Y}$ and $\boldsymbol{C}$ depends on $\boldsymbol{R}$, which can be defined hierarchically as $f(\boldsymbol{Y}, \boldsymbol{C}|\boldsymbol{R}) = f(\boldsymbol{Y}|\boldsymbol{C}, \boldsymbol{R})p(\boldsymbol{C}|\boldsymbol{R})$. For $f(\boldsymbol{Y}|\boldsymbol{C}, \boldsymbol{R})$: $\boldsymbol{y_{ij}}$'s are assumed to follow independent multivariate normal distributions with mean and covariance matrix varying by cancer status, $c_{ij} \in \{0, 1\}$, and region indicator, $r_{ij} \in \{0, 1\}$. For $p(\boldsymbol{C}|\boldsymbol{R})$: c_{ij} is assumed to follows a Bernoulli distribution, $Ber(p_{r_{ij}})$, with cancer probability, $p_{r_{ij}}$, varying by region indicator, $r_{ij} \in \{0, 1\}$:

$$\begin{aligned} \boldsymbol{y_{ij}}|c_{ij}, r_{ij} &\overset{ind}{\sim} \mathcal{MVN}(\boldsymbol{\mu_{c_{ij}, r_{ij}}}, \boldsymbol{\Gamma_{c_{ij}, r_{ij}}}), \\ c_{ij}|r_{ij} &\overset{ind}{\sim} Ber(p_{r_{ij}}). \end{aligned} \tag{3.1}$$

Model (3.1) is the basis for the construction of the models proposed in this chapter, and will be used as the baseline for model evaluation.

3.2.2 General Model Accounting for Between-patient Heterogeneity and Spatial Correlation

First, we introduce our general approach for modeling the between-patient heterogeneity in the mpMRI parameters and spatial correlation in the voxel-wise cancer status.

Between-patient Heterogeneity

In the Bayesian hierarchical model framework of the baseline model (3.1), the between-patient heterogeneity in the mpMRI parameters can be incorporated by a random intercept, $\boldsymbol{\delta_i} \sim \mathcal{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$, on $\{\boldsymbol{y_{ij}}|j = 1, 2, \ldots, n_i\}$, so that $\boldsymbol{y_{ij}}|c_{ij}, r_{ij}, \boldsymbol{\delta_i} \overset{ind}{\sim} \mathcal{MVN}(\boldsymbol{\mu_{c_{ij},r_{ij}}} + \boldsymbol{\delta_i}, \boldsymbol{\Gamma_{c_{ij},r_{ij}}})$. In our setting, $\boldsymbol{\delta_i}$ represents the subject-specific random shift of the i^{th} patient from the overall mean, with respect to the mpMRI parameters.

General Gaussian Process Model for Spatial Correlation

The baseline model (3.1) assumes that all voxels are independent, with voxel-wise cancer probability $p_{r_{ij}}$ only depending on the region indicator of the voxel $r_{ij} \in \{0, 1\}$. In reality, however, substantial spatial correlation has been observed in the voxel-wise cancer status within each image. To account for this data feature, we instead assume that the voxel-wise cancer probabilities in the i^{th} image, which we now denote as $\{p_{ij}|j = 1, \ldots, n_i\}$, are dependent on the location/coordinates of the voxels, and are spatially correlated according to the distance between 2-D coordinates $\{\boldsymbol{s_{ij}}|j = 1, \ldots, n_i\}$. Notice that this only implies spatial correlation between voxels within an image and that voxels from different images (patients) are assumed independent. For Gaussian distributed geostatistical data, Gaussian process models are widely used to model the spatial correlation between locations assuming correlated Gaussian spatial random effects. The p_{ij}'s, however, are restricted to the unit interval, and Gaussian spatial random effects cannot be added directly. Instead, we apply a probit transformation and specify $q_{ij} = \Phi^{-1}(p_{ij})$, which has a support of $(-\infty, +\infty)$,

where Φ^{-1} denotes the probit function, which is also the inverse CDF of a standard normal distribution, and introduce spatial correlation by adding spatially correlated Gaussian random effects $\boldsymbol{w_i} = (w_{i1}, w_{i2}, \ldots, w_{i,n_i})^T$ on $\boldsymbol{q_i} = (q_{i1}, q_{i2}, \ldots, q_{i,n_i})^T$.

Under the probit model structure, we introduce a latent variable, $\kappa_{ij} \sim N(q_{r_{ij},0}+w_{ij}, 1)$, where $q_{r_{ij},0}$ denotes the probit of the region-specific cancer prevalence. The cancer status, c_{ij}, then follows a Bernoulli distribution, where $c_{ij} = 1$ if $\kappa_{ij} > 0$, and 0 otherwise. The full model that accounts for both between-patient heterogeneity in the mpMRI parameters and spatial correlation in voxel-wise cancer risk becomes:

$$\begin{aligned} \boldsymbol{y_{ij}} &\overset{ind}{\sim} \mathcal{MVN}(\boldsymbol{\mu_{c_{ij},r_{ij}}}+\boldsymbol{\delta_i}, \boldsymbol{\Gamma_{c_{ij},r_{ij}}}), \;\; \boldsymbol{\delta_i} \sim \mathcal{MVN}(\mathbf{0}, \boldsymbol{\Sigma}), \\ c_{ij} &= I(\kappa_{ij} > 0), \;\; \kappa_{ij} \sim N(q_{r_{ij},0} + w_{ij}, 1), \;\; \boldsymbol{w_i} \sim \mathcal{MVN}(\mathbf{0}, \boldsymbol{C}(\boldsymbol{S_i}, \boldsymbol{S_i}|\boldsymbol{\theta})), \end{aligned} \tag{3.2}$$

where $\boldsymbol{S_i} = (\boldsymbol{s_{i1}}, \boldsymbol{s_{i2}}, \ldots, \boldsymbol{s_{i,n_i}})^T$, $\boldsymbol{C}(\boldsymbol{S_i}, \boldsymbol{S_i}|\boldsymbol{\theta})$ denotes the spatial covariance matrix of $\boldsymbol{w_i}$ and $\boldsymbol{\theta}$ denotes the set of spatial parameters shared by all images. For the construction of $\boldsymbol{C}(\boldsymbol{S_i}, \boldsymbol{S_i}|\boldsymbol{\theta})$ in image i, we assume that $\boldsymbol{w_i}$ is the realization of a zero-mean Gaussian process GP$(\mathbf{0}, \boldsymbol{C}(\cdot, \cdot|\boldsymbol{\theta}))$ on locations $\boldsymbol{S_i}$. We define $\boldsymbol{C}(\boldsymbol{S_i}, \boldsymbol{S_i}|\boldsymbol{\theta}) = \sigma^2 \boldsymbol{\rho}(\boldsymbol{S_i}, \boldsymbol{S_i}|\boldsymbol{\zeta})$, where $\rho(\cdot, \cdot|\boldsymbol{\zeta})$ is a correlation function with the set of correlation parameters $\boldsymbol{\zeta}$, σ^2 denotes the spatial variance and $\boldsymbol{\theta} = \{\sigma^2, \boldsymbol{\zeta}\}$. In this chapter, we employ a Matérn stationary correlation function (Stein, 2012), which is one of the most widely used correlation functions in spatial statistics which flexibly covers various types of stationary spatial correlation patterns. Given the Matérn correlation function, the spatial correlation between two locations $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^2$ is defined as:

$$\rho(\boldsymbol{s}, \boldsymbol{t}|\boldsymbol{\zeta}) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \Big(\frac{2\nu^{1/2}\|\boldsymbol{s} - \boldsymbol{t}\|}{\phi} \Big)^{\nu} \boldsymbol{J}_{\nu} \Big(\frac{2\nu^{1/2}\|\boldsymbol{s} - \boldsymbol{t}\|}{\phi} \Big), \tag{3.3}$$

where $\boldsymbol{\zeta} = \{\phi, \nu\}$, ϕ is the spatial range parameter where larger ϕ indicates larger-scale spatial correlation, ν is the smoothness parameter controlling the degree of differentiabil-

ity, where larger ν indicates smoother correlation, $\Gamma(\cdot)$ is the gamma function, $\boldsymbol{J}_\nu(\cdot)$ is the modified Bessel function of the second kind with order ν, and $\|\cdot\|$ denotes the Euclidean distance. The matérn correlation function covers several commonly seen classes of correlation functions, including the exponential ($\nu = 0.5$) and squared exponential ($\nu \to \infty$) correlation functions.

3.2.3 Computationally Efficient Modeling Approaches for Spatial Correlation

The full spatial process model in (3.2) becomes computationally infeasible in the case of our motivating data, where there are a large number of images, each involving a separate spatial process over thousands of voxels. In this sub-section, we propose three different approaches that ensure scalable spatial modeling of the mpMRI data.

Nearest Neighbor Gaussian Process (NNGP)

Our first approach to modeling spatial correlation in a computationally efficient manner is through sparse approximation via NNGP. Most sparse approximation approaches do not necessarily define a valid spatial process, and prediction is through interpolation from a different spatial process that may not reflect the true predictive uncertainty. To deal with issue, Datta et al. (2016) proposed an NNGP approach for fully process-based modeling of large spatial data sets, which was shown to significantly outperform competing approaches in terms of inference and scalability, and thus we believe it holds promise in our setting.

The construction of NNGP was discussed in detail by Datta et al. (2016), with applications to spatially-correlated, normally distributed observations on a single map under the linear regression model framework. In our setting, we extend NNGP to Bayesian hierarchical modeling of multi-image data with spatially-correlated binary outcomes. We apply a separate NNGP to each image. Take image i as an example: the joint density of $\boldsymbol{w_i}$, the

voxel-wise spatial random effects on the probit-transformed cancer probabilities in model (3.2), can be written as the product of conditional densities:

$$f(\boldsymbol{w_i}) = \prod_{j=1}^{n_i} f(w_{ij}|w_{i1}, w_{i2}, \ldots, w_{i,j-1}). \tag{3.4}$$

To reduce the computational burden, we replace the large conditioning set $\{w_{i1}, w_{i2}, \ldots, w_{i,j-1}\}$ with a smaller set of size at most m on $N(\boldsymbol{s_{ij}}) \subseteq \boldsymbol{S_i} \setminus \{\boldsymbol{s_{ij}}\}$, where $m \ll \min_i n_i$, to construct an alternative density:

$$\tilde{f}(\boldsymbol{w_i}) = \prod_{j=1}^{n_i} f(w_{ij}|\boldsymbol{w_{N(s_{ij})}}), \tag{3.5}$$

where $\boldsymbol{w_{N(s_{ij})}}$ denotes the spatial random effects on $N(\boldsymbol{s_{ij}})$, the m nearest neighbors of $\boldsymbol{s_{ij}}$. For each image i, we view $\{\boldsymbol{S_i}, N_{\boldsymbol{S_i}}\}$ as a directed graph $\mathcal{G}$, with $\boldsymbol{S_i}$ being the set of nodes and $\boldsymbol{N_{S_i}}$ the set of directed edges. It has been proven that if $\mathcal{G}$ is a directed acyclic graph, then $\tilde{f}(\boldsymbol{w_i})$ in (3.5) will be a proper multivariate joint density. Specifically, let $C_{\boldsymbol{s_{ij}}}$ denote the variance of w_{ij}, $\boldsymbol{C_{N(s_{ij})}}$ the $m \times m$ covariance matrix of $\boldsymbol{w_{N(s_{ij})}}$, and $\boldsymbol{C_{s_{ij},N(s_{ij})}}$ the $1 \times m$ cross-covariance matrix between w_{ij} and $\boldsymbol{w_{N(s_{ij})}}$. By the property of the multivariate Gaussian distribution:

$$\tilde{f}(\boldsymbol{w_i}) = \prod_{j=1}^{n_i} N(w_{ij}|\boldsymbol{B_{s_{ij}}}\boldsymbol{w_{N(s_{ij})}}, F_{\boldsymbol{s_{ij}}}), \tag{3.6}$$

where $\boldsymbol{B_{s_{ij}}} = \boldsymbol{C_{s_{ij},N(s_{ij})}}\boldsymbol{C^{-1}_{N(s_{ij})}}$, $F_{\boldsymbol{s_{ij}}} = C_{\boldsymbol{s_{ij}}} - \boldsymbol{C_{s_{ij},N(s_{ij})}}\boldsymbol{C^{-1}_{N(s_{ij})}}\boldsymbol{C_{N(s_{ij}),s_{ij}}}$. In fact, $\tilde{f}(\boldsymbol{w_i})$ is the probability density function (pdf) of a multivariate normal distribution, which we denote as $\mathcal{MVN}(\boldsymbol{0}, \tilde{\boldsymbol{C}}_{\boldsymbol{S_i}})$. Given that each $N(\boldsymbol{s_{ij}})$ has at most m ($m \ll \min_i n_i$) members, it can be shown that the precision matrix $\tilde{\boldsymbol{C}}^{-1}_{\boldsymbol{S_i}}$ is sparse with at most $m(m+1)n_i/2$ non-zero entries.

To construct the neighbor sets $\{N(\boldsymbol{s_{ij}})|j = 1, \ldots, n_i\}$ in the i^{th} image, we first order

the voxels by x-coordinate then y-coordinate, and denote the ordered voxels as $\{\boldsymbol{s_{i1}}, \boldsymbol{s_{i2}}, ..., \boldsymbol{s_{i,n_i}}\}$, then define $N(\boldsymbol{s_{ij}})$ as the set of m voxels in $\{\boldsymbol{s_{i1}}, \boldsymbol{s_{i2}}, ..., \boldsymbol{s_{i,j-1}}\}$ with the shortest Euclidean distance from $\boldsymbol{s_{ij}}$. The ordering of voxels has been shown to have no discernible impact on the approximation of (3.4) by (3.5). The choice of m can follow standard model comparison metrics such as DIC, but typically a small value between $10 \sim 15$ can provide inference almost indistinguishable to full spatial models for an image with thousands of voxels. Our proposed model with NNGP for spatial modeling follows the structure of the full model (3.2), except that we replace the original prior for $\boldsymbol{w_i}$, $\mathcal{MVN}(\boldsymbol{0}, \boldsymbol{C_{S_i}})$, by the NNGP prior $\mathcal{MVN}(\boldsymbol{0}, \tilde{\boldsymbol{C}}_{\boldsymbol{S_i}})$, $i = 1, 2, \ldots, N+1$.

Knot-based Reduced-rank Approximation

The NNGP-based approach proposed in the previous paragraphs reduces computational intensity by inducing sparsity in the large spatial precision matrix, which has been proven to perform well in capturing local spatial dependence structures. An alternative approach is through reduced-rank approximation, which is better equipped to capture large-scale, global spatial dependency (Stein, 2008; Finley et al., 2009).

Among the various reduced-rank approximation techniques, we considered a knot-based method proposed by Banerjee et al. (2008). The basic idea is to construct a new spatial process by regressing the original process on its realizations over a smaller set of locations, which are referred to as "knots". Take the i^{th} image as an example: we first select a set of a knots $\boldsymbol{S_i^*} = \{\boldsymbol{s_{i,1}^*}, ..., \boldsymbol{s_{i,a}^*}\} \subset \boldsymbol{S_i}$, where $a \ll \min_i n_i$, with corresponding spatial random effects $\boldsymbol{w_i^*} = (w_{\boldsymbol{s_{i,1}^*}}, ..., w_{\boldsymbol{s_{i,a}^*}})^T$. The original Gaussian process in model (3.2) yields $\boldsymbol{w_i^*} \sim \mathcal{MVN}(\boldsymbol{0}, \boldsymbol{C_{S_i^*}})$, where $\boldsymbol{C_{S_i^*}} = C(\boldsymbol{S_i^*}, \boldsymbol{S_i^*}|\boldsymbol{\theta})$. If we use $\boldsymbol{w_i^*}$ as the basis, then for any single site $\boldsymbol{s_{ik}}$, the corresponding spatial interpolant is given by $\tilde{w}(\boldsymbol{s_{ik}}) = E[w(\boldsymbol{s_{ik}})|\boldsymbol{w_i^*}] = \boldsymbol{C_{s_{ik},S_i^*}}\boldsymbol{C_{S_i^*}^{-1}}\boldsymbol{w_i^*}$, where $\boldsymbol{C_{s_{ik},S_i^*}} = \boldsymbol{C}(\boldsymbol{s_{ik}}, \boldsymbol{S_i^*}|\boldsymbol{\theta})$ is the $1 \times a$ cross-covariance matrix of $\boldsymbol{w_{s_{ik}}}$ and $\boldsymbol{w_i^*}$. This single-site interpolator defines a new spatial process: $\tilde{\tilde{\boldsymbol{w}}}(\boldsymbol{S_i}) \sim \mathbf{GP}(\boldsymbol{0}, \tilde{\tilde{\boldsymbol{C}}}(\cdot, \cdot|\boldsymbol{\theta}, \boldsymbol{S_i^*}))$, with $\tilde{\tilde{\boldsymbol{C}}}(\boldsymbol{S_i}, \boldsymbol{S_i}|\boldsymbol{\theta}, \boldsymbol{S_i^*}) = \boldsymbol{C_{S_i,S_i^*}}\boldsymbol{C_{S_i^*}^{-1}}\boldsymbol{C_{S_i^*,S_i}}$. Un-

der this newly defined spatial process, we can replace the original spatial random effects $\boldsymbol{w_i}$ by $\widetilde{\widetilde{\boldsymbol{w}}}_{\boldsymbol{i}} = E[\boldsymbol{w_i}|\boldsymbol{w_i^*}] = \boldsymbol{C_{S_i,S_i^*}}\boldsymbol{C_{S_i^*}^{-1}}\boldsymbol{w_i^*}$. Since the resulting covariance matrix has a fixed rank a much smaller than $\min_i n_i$, faster computation can be achieved by avoiding inverting spatial covariance matrices of size larger than $a \times a$.

CAR Model

The final approach we consider uses Gaussian Markov random fields (GMRF). Different from the NNGP-based and reduced-rank approaches, GMRF does not specify a spatial correlation function. Instead, spatial dependency is introduced by specifying the conditional distributions of the spatial random effects $\{f(w_{ij}|\boldsymbol{w_{i,-j}})|j = 1, \ldots, n_i\}$. Here we apply a popular application of GMRF, the conditional autoregressive (CAR) model, on $\boldsymbol{w_i}$:

$$w_{ij}|\boldsymbol{w_{i,-j}} \sim N(\sum_{k \neq j} b_{ijk} w_{ik}, \sigma^2), \tag{3.7}$$

where b_{ijk} is the weight of w_{ik} on w_{ij} that is specified by the user, and we define $b_{ijk} = \frac{d_{ijk}^{-1}}{\sum_{l \neq j} d_{ijl}^{-1}}$, with d_{ijk} denoting the Euclidean distance between voxel j and k in image i. Using Brook's lemma (Rue and Held, 2005), we can obtain that $\boldsymbol{w_i} \sim \mathcal{MVN}(\boldsymbol{0}, \sigma^2(\boldsymbol{I} - \boldsymbol{B_i})^{-1})$, where $\boldsymbol{B_i} = [b_{ijk}]_{j,k=1}^{n_i}$. Since the precision matrix of $\boldsymbol{w_i}$ has a closed form $\sigma^{-2}(\boldsymbol{I} - \boldsymbol{B_i})$, we have avoided the need to invert large covariance matrices. However, since GMRF does not allow inference on the spatial correlation structure, this model may not reveal the true spatial dependence structure, which may limit accuracy of the resulting classifier.

3.2.4 MCMC Algorithm for Bayesian Inference and Classification

Bayesian inference and classification of the various models were implemented using MCMC algorithms via Gibbs sampler with a Metropolis-Hastings sampling step. The model parameters are estimated using data from all images, and the voxel-wise cancer status indicators for a new patient, $\{c_{N+1,j}|j = 1, \ldots, n_{N+1}\}$, are classified by treating them as unknown

parameters and updating them along with the other model parameters. Derivation of the MCMC algorithm is based on the joint density:

$$\prod_{i=1}^{N+1}\prod_{j=1}^{n_i}\left\{\mathcal{MVN}(\boldsymbol{y_{ij}}|\boldsymbol{\mu_{c_{ij},r_{ij}}}+\boldsymbol{\delta_i},\boldsymbol{\Gamma_{c_{ij},r_{ij}}})P(\kappa_{ij}>0)^{c_{ij}}P(\kappa_{ij}\leqslant 0)^{1-c_{ij}}\right\}\times$$
$$\prod_{i=1}^{N+1}\mathcal{MVN}(\boldsymbol{\delta_i}|\mathbf{0},\boldsymbol{\Sigma})\times\prod_{i=1}^{N+1}\prod_{j=1}^{n_i}N(\kappa_{ij}|q_{r_{ij},0}+w_{ij},1)\times\prod_{i=1}^{N+1}\mathcal{MVN}(\boldsymbol{w_i}|\mathbf{0},\boldsymbol{C_i})$$
$$\times\prod_{c=0}^{1}\prod_{r=0}^{1}\mathcal{W}^{-1}(\boldsymbol{\Gamma_{c,r}}|d-1,\boldsymbol{I_d})\times\mathcal{W}^{-1}(\boldsymbol{\Sigma}|d-1,\boldsymbol{I_d})\times\pi(\boldsymbol{\theta}), \tag{3.8}$$

where $\pi(\boldsymbol{\theta})$ denotes the prior distribution of $\boldsymbol{\theta}$, and $\boldsymbol{C_i}$ is a general notation for the spatial covariance matrix of $\boldsymbol{w_i}$ in the different models, which is replaced by $\boldsymbol{C_{S_i}}$ in the full model, $\tilde{\boldsymbol{C}}_{\boldsymbol{S_i}}$ in the NNGP model, $\tilde{\tilde{\boldsymbol{C}}}_{\boldsymbol{S_i}}$ in the reduced-rank model, and $(\boldsymbol{I}-\boldsymbol{B_i})^{-1}\boldsymbol{D_i}$ in the CAR model, with $\boldsymbol{B_i}$ and $\boldsymbol{D_i}$ defined in the previous paragraphs. Our initial simulation results showed that when the spatial variance, σ^2, was large (e.g. 50), and spatial range, ϕ, was large (e.g. 5), given a flat prior, $\{q_{r,0}|r=0,1\}$, the probit of the cancer prevalence in region r, converged to some value close to 0 or 1, which were far from their true values. This was possibly due to the large-scale spatial correlation and large spatial variance, which resulted in a small effective sample size for estimating the overall cancer prevalence. Therefore, instead of updating $\{q_{r,0}|r=0,1\}$ along with the other parameters, we set them to be the sample prevalence $q_{r,0}=\Phi^{-1}(\sum_{i=1}^{N}\sum_{j:r_{ij}=r}c_{ij}/\sum_{i=1}^{N}\sum_{j:r_{ij}=r}1)$. Our simulation results indicate that this substitution does not degrade our ability to appropriately model the other model parameters.

The MCMC algorithm can be split into two parts. One is for updating the non-spatial model parameters via Gibbs sampling, which is shared by all models. The other is for updating the spatial model parameters, specifically, with Gibbs sampling for the update of w_{ij}'s and k_{ij}'s, and Metropolis-Hastings sampling for the block update of $\boldsymbol{\theta}$. This spatial part varies between models. In addition, for the reduced-rank model, updating w_{ij}'s se-

quentially is computationally challenging, and thus we marginalize over w_{ij}'s and update $\boldsymbol{w}_i^*$'s, the spatial random effects on the selected knots within each image, instead.

While implementing the spatial part of the MCMC algorithm, a crucial strategy that facilitates computational efficiency is to conduct partial parallelization in the posterior computation. The update of the voxel-level spatial components, including k_{ij}'s, w_{ij}'s, and the corresponding spatial matrices involved in their full conditional distributions, are of extreme high dimension and contribute the most to the computational intensity of the algorithm. In fact, these model components are independent between subjects conditional on the other model parameters, and thus can be updated independently within each subject. To make use of this property and further reduce the computational burden, we utilize the partial parallelization strategy in posterior sampling, where in each MCMC iteration, we parallelize the update of these voxel-wise spatial parameters and matrices with respect to subjects due to their conditional independence. Full details of the MCMC algorithms are available in Appendix B.1.

3.3 Application

3.3.1 Simulation Studies

Simulation Settings

We conducted simulation studies to evaluate the classification accuracy of the models discussed in Section 3.2. To simulate a prostate image, a mask (shape of the prostate image, including the voxel-wise region indicators and 2-D coordinates) was sampled with replacement from the prostate images in the real data set. Voxel-wise mpMRI parameters and cancer statuses were then simulated according to the full model (3.2), with a spatial correlation structure following the Matérn correlation function (3.3). We then applied the proposed models, and compared their performance with the baseline model (3.1) to evalu-

ate the improvement in classification accuracy obtained from modeling the between-patient heterogeneity and spatial correlation. We also compared our proposed models with computationally efficient spatial modeling approaches with the full model (3.2), to evaluate their performance in approximating the classification accuracy of full spatial modeling. Since the full model is computationally infeasible on the original-size images, we generated reduced-size images by taking every third row and column of the full-size images, resulting in an image resolution of $18 \times 18 \sim 28 \times 28$.

We conducted the following data generation processes: the general model parameters, including the mean, within-patient and between-patient covariance of the mpMRI parameters, and cancer prevalence in the PZ and CG, were set to equal the estimates from the real data. We set σ^2 (spatial variance) to be 1, 5 or 20, ϕ (range parameter, larger ϕ indicates larger scale correlation) to be 0.1, 0.25, 0.5 or 2, and ν (smoothness parameter, smaller ν indicates larger differentiability) to be 0.5 or 1.5. Figure 3.1 shows the spatial correlation versus distance between voxels under different settings of ϕ and ν. In each simulation, we trained the classifiers using 34 training images, and evaluated the performance using 10 test images.

Simulation Results

Table 3.1 reports the performance of all candidate models under different data scenarios. We considered six candidate models: M-base: the baseline model; M-sse: M-base plus subject specific effects (SSE) accounting for between-patient heterogeneity in the mpMRI parameters; M-sse-nngp, M-sse-rr, M-sse-car and M-sse-full: M-sse plus spatial modeling using NNGP, reduced-rank approximation, CAR and full model, respectively. We set both m (the number of nearest neighbors in NNGP) and a (the number of knots in the reduced-rank model) equal to 10 to make the NNGP and reduced-rank models comparable. The nearest neighbor sets for the NNGP were constructed following the procedure introduced in Section 3.2.3, and we selected a set of equally spaced grids in each image to be the

knots for the reduced-rank model. All samplers were programmed using R with the Rcpp package (Eddelbuettel et al., 2011). The R code for implementing our proposed models is available at https://github.com/Jin93/PCa-mpMRI2.

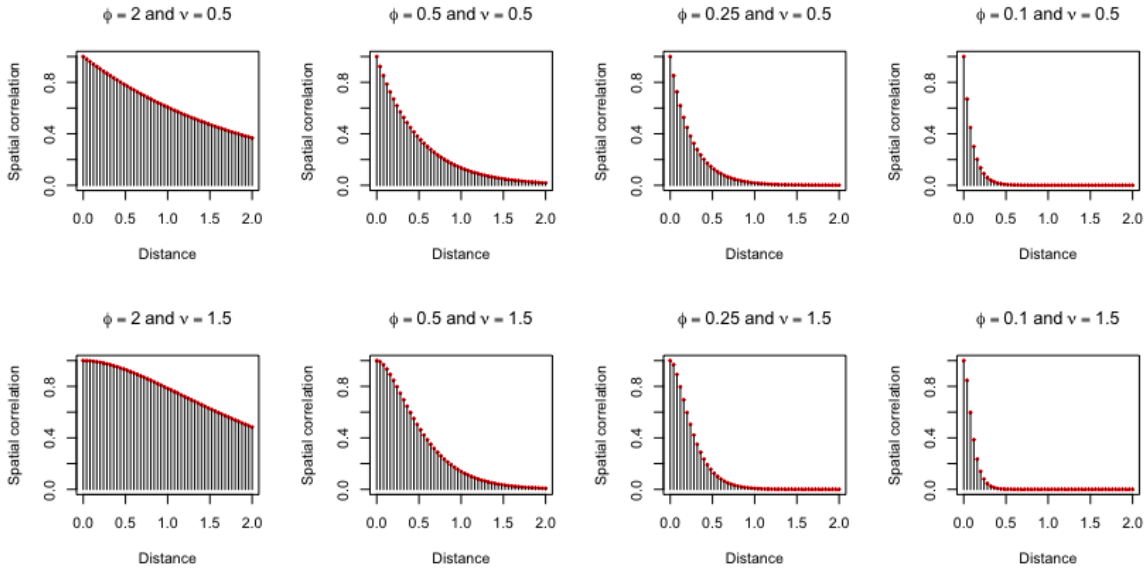

Figure 3.1: Matérn correlation versus spatial distance assuming different values for ϕ (range parameter, larger ϕ indicates larger-scale correlation) and ν (smoothness parameter, smaller ν indicates larger differentiability). The difference in x-axes between two neighboring vertical lines is equal to 0.04, which is the average distance between two neighboring voxels in the motivating data set. The range of the x-axes is $(0, 2)$, which is the scale of the pairwise distances between voxels in the motivating data set.

Table 3.1: Performance of the candidate models under different settings for σ^2 (spatial variance), ϕ (spatial range parameter, larger ϕ indicates larger-scale correlation), and ν (spatial smoothness parameter, smaller ν indicates larger differentiability). "M-base": the baseline model (3.1) in Section 3.2.1; "M-sse": the baseline model plus subject-specific effects (SSE) accounting for between-patient heterogeneity in the mpMRI parameters; "M-sse-nngp", "M-sse-rr", "M-sse-car" and "M-sse-full": models that account for patient heterogeneity, and also spatial correlation using NNGP, reduced-rank model, CAR model and full spatial model, respectively. Bayesian inference and classification were based on two chains of 25000 MCMC iterations after a burn-in stage of 5000 iterations. AUCs are summarized as means with standard deviations in the parentheses, which were obtained from 100 simulations for each data scenario. The "Time" row lists the average computing time in hours for each simulation.

Parameters			AUC					
σ^2	ϕ	ν	M-base	M-sse	M-sse-nngp	M-sse-rr	M-sse-car	M-sse-full
1	2.0	0.5	.755 (.023)	.820 (.015)	.847 (.011)	.851 (.012)	.835 (.010)	.852 (.011)
		1.5	.750 (.032)	.823 (.020)	.833 (.012)	.851 (.014)	.837 (.013)	.851 (.013)
	0.5	0.5	.766 (.022)	.819 (.017)	.859 (.011)	.852 (.011)	.834 (.010)	.861 (.011)
		1.5	.763 (.021)	.819 (.019)	.849 (.011)	.872 (.011)	.838 (.011)	.876 (.010)
	0.25	0.5	.770 (.015)	.815 (.018)	.849 (.010)	.800 (.019)	.830 (.010)	.850 (.011)
		1.5	.767 (.019)	.820 (.011)	.858 (.010)	.822 (.016)	.833 (.010)	.865 (.010)
	0.1	0.5	.769 (.016)	.817 (.013)	.834 (.007)	.627 (.014)	.827 (.008)	.818 (.008)
		1.5	.768 (.015)	.815 (.017)	.837 (.008)	.600 (.015)	.828 (.008)	.831 (.009)
5	2.0	0.5	.760 (.013)	.819 (.014)	.888 (.011)	.888 (.013)	.842 (.009)	.892 (.012)
		1.5	.751 (.027)	.809 (.027)	.821 (.023)	.883 (024)	.832 (.020)	.882 (.024)
	0.5	0.5	.770 (.017)	.807 (.018)	.892 (.010)	.885 (.012)	.835 (.009)	.896 (.010)
		1.5	.767 (.025)	.808 (.020)	.906 (.012)	.919 (.011)	.841 (.010)	.925 (.011)
	0.25	0.5	.776 (.015)	.812 (.011)	.876 (.005)	.848 (.010)	.835 (.005)	.880 (.005)
		1.5	.777 (.013)	.808 (.021)	.901 (.008)	.873 (.009)	.835 (.008)	.907 (.008)
	0.1	0.5	.777 (.016)	.803 (.017)	.837 (.007)	.688 (.017)	.824 (.006)	.839 (.008)
		1.5	.778 (.011)	.810 (.009)	.851 (.007)	.688 (.015)	.827 (.007)	.853 (.007)
20	2.0	0.5	.765 (.012)	.811 (.010)	.902 (.009)	.906 (.010)	.839 (.009)	.910 (.009)
		1.5	.755 (.027)	.811 (.024)	.843 (.022)	.923 (025)	.843 (.018)	.924 (.026)
	0.5	0.5	.771 (.015)	.802 (.014)	.903 (.010)	.897 (.013)	.836 (.008)	.914 (.010)
		1.5	.770 (.021)	.805 (.020)	.940 (.010)	.945 (.011)	.844 (.010)	.952 (.010)
	0.25	0.5	.782 (.009)	.801 (.016)	.882 (.009)	.856 (.013)	.833 (.006)	.893 (.009)
		1.5	.761 (.026)	.830 (.014)	.924 (.012)	.918 (.015)	.849 (.012)	.925 (.014)
	0.1	0.5	.782 (.013)	.804 (.014)	.837 (.009)	.698 (.021)	.824 (.008)	.845 (.010)
		1.5	.786 (.009)	.808 (.011)	.852 (.008)	.710 (.018)	.826 (.007)	.860 (.008)
Time			0.31	0.35	2.39	3.75	0.77	—

Note: M-sse-full is computationally intensive under a fully Bayesian framework for simulation studies, so we fixed the spatial parameters as the true values, and the corresponding computing time is not listed here for comparison.

From Table 3.1, we observed that the AUC was substantially improved by modeling the patient heterogeneity in the mpMRI parameters (M-base versus M-sse) in all scenarios. We observed big improvement in AUC due to modeling the spatial correlation between voxels (M-sse versus the four models with spatial modeling) given large σ^2 (large spatial variance), large ϕ (larger scale spatial correlation) and large ν (less differentiability). Relative performance of the models with spatial modeling varied by scenario (Table 3.1, columns 6-9), with σ^2 having little effect but ϕ and ν having a noticeable effect on performance. The M-sse-nngp model had similar AUC as the full model (M-sse-full) when ϕ equaled 0.1 and 0.25 (small-scale, local correlation) regardless of ν, and also performed well when ϕ equaled 0.5 and 2 with the exception of the scenario when $\nu = 1.5$. The M-sse-rr model, on the other hand, had an average AUC closer to that of the full model when ϕ equaled 0.5 or 2 (larger-scale correlation), but performed poorly when ϕ equaled 0.1 or 0.25 (local correlation), having worse performance than M-base in some cases (see the results when $\sigma^2 = 1$, $\phi = 0.1$ and $\nu = 1.5$). Both M-sse-nngp and M-sse-rr had an average AUC closer to that of the full model when ν was smaller (larger differentiability). In general, M-sse-nngp had an average AUC close to that of the full model under most data scenarios except when $\phi = 2$ and $\nu = 1.5$, where there was smooth and very strong correlation across the whole image (see Figure 3.1); in contrast, M-sse-nngp did not approximate the performance of the full model well unless the spatial correlation had a large scale ($\phi = 2$). One possible explanation for the poor performance of M-sse-rr under certain scenarios is that 10 knots is not adequate for the reduced-rank approximation to capture the underlying spatial structure well, unless there is strong, large-scale spatial correlation across the whole image. In fact, our simulation results indicated that a much larger number of knots ($a \geq 100$) is required for good performance, which would lead to a much longer computation time and thus was not conducted as part of our simulation study. The M-sse-nngp model, on the other hand, demonstrated its advantage in that a much smaller number of nearest neighbors, in other words, much less computation time, is required, to approximate the classification accuracy

of the full model. The M-sse-car model had improved AUC compared with M-sse, but overall did not approximate the full model well. The performance of M-sse-car was not affected by σ^2, ϕ or ν since the CAR model assumes a fixed spatial structure with no spatial correlation parameters. In summary, the M-sse-nngp model demonstrated robust performance, offering the highest classification accuracy and the best approximation to the full model compared with M-sse-rr and M-sse-car in the majority of the considered scenarios.

3.3.2 Application to in Vivo Data

We next illustrate the performance of the candidate models on the motivating mpMRI data set described in Section 2.2. Our preliminary analyses suggest that the between-patient variation in mpMRI parameters has a complex pattern that might be hard to estimate with a limited sample (34 subjects). Therefore, we applied each proposed spatial modeling approach both with and without subject specific effects (SSE) accounting for the between-patient variability, which lead to M-nnngp and M-sse-nngp: spatial modeling using NNGP without and with SSE; M-rr and M-sse-rr: spatial modeling using reduce-rank approximation without and with SSE; and M-car and M-sse-car: spatial modeling using the CAR model without and with SSE. Summaries of the ROC curve were obtained using 5-fold Cross-Validation to account for over-fitting due to training and evaluating the model on the same data set (Hastie et al., 2009).

Table 3.2: Model performance on the motivating mpMRI data set. The candidate models include: "M-base": the baseline model (3.1) in Section 3.2.1, "M-nnngp" and "M-sse-nngp": spatial modeling using NNGP without and with SSE, respectively, "M-rr" and "M-sse-rr": spatial modeling using reduce-rank approximation without and with SSE, respectively, and "M-car" and "M-sse-car": spatial modeling using the CAR model without and with SSE, respectively. Bayesian inference and classification were based on two chains of 75000 MCMC iterations, after a burn-in stage of 5000 iterations. The "AUC" row lists the average AUC obtained from 5-fold Cross-Validation, the "S80" row lists the sensitivity corresponding to 80% specificity, and the "Time" row lists the computing time in hours to complete the 5-fold Cross-Validation.

	M-base	NNGP		Reduced-rank		CAR	
		M-nngp	M-sse-nngp	M-rr	M-sse-rr	M-car	M-sse-car
AUC	0.763	0.808	0.785	0.807	0.774	0.764	0.765
S80	0.615	0.673	0.641	0.680	0.637	0.607	0.605
ϕ	—	1.67 (1.49, 1.82)	1.64 (1.43, 1.85)	0.71 (0.60, 0.81)	0.76 (0.63, 0.85)	—	—
ν	—	0.88 (0.85, 0.92)	0.89 (0.82, 0.98)	15.89 (13.92, 18.21)	16.03 (13.97, 18.56)	—	—
σ^2	—	657.9 (619.4, 708.9)	632.2 (598.3, 694.0)	765.6 (702.2, 826.1)	720.5 (562.9, 791.3)	4.7 (3.5, 6.3)	4.5 (3.4, 6.3)
Time	8.52	59.78	59.96	96.91	97.19	19.02	19.24

Note: Results for the NNGP and reduced-rank models were obtained using 10 nearest neighbors and 10 knots, respectively. The NNGP using more than 10 neighbors gave similar AUC and S80. The reduced-rank model could potentially have improved classification if we use much more knots ($a = 100$, for example). However, the MCMC algorithm becomes much more computationally intensive, and in fact is unrealistic to apply to the data.

Table 3.2 summarizes the performance of the candidate models, including AUC, S80 (sensitivity corresponding to 80% specificity), posterior mean of the spatial parameters and computing time in hours to complete the 5-fold Cross-Validation. Regarding the AUC and S80, spatial modeling with NNGP and reduced-rank approximation demonstrated improvements both with and without SSE compared with M-base: the AUC increased from 0.763 to 0.808 (M-nngp), 0.785 (M-sse-nngp), 0.807 (M-rr) and 0.774 (M-sse-rr), and S80 was increased from 0.615 to 0.673 (M-nngp), 0.641 (M-sse-nngp), 0.680 (M-rr) and 0.637 (M-sse-rr). This suggests that spatial modeling improved the classification accuracy by successfully modeling the spatial correlation in the data. The M-car and M-sse-car models, however, had performance almost equivalent to M-base both without SSE (AUC: 0.764, S80: 0.607) and with SSE (AUC: 0.765, S80: 0.605). For all three spatial modeling approaches, adding SSE to account for patient heterogeneity did not result in additional improvement in AUC or S80, and, in fact, lowered the AUC and S80 of M-nngp and M-rr. One possible explanation is that the SSE might have a complex distribution, with features that cannot be reflected by the assumed multivariate normal distribution.

The three different spatial modeling approaches lead to different posterior estimates for the spatial parameters. For the NNGP-based models, the posterior mean spatial range ($\hat{\phi}$), smoothness ($\hat{\nu}$) and variance ($\hat{\sigma^2}$) were 1.67, 0.88 and 657.9, respectively, without SSE, and 1.64, 0.89 and 632.2, respectively, with SSE. For the models using reduced-rank approximation, $\hat{\phi}$, $\hat{\nu}$ and $\hat{\sigma^2}$ were 0.71, 15.89 and 765.6, respectively, without SSE, and 0.76, 16.03 and 720.5, respectively, with SSE. For the models using CAR for spatial modeling, there is no assumed ϕ or ν, and the posterior mean of σ^2 was 4.7 without SSE and 4.5 with SSE. It can be observed that, for each spatial modeling approach, the estimated spatial parameters were similar with or without SSE. For the models without SSE, M-nngp and M-rr both gave large estimated spatial variances $\hat{\sigma^2}$, indicating strong spatial variation among voxels within each image. Although the estimated range $\hat{\phi}$ differed (1.67 for M-nngp and 0.71 for M-rr), both indicated large-scale spatial dependency (see Figure 3.1). The M-nngp

model gave a smaller $\hat{\nu} = 0.88$, possibly because the NNGP tends to capture local spatial dependencies that are less smooth. The M-rr model, on the other hand, gave a larger $\hat{\nu} = 15.89$ indicating a much smoother correlation structure, possibly because reduced-rank approximation tends to capture large-scale, smooth spatial dependencies. In contrast, the CAR model assumed a fixed spatial dependence structure that was determined by the distance between voxels, which was quite different from the estimated spatial dependence structures using NNGP or reduced-rank approximation, and thus the M-car model gave a quite different estimated spatial variance $\hat{\sigma^2} = 4.7$. There was also substantial variability in the computational intensity. For the 5-fold Cross-Validation with 75000 MCMC iterations after a burn-in stage of 5000 iterations, the NNGP-based models were approximately 3.1 times slower than the CAR-based models, and the models using reduced-rank approximation were approximately 4.9 times slower than the CAR-based models. Computing time for the various models are reported in Table 3.2.

Figure 3.2 and Figure 3.3 show heatmaps and classification maps, respectively, for two representative prostate images. In the heatmaps, warmer color indicates higher posterior predictive cancer probability, and in the classification maps, red, yellow, grey and blue indicate correctly identified cancer regions (true positive), incorrectly identified cancer regions (false positive), correctly identified noncancer regions (true negative), and incorrectly identified noncancer regions (false negative), respectively, using the probability cut-off corresponding to 80% specificity. Compared with M-base, which generated diffuse areas of high posterior cancer probabilities scattered throughout the images, all spatial models identified larger and more integrated true positive areas. The models using NNGP and reduced-rank approximation produced prediction maps with clearly distinguished cancer and noncancer regions. This is likely because the estimated ϕ and ν for the two models indicate strong, large-scale spatial correlation in the voxel-wise cancer probabilities, which leads to strongly clustered regions of cancer or noncancer voxels. Prediction maps from the CAR-based models are much more smooth, and it is hard to distinguish the cancer and

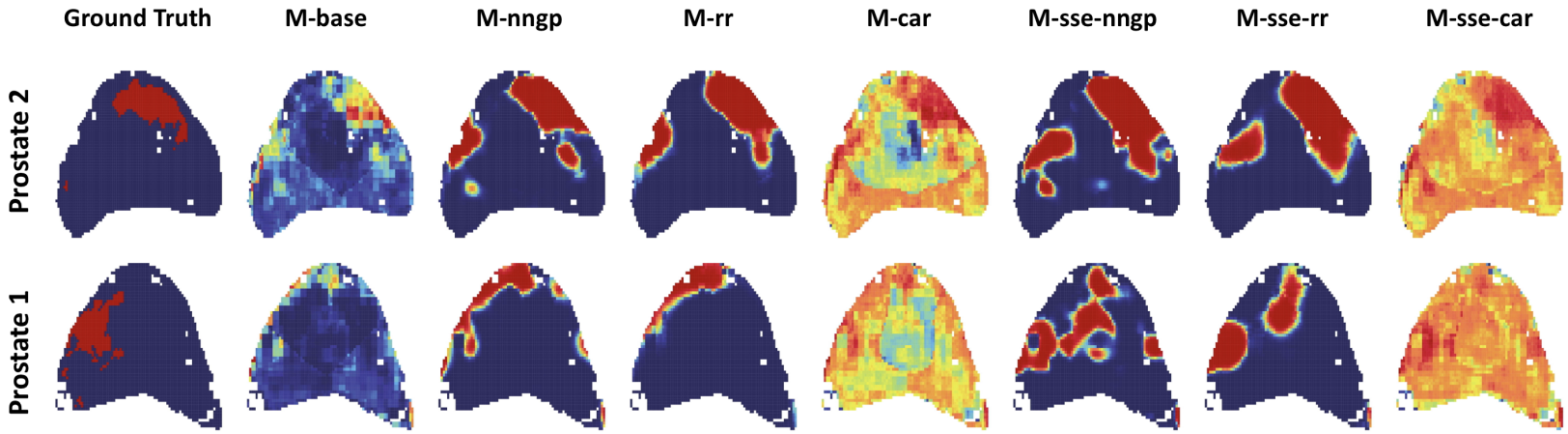

Figure 3.2: Maps of two representative prostates: ground truth (column 1, red regions represent the registered areas of cancer), heatmaps of posterior predictive cancer probabilities using M-base (baseline model), M-nngp (NNGP without SSE), M-rr (reduced-rank without SSE), M-car (CAR without SSE), M-sse-nngp (NNGP with SSE), M-sse-rr (reduced-rank with SSE), M-sse-car (CAR with SSE), respectively (columns 2-8, where warmer color indicates higher posterior predictive cancer probability, and the color was scaled by the range of the posterior predictive cancer probabilities within each image). The white dots in the heatmaps indicate missing values for at least one mpMRI parameter for the voxels.

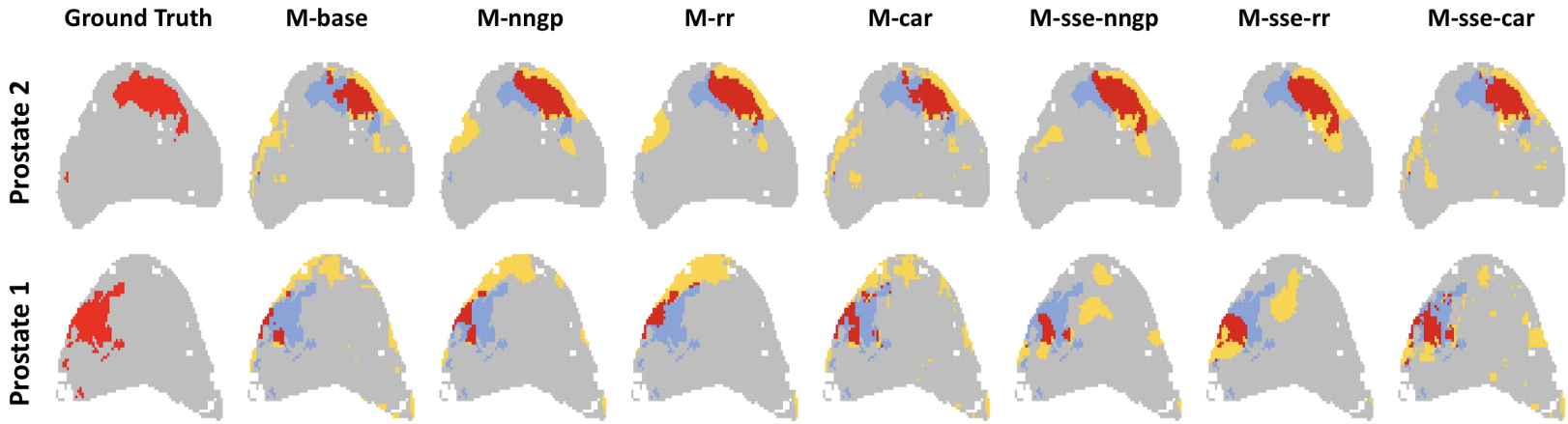

Figure 3.3: Classification maps of two representative prostates that categorize voxels into true positive (red), false positive (yellow), true negative (grey) and false negative (blue), using the probability cut-off corresponding to 80% specificity. Models include M-base (baseline model), M-nngp (NNGP without SSE), M-rr (reduced-rank without SSE), M-car (CAR without SSE), M-sse-nngp (NNGP with SSE), M-sse-rr (reduced-rank with SSE), M-sse-car (CAR with SSE), respectively (columns 2-8).

noncancer regions by eye. Using a probability cut-off that corresponds to 80% specificity, however, we can see from Figure 3.3 that the CAR-based models had similar classification results as M-base. In addition, we see that models that did not account for patient heterogeneity (no SSE) generated prediction maps closer to the ground truth than models with SSE: compared with M-nngp and M-rr, M-sse-nngp and M-sse-rr identified more areas of high posterior cancer probabilities which lead to more true positive areas, but also larger regions of false positives around true positives.

3.4 Discussion

In this chapter, we proposed Bayesian hierarchical models for high-resolution mpMRI data, which aim to improve the voxel-wise classification accuracy of prostate cancer by modeling the between-patient heterogeneity in the mpMRI parameters and spatial correlation in the voxel-wise cancer status. Conventional spatial models become computationally infeasible due to the large size of the multi-image mpMRI data. We consider three computationally efficient approaches to introduce spatial correlation at the voxel-level cancer classification, which use a sparse approximation via the NNGP prior, a reduced-rank approximation through a knot-based approach and GMRF with a CAR model.

Simulation results indicate that classification accuracy can be substantially improved by modeling both the patient heterogeneity in the mpMRI parameters and spatial correlation structure within an image. The proposed M-sse-nngp model performed better under more local, smaller-scale spatial dependency, but was robust to the true spatial structure and had the best overall performance under the considered data scenarios. The proposed M-sse-rr model outperformed M-sse-nngp only when there was strong correlation across the whole image. Otherwise, it performed poorly, performing worse than non-spatial models when there was only small-scale, local correlation. The proposed M-sse-car model improved the AUC relative to the non-spatial models, but performed worse than M-sse-nngp under all

considered data scenarios. Application on our motivating data showed that spatial modeling using NNGP and reduced-rank approximation improved the average AUC and S80 compared with the base-model, while spatial modeling using the CAR model did not. Modeling between-patient heterogeneity in the mpMRI parameters on top of spatial modeling did not further improve the average AUC and S80, and, in some cases, actually decreased performance. This is possibly because the between-patient variation has a complex structure that is hard to estimate with only 34 patients. In fact, our preliminary analysis suggests that the distribution of the subject specific effects may be bimodal, in which case a more complex random effects distribution would be required to properly model the data, which, in turn, would be more difficult to estimate. Further investigation is required for the proper modeling of patient heterogeneity in the mpMRI parameters.

A major challenge for model application is the computational burden brought by simultaneously modeling and classifying multiple high-dimensional images with spatial correlation structures. Among the proposed spatial modeling approaches, CAR demonstrated the highest computational efficiency, but had a lower classification accuracy compared with NNGP and reduced-rank approximation. Although the proposed M-nngp and M-rr models demonstrated similar classification accuracy on the motivating data, M-nngp was approximately 1.6 times faster than M-rr with equal number of nearest neighbors ($m = 10$) and knots ($a = 10$), indicating an advantage of M-nngp in practical applications. The computational intensity increases linearly with the increase in m for NNGP, and increases slightly faster with the increase in a for the reduced-rank model. Application results showed that M-nngp using more than 10 neighbors did not add much improvement in AUC and S80. The M-rr model, on the other hand, could potentially have improved classification if we use many more knots (e.g. $a = 100$). However, the MCMC algorithm will become computational intensive and unrealistic to apply to our data. In summary, the M-nngp model is preferred for real-world application, with robust performance and high computational efficiency. Regarding the model implementation, parallel computing also plays an impor-

tant role, where in each MCMC iteration, it significantly reduces the computation time by updating model components for each subject (image) simultaneously using multiple cores.

In this chapter, spatial modeling was conducted assuming a stationary spatial structure, while in reality the spatial dependency may show a more complex pattern. A future extension to the proposed models is to introduce a non-stationary spatial structure. For example, the spatial correlation may change as a function of the location in the prostate and properly modeling this complex structure could improve performance. Another future direction is lesion-wise cancer detection, which has so far received limited discussion in the literature (Litjens et al., 2014; Leng et al., 2018). While the majority of current quantitative mpMRI classifiers for prostate cancer focus on voxel-wise detection, clinical practice requires that the results are ultimately translated into detection of cancer lesions, and determining how best to identify lesions using voxel-wise data is worthy of future investigation.

Chapter 4

A Multi-resolution Super Learner Algorithm for General Voxel-wise Cancer Classification

4.1 Introduction

The goal of this dissertation is to improve voxel-wise classification of prostate cancer by simultaneously modeling the various sources of variability in the mpMRI data. The Bayesian methods in Chapters 2 and 3 provide powerful tools to do this, but also have limitations, including computational costs and challenges associated with extending to more complex cancer classification/prediction problems. For example, there is a need to extend our existing classifiers to predict ordinal outcomes, such as the Gleason score, which is challenging with the framework proposed in Chapters 2 and 3 because of the large number of model parameters.

In this chapter, we propose a machine learning-based approach to voxel-wise classification, which can flexibly account for the various features of the mpMRI data. The method will account for regional heterogeneity in the mpMRI data via ensemble learning, based on the super learner algorithm (Van der Laan et al., 2007). Briefly, we first train a selected classifier globally on the whole prostate gland, then train local classifiers using sub-regions

of the whole gland, under prostate segmentation of different resolutions: 2×2, 3×3, etc., therefore allowing classification to vary between sub-regions. In the end, we average across the global and multi-resolution local classifiers using the super learner algorithm. The spatial correlation between voxels is accounted for by applying the spatial Gaussian kernel smoothing technique in Chapter 2 to the voxel-wise cancer probabilities from the multi-resolution classifiers. This spatial smoothing technique not only accounts for the spatial correlation between voxels, but also reduces random noise in the data. Regarding the between-patient variability in the observed mpMRI parameters, Chapter 3 presented evidence that the current data set is not sufficient for sophisticated modeling of the subject-specific effects, and therefore we put the modeling of between-patient variability aside and do not discuss it in this chapter.

The proposed algorithm will be introduced in detail for the voxel-wise classification of binary prostate cancer status. To illustrate its flexibility, we will also discuss its application to classifying the ordinal clinical significance of prostate cancer. Simulation studies and application to the motivating data set were conducted for both binary and ordinal classification problems. Results illustrate the benefit of the proposed multi-resolution modeling technique in combination with spatial smoothing, in terms of classification accuracy.

The rest of the Chapter is organized as follows. In Section 4.2, we introduce the proposed method for binary cancer classification. In Section 4.3, we present simulation studies and application to the motivating data to illustrate the model performance on classifying binary cancer status. In Section 4.4, we discuss the extension to ordinal outcomes, present simulation results illustrating its performance, and apply our classifier to the motivating data. Section 4.5 summarizes the chapter and discusses related future research.

4.2 Methods

The proposed method accounts for the heterogeneity both between and within different regions of a prostate via an ensemble learning technique called super learner (van der Laan et al., 2006; Van der Laan et al., 2007). The method is composed of two stages. In stage one, we first select a base learner, which can be a statistical model or machine learning algorithm for classification. Next, we segment the 2-D support of the prostate gland, $(-1,1)\times(-1,1)$, into 1×1, 2×2, 3×3, ..., regular grids. Under the $k\times k$ segmentation, $k=1,2,3,\ldots$, we train the selected base learner locally in each of the k^2 sub-regions. In stage two, we use the summary statistics of the classification results from the multi-resolution base learners in stage one as the new covariates, and train a new classification model, which is a weighted combination of the multi-resolution base learners in stage one.

4.2.1 Standard Super Learner Algorithm

First, we provide a brief overview of the standard super learner algorithm. The super learner algorithm was originally proposed by Van der Laan et al. (2007) as a prediction method using ensemble learning. The super learner selects an optimal weighted combination of multiple candidate learners using Cross-Validation, and it has been shown that the super learner performs asymptotically as well as the oracle learner (i.e. the learner that minimizes risk under the true data-generating distribution) in terms of expected risk difference among the family of candidate learners, if the number of candidate learners, K, is polynomial in sample size, n, i.e. $K \leqslant n^q$ for some $q < \infty$. Details of the super learner and its asymptotic properties can be found in van der Laan et al. (2006) and Van der Laan et al. (2007). Intuitively, super learner uses ensemble learning to "average" over multiple candidate learners, therefore capturing the data features extracted by each base learner.

The original super learner algorithm proceeds as follows. Suppose we have n i.i.d. observations, which are denoted as $O_i = (X_i, Y_i) \sim P_0$, $i = 1,2,\ldots,n$. The goal is

to train a regression model, $\hat{\psi}_0(X) = E_0(Y|X)$, of $Y \in \mathcal{Y}$ on $X \in \mathcal{X}$, which is the minimizer of the expectation of a loss function, $E_0 L(O, \psi)$. Note that the super learner is applicable to any parameters that can be defined as minimizers of a loss function, $L(O, \psi)$, over a parameter space, $\mathbf{\Psi}$. Assume that there are K candidate prediction models, Ψ_k, $k = 1, 2, \ldots, K$, each representing a different mapping from the data, $P_n = \{(Y_i, X_i), i = 1, 2, \ldots, n)$, to $\mathbf{\Psi}$, the parameter space of the functions of X. The algorithm conducts a V-fold Cross-Validation to determine the weight of each candidate model in the super learner. Suppose that $v \in \{1, 2, \ldots, V\}$ denotes a split of the sample into a training sample $T(v) \subset \{1, 2, \ldots, n\}$ and a validation sample $V(v) = \{1, 2, \ldots, n\} \backslash T(v)$, with $\bigcup_{v=1}^{V} V(v) = \{1, 2, \ldots, n\}$, and $V(v_1) \bigcap V(v_2) = \emptyset$, $\forall v_1 \neq v_2$. For the k^{th} candidate model, Ψ_k, $k = 1, 2, \ldots, K$, we define the realization of $\hat{\Psi}_k$ on the training set $P_{n,T(v)}$, $v = 1, 2, \ldots, V$, as $\hat{\psi}_{k,v} = \hat{\Psi}_k(P_{n,T(v)})$. For each $v \in \{1, 2, \ldots, V\}$, we then apply $\{\hat{\psi}_{k,v}, k = 1, 2, \ldots, K\}$ to the corresponding validation set $V(v)$ to get predictions $\{Z_i = (\hat{\psi}_{k,v}(X_i), k = 1, 2, \ldots, K)^T, i \in V(v)\}$, i.e. each sample will have a vector of K Cross-Validated predictions Z_i obtained from the K candidate models. After the Cross-Validation process, a new data set $\{(Y_i, Z_i), i = 1, 2, \ldots, n\}$ is constructed, with Z_i being treated as the $K \times 1$ vector of new covariates. This new data set is then used to train a stage-two prediction algorithm, $\tilde{\Psi}$, which is a mapping from $\{(Y_i, Z_i), i = 1, 2, \ldots, n)\}$ to $\tilde{\Psi}(\{(Y_i, Z_i), i = 1, 2, \ldots, n)\}) : \mathcal{Z} \rightarrow \mathcal{Y}$, and will be referred to as the minimum Cross-Validated risk predictor. Based on $\tilde{\Psi}$ and $\{\hat{\psi}_{k,v}, k = 1, 2, \ldots, K, v = 1, 2, \ldots, V\}$, we define a new mapping $\hat{\Psi}^*$ from the original data $P_n = \{(Y_i, X_i), i = 1, 2, \ldots, n\}$ to the predictor $\tilde{\Psi}\big(\{(Y_i, Z_i = (\hat{\Psi}_k(P_{n,T(v_i)})(X_i), k = 1, 2, \ldots, K)^T), i = 1, 2, \ldots, n\}\big)$. The super learner for a new observation X based on the original data P_n is given by:

$$\hat{\Psi}(P_n)(X) = \hat{\Psi}^*\big(P_n\big)\big(\hat{\Psi}_k(P_n)(X), k = 1, 2, \ldots, K\big). \tag{4.1}$$

4.2.2 The Proposed Two-stage Algorithm

The original super learner was proposed to generate an optimal prediction algorithm through weighted combination of multiple candidate learners. Here we propose a general two-stage algorithm for the voxel-wise classification of prostate cancer, which adopts the idea of super learner to account for regional heterogeneity in the mpMRI data. The basic model framework is similar to that of Van der Laan et al. (2007), but, instead of combining learners based on different classification techniques, we combine learners at different resolutions, which allows us to account for regional heterogeneity without a formal hierarchical model.

Suppose that the outcome of interest is the voxel-wise, binary cancer status, c_{ij}'s, which were defined in Section 2.2. The algorithm proceeds as follows. We first select a base learner Ψ, which can be any classification model or algorithm for the voxel-wise cancer status. To account for the potential regional heterogeneity across the prostate gland, we propose to train Ψ separately within each sub-region of the prostate based on multi-resolution segmentations of the whole gland (WG): first, we train the base learner using all the data on the WG (which we denote as A^1), and denote the trained learner as $\hat{\Psi}_1(\boldsymbol{y})$, with $\boldsymbol{y}$ being a vector of mpMRI parameters. The Cross-Validated classification result for the j^{th} voxel in the i^{th} image is then obtained from $\hat{\Psi}_1$ and is denoted as $x_{ij}^1 = \hat{\Psi}_1(\boldsymbol{y_{ij}})$. Second, we segment the 2-D support of the WG, $(-1,1)\times(-1,1)$, into 2×2 equal-sized sub-regions: $A_1^2 = (-1,0)\times(-1,0)$, $A_2^2 = (-1,0)\times(0,1)$, $A_3^2 = (0,1)\times(-1,0)$, $A_4^2 = (0,1)\times(0,1)$. Note that the voxels that fall on the edges of a sub-region can be included in any sub-region that share the edge. We train the base learner within each region separately to obtain the region-specific learners $\{\hat{\Psi}_{2,l}, l = 1,2,3,4\}$. For a new voxel, the classification under the segmentation of resolution 2×2 will be $\hat{\Psi}_2(\boldsymbol{y}) = \sum_{l=1}^4 I(\boldsymbol{s}\in A_l^2)\hat{\Psi}_{2,l}(\boldsymbol{y})$, where $\boldsymbol{s}$ denotes the standardized 2-D coordinate for the voxel. The Cross-Validated classification for the j^{th} voxel in the i^{th} image under the 2×2 segmentation is then obtained and denoted as $x_{ij}^2 = \hat{\Psi}_2(\boldsymbol{y_{ij}})$. Third, we segment the 2-D support of the WG, $(-1,1)\times(-1,1)$, into 3×3

sub-regions: $\{A_l^3 = (a, a+2/3) \times (b, b+2/3), a, b \in \{-1, -1/3, 1/3\}, l = 1, 2, \ldots, 9\}$, train the base learner within each region separately, and get the trained learner under the segmentation of resolution 3×3: $\hat{\Psi}_3(\boldsymbol{y}) = \sum_{l=1}^{9} I(\boldsymbol{s} \in A_l^3)\hat{\Psi}_{3,l}(\boldsymbol{y})$. The Cross-Validated classification for the j^{th} voxel in the i^{th} image under the 3×3 segmentation is then obtained and denoted as $x_{ij}^3 = \hat{\Psi}_3(\boldsymbol{y_{ij}})$. Theoretically, we can continue the segmentation process with $k = 4, 5, 6, \ldots$, i.e. segmenting the whole prostate gland into $\{k^2 | k \in \mathbb{N}^+\}$ sub-regions: $\{A_l^k = (a, a+2/k) \times (b, b+2/k), a, b \in \{-1+2i/k, i = 0, 1, 2, \ldots, k-1\}, l = 1, 2, \ldots, k^2\}$ For now, we only consider $k = 1, 2, 3$, i.e. segmentation of the lowest $K = 3$ resolutions, for illustration.

Step 1 will result in Cross-Validated voxel-wise classification results under each $k \times k$ resolution for $k = 1, 2, 3$, etc. Considering the observed spatial dependency between voxels within each image, the spatial correlation should be accounted for between $\{x_{ij}^k, j = 1, 2, \ldots, n_i\}$, in each image $i = 1, 2, \ldots, N$. To avoid formally modeling the spatial correlation and to maintain scalability, we propose to apply the spatial Gaussian kernel smoother proposed in Section 2.3.5 to $\{x_{ij}^k, j = 1, 2, \ldots, n_i\}$ for each image i, so as to account for the potential spatial correlation structure between voxels, as well as remove random noise from the image.

After the spatial smoothing step, we obtain a vector of spatially smoothed, Cross-Validated classification results for each voxel, which is denoted as $\tilde{\boldsymbol{x}}_{\boldsymbol{ij}} = (\tilde{x}_{ij}^1, \tilde{x}_{ij}^2, \ldots, \tilde{x}_{ij}^K)^T$. We now proceed to the second stage, where a new classification model for c_{ij} is developed using $\tilde{\boldsymbol{x}}_{\boldsymbol{ij}}$ as the vector of new covariates. In other words, we use a weighted combination of the classification results from the multi-resolution base learners in stage one as the final classifier. We denote the trained model as $\Psi^*(\tilde{\boldsymbol{x}})$, where $\tilde{\boldsymbol{x}}$ is the vector of spatially smoothed classification results from the stage-one multi-resolution base learners. Given that the outcome is binary, a generalized linear model with either a logistic or a probit link function can be used as Ψ^*. In this dissertation, we apply the probit regression.

Figure 4.1 summarizes the work flow of the proposed algorithm. An advantage of the

proposed super learner classification algorithm is that, unlike the Bayesian hierarchical models proposed in Chapters 2-3, the super learner algorithm can use any classification method, including the "black-box" machine learning algorithms, as the base learner, while the mpMRI features can still be accounted for in the classification process.

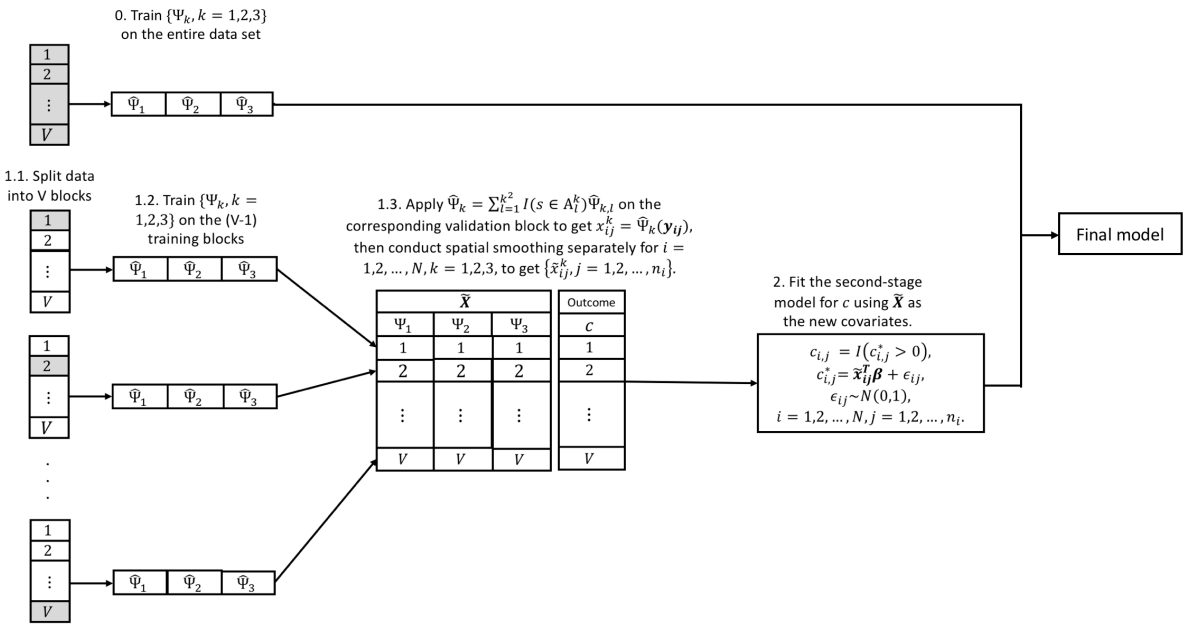

Figure 4.1: Flow chart of the proposed algorithm. The number of resolutions used for prostate segmentation, K, is fixed to 3.

We have only considered $k = 1, 2, 3$, but the proposed segmentation process could continue to finer resolution depending on the availability of the training data and the locality of spatial heterogeneity. A practical problem is that, when k is large, some sub-regions might only have cancer or noncancer voxels. When implementing the proposed super learner algorithm, K, the maximum resolution considered for prostate segmentation, should be chosen to avoid having too many sub-regions with this issue. However, with a few sub-regions having only one cancer status, the super learner algorithm can still be implemented: suppose that under $k \times k$ segmentation, one sub-region, l_k^*, only has voxels with cancer status

$c \in \{0, 1\}$, we then set $\hat{\Psi}_{K,l_k^*}(\boldsymbol{y}) \equiv c$, i.e. the resulting classification in this sub-region will be c with probability 1. Note that this prediction will be averaged with the predictions from lower-resolution classifiers that presumably will not have this issue, therefore resulting in a prediction that is not constantly equal to c.

4.3 Application

4.3.1 Simulation Studies

We conducted simulation studies to illustrate the performance of the proposed classifier under different data generation mechanisms. We considered multiple choices for the base learner, including the generalized linear model with probit link function (which we call "GLM" for simplicity), quadratic discriminant analysis (QDA), and random forest (RF)(Hastie et al., 2009). Given each base learner, we applied the following models to the simulated data sets: (1) the baseline model, which is simply the base learner; (2) the proposed algorithm, but without the step of spatial Gaussian kernel smoothing, which we call "SL0"; and (3) the proposed algorithm, which we call "SL". We hope to evaluate the improvement in classification obtained from the proposed multi-resolution modeling strategy by comparing models (1) and (2), and evaluate the improvement obtained from spatial smoothing by comparing models (2) and (3). In addition, we considered another model, "GLM + QDA + RF", which combines the classification results from multi-resolution GLM, QDA and RF together using super learner. For this model, we also considered the ensemble learner with and without the spatial smoothing step, denoted as "SL" and "SL0", respectively.

The simulated data were generated as follows. The prostate images, including the location of PZ and CG (i.e. voxel-wise region indicators) and standardized 2-D coordinates, were selected with replacement from the images in the motivating data set. Within each image, the voxel-wise cancer status and mpMRI parameters were simulated according to

model (4.2) to introduce regional heterogeneity and spatial correlation between voxels:

$$\begin{aligned}
\boldsymbol{w_i} \sim & \mathcal{MVN}(\mathbf{0}, \boldsymbol{C}(\boldsymbol{S_i}, \boldsymbol{S_i}|\boldsymbol{\theta})), \\
c_{ij}^* \sim & N(q_{r_{ij},0} + w_{ij}, 1), \\
c_{ij} = & I(c_{ij}^* > 0), \\
\boldsymbol{e}_{\boldsymbol{a^k}}^{\boldsymbol{k}} \sim & \mathcal{MVN}(\mathbf{0}, \boldsymbol{\Lambda}), \\
\boldsymbol{\delta_i} \sim & \mathcal{MVN}(\mathbf{0}, \boldsymbol{\Sigma}), \\
\boldsymbol{y_{ij}} \overset{ind}{\sim} & \mathcal{MVN}(\boldsymbol{\mu}_{\boldsymbol{c_{ij}},\boldsymbol{r_{ij}}} + \sum_{k=1}^{K} \boldsymbol{e}_{\boldsymbol{a}_{\boldsymbol{ij}}^{\boldsymbol{k}}}^{\boldsymbol{k}} + \boldsymbol{\delta_i}, \boldsymbol{\Gamma}_{\boldsymbol{c_{ij}},\boldsymbol{r_{ij}}}).
\end{aligned} \tag{4.2}$$

Specifically, we first simulated the spatially correlated $\boldsymbol{w_i} = (w_{i1}, w_{i2}, \ldots, w_{i,n_i})^T$ from $\mathcal{MVN}(\mathbf{0}, \boldsymbol{C}(\boldsymbol{S_i}, \boldsymbol{S_i}|\boldsymbol{\theta}))$ assuming a Matérn correlation function, i.e. the (i, j)-th entry of the spatial covariance matrix $\boldsymbol{C}(\boldsymbol{S_i}, \boldsymbol{S_i}|\boldsymbol{\theta})$ was defined as

$$\boldsymbol{C}(\boldsymbol{s_{ij}}, \boldsymbol{s_{ik}}|\boldsymbol{\theta}) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \Big(\frac{2\nu^{1/2} \big\| \boldsymbol{s_{ij}} - \boldsymbol{s_{ik}} \big\|}{\phi} \Big)^{\nu} \boldsymbol{J}_{\nu} \Big(\frac{2\nu^{1/2} \big\| \boldsymbol{s_{ij}} - \boldsymbol{s_{ik}} \big\|}{\phi} \Big), \tag{4.3}$$

with $\boldsymbol{\theta} = (\sigma^2, \phi, \nu)$. This step was conducted to introduce spatial correlation to the voxel-wise cancer status. Second, we simulated c_{ij}'s independently from $N(q_{r_{ij},0} + w_{ij})$, with $q_{r_{ij}}$ denoting the probit of the overall cancer prevalence in region $r_{ij} = 0$ or 1. Third, we set the voxel-wise cancer status, c_{ij}, to equal to $I(c_{ij}^* > 0)$. The distribution of the mpMRI parameters, $\boldsymbol{y_{ij}}$, were assumed to vary by cancer status, c_{ij}, and region indicator, r_{ij}. We also assumed that the distribution of $\boldsymbol{y_{ij}}$ varied between sub-regions, $\{A_l^k, k = 2, 3, l = 1, 2, \ldots, k^2\}$. Note that K was set to 3. To introduce this regional heterogeneity, for $k = 2, 3$, we generated region-specific random shifts, $\{\boldsymbol{e}_{\boldsymbol{a^k}}^{\boldsymbol{k}}, a^k = 1, 2, \ldots, k^2\}$, from $\mathcal{MVN}(\mathbf{0}, \boldsymbol{\Lambda})$, where a^k is the indicator for sub-region under $k \times k$ segmentation of the WG. In addition, we assumed that there exist subject-specific random shifts from the mean for the i^{th} image/subject, which was denoted as $\boldsymbol{\delta_i}$ and generated from $\mathcal{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$. The

mpMRI parameters were then simulated based on the cancer status c_{ij}, the region indicator r_{ij}, and the sub-region indicator $a_{ij}^k \in \{1, 2, \ldots, k^2, k = 2, 3\}$, from $\mathcal{MVN}(\boldsymbol{\mu}_{\boldsymbol{c_{ij}},\boldsymbol{r_{ij}}} + \sum_{k=1}^{K} \boldsymbol{e}^{\boldsymbol{k}}_{\boldsymbol{a}^{\boldsymbol{k}}_{\boldsymbol{ij}}} + \boldsymbol{\delta_i}, \boldsymbol{\Gamma}_{\boldsymbol{c_{ij}},\boldsymbol{r_{ij}}})$.

The mean, within-patient and between-patient covariance of the mpMRI parameters, and the cancer prevalence in the PZ and CG, were specified based on the estimates from the motivating data set (see Appendix C.1). We varied $q_{r,0}$, $r = 0, 1$, and $\boldsymbol{\Lambda}$, to simulate different magnitudes for the regional heterogeneity. To generate weak regional heterogeneity, we set the cancer prevelance in the PZ to 0.3, i.e. $q_{1,0} = \Phi^{-1}(0.3)$, the cancer prevelance in the CG to 0.25, i.e. $q_{0,0} = \Phi^{-1}(0.25)$, and $\boldsymbol{\Lambda}$ equal to a $d \times d$ diagonal matrix ($d = 4$), with diagonal entries 1, 0.05, 0.05, and 0.05. To simulate moderate regional heterogeneity, we set the cancer prevelance in the PZ and CG to 0.4 and 0.2, respectively, and $\boldsymbol{\Lambda}$ equal to a $d \times d$ diagonal matrix with diagonal entries 5, 0.18, 0.18, and 0.18. To simulate strong regional heterogeneity, we set the cancer prevelance in the PZ and CG to 0.55 and 0.15, respectively, and set $\boldsymbol{\Lambda}$ equal to a $d \times d$ diagonal matrix with diagonal entries 10, 0.36, 0.36, and 0.36. We also varied $\boldsymbol{\theta} = (\sigma^2, \phi, \nu)$ to simulate different spatial correlation structures. For moderate spatial correlation, we set σ^2 (spatial variance) equal to 10, ϕ (range parameter, larger ϕ indicates larger-scale correlation) equal to 2/3, and ν (smoothness parameter, smaller ν indicates larger differentiability) equal to 1. To simulate strong spatial correlation, we set $\sigma^2 = 20$, $\phi = 2$, and $\nu = 1.5$.

In each simulation, we simulated data for 40 prostate images, and obtained the classification results using 5-fold Cross-Validation. In each data scenario, classification results were summarized by the mean and standard deviation of AUC, S80 and S90, across 100 simulations. Given each base learner, the bandwidth for the spatial Gaussian kernel smoother was selected by 5-fold Cross-Validation to maximize the AUC of the corresponding SL model.

Tables 4.1-4.3 present simulation results assuming week, moderate and strong regional heterogeneity, respectively. Among the three considered base learners, GLM and QDA

show similar classification accuracy and outperform RF. As expected, the proposed multi-resolution modeling strategy improves the AUC, S80 and S90 of the base learner ("baseline" v.s. "SL0"), and the improvement increases as the magnitude of the regional heterogeneity increases. The multi-resolution modeling strategy also reduces the standard

Table 4.1: Simulation results for the classification of binary cancer status, assuming that there is weak regional heterogeneity in the data. "Baseline": the base learner; "SL0": the proposed super learner algorithm without the intermediate Gaussian kernel smoothing step; "SL": the proposed super learner algorithm. Results are summarized by the mean and standard deviation (in the parentheses).

Regional Heterogeneity	Spatial Correlation	Base Learner	Method	Classification Results		
				AUC	S80	S90
Weak	Moderate $\sigma^2 = 10$, $\phi = 2/3$, $\nu = 1$.	GLM	Baseline	.785 (.045)	.616 (.082)	.459 (.087)
			SL0	.789 (.022)	.622 (.040)	.464 (.041)
			SL	.808 (.034)	.651 (.068)	.505 (.072)
		QDA	Baseline	.783 (.046)	.612 (.084)	.457 (.087)
			SL0	.784 (.024)	.616 (.044)	.459 (.046)
			SL	.805 (.036)	.648 (.070)	.498 (.075)
		RF	Baseline	.757 (.041)	.567 (.075)	.410 (.077)
			SL0	.766 (.026)	.586 (.047)	.428 (.048)
			SL	.800 (.038)	.638 (.075)	.489 (.078)
		GLM + QDA + RF	SL0	.785 (.024)	.616 (.043)	.457 (.045)
			SL	.796 (.040)	.633 (.077)	.481 (.081)
	Strong $\sigma^2 = 20$, $\phi = 2$, $\nu = 1.5$.	GLM	Baseline	.778 (.070)	.604 (.129)	.453 (.135)
			SL0	.786 (.033)	.618 (.059)	.465 (.062)
			SL	.831 (.045)	.697 (.089)	.571 (.094)
		QDA	Baseline	.772 (.071)	.596 (.127)	.442 (.135)
			SL0	.776 (.037)	.605 (.065)	.453 (.067)
			SL	.820 (.053)	.682 (.097)	.550 (.105)
		RF	Baseline	.746 (.066)	.548 (.115)	.395 (.116)
			SL0	.759 (.040)	.576 (.070)	.422 (.070)
			SL	.815 (.063)	.674 (.110)	.546 (.112)
		GLM + QDA + RF	SL0	.776 (.037)	.603 (.066)	.447 (.069)
			SL	.802 (.061)	.650 (.114)	.514 (.124)

deviation of AUC, S80 and S90 by approximately 50%. The intermediate spatial smoothing step further improves classification, and the improvement increases as the increase in the magnitude, scale and smoothness of the spatial correlation increase. Implementation of the spatial smoothing step increases the standard deviation of AUC, S80 and S90, but the

standard deviation remains smaller than that of the base learner.

Table 4.2: Simulation results for the classification of binary cancer status, assuming that there is moderate regional heterogeneity in the data.

Regional Heterogeneity	Spatial Correlation	Base Learner	Method	Classification Results		
				AUC	S80	S90
Moderate	Moderate $\sigma^2 = 10$, $\phi = 2/3$, $\nu = 1.$	GLM	Baseline	.761 (.048)	.572 (.085)	.410 (.086)
			SL0	.788 (.021)	.620 (.039)	.459 (.039)
			SL	.807 (.033)	.655 (.066)	.498 (.066)
		QDA	Baseline	.759 (.047)	.570 (.082)	.409 (.084)
			SL0	.782 (.022)	.611 (.040)	.451 (.040)
			SL	.801 (.033)	.647 (.064)	.488 (.066)
		RF	Baseline	.731 (.045)	.518 (.076)	.359 (.074)
			SL0	.761 (.023)	.576 (.041)	.415 (.040)
			SL	.796 (.033)	.636 (.064)	.475 (.066)
		GLM + QDA + RF	SL0	.784 (.022)	.615 (.040)	.453 (.040)
			SL	.796 (.037)	.637 (.070)	.474 (.072)
	Strong $\sigma^2 = 20$, $\phi = 2$, $\nu = 1.5.$	GLM	Baseline	.748 (.074)	.553 (.131)	.401 (.134)
			SL0	.775 (.037)	.599 (.067)	.443 (.071)
			SL	.820 (.057)	.685 (.114)	.544 (.126)
		QDA	Baseline	.744 (.075)	.546 (.131)	.393 (.133)
			SL0	.765 (.040)	.586 (.072)	.431 (.075)
			SL	.808 (.059)	.666 (.117)	.523 (.126)
		RF	Baseline	.718 (.072)	.501 (.121)	.349 (.118)
			SL0	.743 (.042)	.551 (.073)	.396 (.073)
			SL	.802 (.061)	.656 (.116)	.511 (.126)
		GLM + QDA + RF	SL0	.766 (.040)	.587 (.070)	.429 (.074)
			SL	.798 (.065)	.648 (.122)	.500 (.133)

An interesting finding from Tables 4.1-4.3 is that, without the intermediate spatial smoothing step, GLM + QDA + RF provides similar classification accuracy to the single-learner based classifier with the highest classification accuracy. This is consistent with our expectation: super learner approximates the optimal learner in terms of minimizing the loss function (not maximizing AUC, S80 or S90), as long as the number of candidate learners is polynomial in sample size. However, after implementing the spatial smoothing step, combining GLM, QDA and RF together does not provide higher classification accuracy than using one single base learner. One possible explanation is that combining the three base learners reduces most of the noise that would otherwise be reduced by spatial smoothing,

and, as a result, there is less room for improvement.

Table 4.3: Simulation results for the classification of binary cancer status, assuming that there is strong regional heterogeneity in the data.

Regional Heterogeneity	Spatial Correlation	Base Learner	Method	Classification Results		
				AUC	S80	S90
Strong	Moderate $\sigma^2 = 10$, $\phi = 2/3$, $\nu = 1$.	GLM	Baseline	.734 (.060)	.526 (.102)	.368 (.096)
			SL0	.778 (.024)	.602 (.043)	.441 (.043)
			SL	.792 (.035)	.627 (.068)	.470 (.071)
		QDA	Baseline	.733 (.059)	.525 (.102)	.368 (.095)
			SL0	.771 (.026)	.594 (.044)	.434 (.044)
			SL	.787 (.037)	.621 (.069)	.461 (.071)
		RF	Baseline	.709 (.050)	.480 (.083)	.324 (.076)
			SL0	.747 (.026)	.554 (.044)	.393 (.044)
			SL	.781 (.036)	.611 (.066)	.452 (.067)
		GLM + QDA + RF	SL0	.775 (.026)	.598 (.045)	.436 (.046)
			SL	.783 (.039)	.612 (.072)	.453 (.074)
	Strong $\sigma^2 = 20$, $\phi = 2$, $\nu = 1.5$.	GLM	Baseline	.740 (.079)	.542 (.127)	.387 (.123)
			SL0	.773 (.045)	.600 (.074)	.444 (.073)
			SL	.816 (.067)	.679 (.113)	.543 (.120)
		QDA	Baseline	.736 (.080)	.538 (.129)	.383 (.125)
			SL0	.762 (.048)	.584 (.077)	.429 (.076)
			SL	.804 (.069)	.660 (.120)	.517 (.128)
		RF	Baseline	.713 (.068)	.491 (.105)	.339 (.100)
			SL0	.739 (.050)	.545 (.077)	.388 (.073)
			SL	.799 (.070)	.650 (.119)	.506 (.124)
		GLM + QDA + RF	SL0	.763 (.050)	.584 (.081)	.427 (.080)
			SL	.791 (.078)	.637 (.132)	.497 (.137)

Note that in Tables 4.1-4.3, the two considered spatial correlation structures both lead to relatively strong correlation between voxels. This is based on our exploratory data analysis of the motivating data set, which indicates strong spatial correlation between voxels. This is not surprising, as prostate cancer arises as clustered cancer lesions, which is consistent with strong spatial correlation in the data generating process. Although results are not reported here, additional simulation studies indicate that there is little additional benefit to spatial smoothing in the presence of weak or no spatial correlation between voxels.

4.3.2 Application to in Vivo Data

In this section, we present the results when our proposed approach is applied to the motivating data set introduced in Section 2.2. Results were obtained using 5-fold Cross-Validation, and are summarized in Table 4.4. When using GLM as the base learner, the multi-resolution modeling strategy improves the AUC from 0.735 to 0.775, the S80 from 0.582 to 0.651, and the S90 from 0.423 to 0.514, and the intermediate spatial smoothing step further improves the AUC to 0.814, the S80 to 0.728, and the S90 to 0.590. Similar improvements can be

Table 4.4: Application results on the motivating data set. "Baseline": base learner; "SL0": the proposed super learner algorithm but without the intermediate spatial Gaussian kernel smoothing step; "SL": the proposed super learner algorithm. The weight for each resolution was averaged by 5-fold Cross-Validation.

Base Learner	Method	Classification Results			Weight for Each Resolution		
		AUC	S80	S90	1×1	2×2	3×3
GLM	Baseline	0.735	0.582	0.423	—	—	—
	SL0	0.775	0.651	0.514	0.818	1.205	1.483
	SL	0.814	0.728	0.590	1.251	1.562	2.104
QDA	Baseline	0.737	0.594	0.431	—	—	—
	SL0	0.761	0.635	0.490	1.301	0.757	0.854
	SL	0.803	0.696	0.569	1.782	1.018	1.329
RF	Baseline	0.692	0.474	0.318	—	—	—
	SL0	0.738	0.557	0.413	0.791	0.639	1.019
	SL	0.831	0.744	0.587	1.879	1.156	1.866
GLM + QDA + RF	SL0	0.778	0.647	0.496	—	—	—
	SL	0.817	0.706	0.550	—	—	—

observed when using QDA or RF as the base learner. Without the spatial smoothing step, the proposed algorithm using GLM performs better than when using QDA or RF as the base learner. However, if the spatial smoothing step is included, the super learner using RF provides higher classification accuracy than the GLM or QDA-based learners (AUC: 0.831 v.s. 0.814 or 0.803, respectively). Without the spatial smoothing step, combining results from multi-resolution GLM, QDA and RF learners provides higher classification accuracy than using a single base learner (AUC: 0.778 (GLM + QDA + RF) v.s. 0.775 (GLM), 0.761

(QDA) and 0.738 (RF)). However, after adding the spatial smoothing step, the super learner using RF as the base learner provides higher classification accuracy than the super learner combing multi-resolution GLM, QDA and RF learners (AUC: 0.831 v.s. 0.817). This is similar to the finding from our simulation studies, which could be explained by the possibility that spatial smoothing reduces variation from spatial correlation and random noise in stage-one, but after combining GLM, QDA, and RF-based learners together, less variation is left in the residual term, and thus the spatial smoothing step provides less improvement.

Table 4.4 also reports the Cross-Validated weight for each resolution, i.e. the estimated coefficients in the stage-two model. In the super learners using GLM and RF as the base learner, the local learner with the highest resolution, 3×3, has the highest weight, while the global learner has the highest weight in the QDA-based learners. Although different resolutions have different weights, the magnitude of the weights are similar, and all weights are significantly nonzero (p-values $< 2 \times 10^{-16}$). This indicates that the base learners trained under all three resolutions have made contributions to the classifiers, which explains why our proposed multi-resolution modeling approach improves classification.

4.4 Extension to Classifying Ordinal Outcomes

In Section 4.2, we introduced a multi-resolution super learner algorithm for classifying binary cancer status. An additional motivation for developing the proposed super learner algorithm is the need to tackle more complex classification/prediction problems for prostate cancer. In this section, we will discuss the extension of our proposed algorithm to ordinal outcomes.

In clinical practice, diagnosing the clinical significance of prostate cancer is a critical step for determining the appropriate treatment. A widely used system for determining the aggressiveness of prostate cancer is the Gleason grading system, which was originally proposed in 1966, refined in 1977, gaining near universal acceptance, and then modified

in 2005 to an updated version that has been widely used since then (Gleason, 1966, 1977; Epstein et al., 2005; Delahunt et al., 2012). The Gleason Score, which ranges from 1 to 5, describes the aggressiveness of prostate cancer based on biopsy (1: healthy tissue, 5: abnormal tissue). For each patient, two Gleason Scores are assigned: the primary score, which is denoted as S_a, describes the largest area of tumor, and the secondary score, which is denoted as S_b, describes the second largest area (Prostate Conditions Education Council, 2019; Jensen et al., 2019). The total Gleason Score, which is written in the form of $S_a + S_b$ and usually ranges from 6 to 10, is then used to evaluate aggressiveness.

Based on the Gleason Score, prostate tissue can be classified into three ordered categories: (1) noncancer tissue, if $S_a \leqslant 3$, and $S_a + S_b < 6$; (2) clinically insignificant cancer tissue, if $S_a + S_b = 3 + 3$ or $3 + 4$; (3) clinically significant cancer, otherwise. The proposed super learner algorithm can be applied to the voxel-wise ordinal classification of the clinical significance of prostate cancer. The statistical problem becomes classifying G_{ij}'s, which, given the corresponding Gleason Score, $S_{a,ij} + S_{b,ij}$, is defined as:

$$G_{ij} = \begin{cases} 1, & \text{if } S_{a,ij} \leqslant 3, \text{ and } S_{a,ij} + S_{b,ij} < 6; \\ 2, & \text{if } S_{a,ij} + S_{b,ij} \in \{3+3, 3+4\}; \\ 3, & \text{if } S_{a,ij} \geqslant 4 \text{ and } S_{b,ij} \geqslant 3. \end{cases} \tag{4.4}$$

4.4.1 Method

Assume that the voxel-wise clinical significance of prostate cancer, G_{ij}'s, have Z ordered levels, as defined above. Implementing the multi-resolution super learner algorithm follows the workflow proposed in Section 4.2.2, but with several modifications. First, the selected base learner should be able to handle ordinal outcome, e.g. ordinal probit regression, QDA, RF. Second, since the outcome is ordinal, there are multiple choices for $\tilde{\boldsymbol{x}}_{\boldsymbol{ij}}$, the summary statistics of the multi-resolution candidate learners, which are used as covariates for the stage-two model. For example, $\tilde{\boldsymbol{x}}_{\boldsymbol{ij}}$ can be either the vector of the predicted probabilities

for the first $Z-1$ cancer categories (or any $Z-1$ out of the Z categories), or the classified cancer category. Third, in order to classify the ordinal outcome, we change the stage-two model, Ψ^*, from probit regression to ordered probit regression:

$$G_{ij} = z \text{ if } a_{z-1} \leqslant G^*_{ij} < a_z, z = 1, 2, \ldots, Z, \tag{4.5}$$

where $\{a_z, z = 0, 1, 2, 3\}$ is the set of boundaries between categories, with $a_0 = -\infty$, $a_Z = \infty$.

Evaluation of the classification accuracy of the proposed classifier is based on the $Z \times Z$ classification table and overall error rate (the percentage of the falsely categorized voxels). In addition, we introduce two more statistics for evaluating the performance on each cancer category $z \in \{1, 2, \ldots, Z\}$: the "False Positive Rate" (FPR), which is the percentage of the voxels truly in category z that are falsely classified as $z' \neq z$, and the "False Discovery Rate" (FDR), which is the percentage of the voxels classified as category z that are actually not in category z:

$$\begin{aligned} FPR(z) &= \frac{\sum_{i=1}^{N}\sum_{j=1}^{n_i} I\left(G_{ij} = z, \hat{G}_{ij} \neq z\right)}{\sum_{i=1}^{N}\sum_{j=1}^{n_i} I(G_{ij} = z)}, \\ FDR(z) &= \frac{\sum_{i=1}^{N}\sum_{j=1}^{n_i} I\left(G_{ij} \neq z, \hat{G}_{ij} = z\right)}{\sum_{i=1}^{N}\sum_{j=1}^{n_i} I(\hat{G}_{ij} = z)}. \end{aligned} \tag{4.6}$$

The prevalence of the Z categories are typically quite different, and detecting the categories of lower prevalence can be difficult due to the limited information provided by the data. In our motivating data set, for example, the prevalence of noncancer voxels ($G = 1$), the clinically insignificant cancer voxels ($G = 2$), and the clinically significant cancer voxels ($G = 3$), are 0.833, 0.058 and 0.109, respectively, and, as a result, the clinically insignificant cancer category can be hard to detect. To increase the power for detecting categories with low prevalence, we propose to implement a weighted likelihood for the or-

dinal probit regression in the second stage of the super learner (Wang, 2001; Hu and Zidek, 2002; Agostinelli and Greco, 2012). Briefly, we denote the likelihood of the ordinal probit regression as $L(\boldsymbol{G}|\boldsymbol{\vartheta}) = \prod_{i=1}^{N}\prod_{j=1}^{n_i} f(G_{ij}|\tilde{\boldsymbol{x}}_{\boldsymbol{ij}}, \boldsymbol{\vartheta})$, with $\boldsymbol{\vartheta}$ denoting the vector of model parameters. The weighted likelihood is defined as:

$$L^w(\boldsymbol{G}|\boldsymbol{\vartheta}) = \prod_{i=1}^{N}\prod_{j=1}^{n_i} f(G_{ij}|\tilde{x}_{ij})^{w_{ij}}, \tag{4.7}$$

where w_{ij} is the user-defined weight for the j^{th} voxel in the i^{th} image. Here we set:

$$w_{ij} = \frac{1}{m_{G_{ij}} \sum_{i=1}^{N}\sum_{j=1}^{n_i} \frac{1}{m_{G_{ij}}}} = \frac{1}{m_{G_{ij}} Z}, \tag{4.8}$$

where m_z denotes the number of voxels in the data set that belong to category $z \in \{1, 2, \ldots, Z\}$. Given this definition, w_{ij} is specified to be proportional to the inverse prevalence of category G_{ij}, so as to upweight the data for the voxels of the rare categories, while downweighting the data for the voxels of the prevalent categories.

4.4.2 Simulation Studies

We conducted a simulation study to evaluate the performance of the proposed super learner algorithm for classifying ordinal outcomes. As in Section 4.3.1, the simulated prostate images were selected with replacement from the images in the motivating data set. Voxel-wise cancer status and mpMRI parameters were simulated according to model (4.9). We set the boundaries, a_1 and a_2, equal to the median and 70-th percentile of the simulated G^*_{ij}'s, respectively, in order to assign the noncancer voxels with the highest prevalence (50%), and the clinically insignificant cancer voxels with the lowest prevalence (20%). We then generated $G_{ij} = I(a_{z-1} \leqslant G^*_{ij} \leqslant a_z)$. The generating process for $\boldsymbol{w_i}$, G^*_{ij}'s, $q_{r,0}$'s, $\boldsymbol{e^k_{a_k}}$'s, $\boldsymbol{\delta_i}$'s and $\boldsymbol{y_{ij}}$'s followed the same procedure as in Section 4.3.1, except that we only

considered moderate or strong regional heterogeneity, and $\boldsymbol{y_{ij}}$ was assumed to vary by the ordinal G_{ij} instead of the binary c_{ij}.

$$
\begin{aligned}
\boldsymbol{w_i} \sim & \mathcal{MVN}(\mathbf{0}, \boldsymbol{C}(\boldsymbol{S_i}, \boldsymbol{S_i}|\boldsymbol{\theta})), \\
G^*_{ij} \sim & N(q_{r_{ij},0} + w_{ij}, 1), \\
G_{ij} = & I(a_{z-1} \leqslant G^*_{ij} < a_z), \\
\boldsymbol{e^k_{a_{k^2}}} \sim & \mathcal{MVN}(\mathbf{0}, \boldsymbol{\Lambda}), \\
\boldsymbol{\delta_i} \sim & \mathcal{MVN}(\mathbf{0}, \boldsymbol{\Sigma}), \\
\boldsymbol{y_{ij}} \overset{ind}{\sim} & \mathcal{MVN}(\boldsymbol{\mu_{G_{ij}, r_{ij}}} + \sum_{k=1}^{K} \boldsymbol{e^k_{a^k_{ij}}} + \boldsymbol{\delta_i}, \boldsymbol{\Gamma_{G_{ij}, r_{ij}}}).
\end{aligned} \tag{4.9}
$$

As in Section 4.3.1, we considered different choices for the base learner, including ordered probit regression (which we denote as "GLM" for simplicity), QDA, and RF. For the stage-two ordered probit model, we considered two sets of weights for the observations: (1) equal weights ($w_{ij} = 1$, $\forall i, j$), which is denoted as "w_1", and (2) weights that are proportional to the inverse prevelance of the corresponding categories (see (4.8)), which is denoted as "w_2". In each simulation, we applied all the following models for each of the three base learners: (1) the baseline model, which is simply the base learner; (2) the proposed algorithm without the step of spatial Gaussian kernel smoothing on $\tilde{x}_{ij}$'s, and with weight w_1, which we call "SL0 + w_1"; (3) the proposed algorithm with weight w_1, which we call "SL + w_1"; (4) the proposed algorithm without the step of spatial Gaussian kernel smoothing on $\tilde{x}_{ij}$'s, and with weight w_2, which we call "SL0 + w_2"; and (5) the proposed algorithm with weight w_2, which we call "SL + w_2". In addition, we considered combining the classification results from the stage-one multi-resolution GLM, QDA and RF as covariates for the stage-two model, which is denoted as "GLM + QDA + RF". For this model, we also considered either conducting or not conducting the intermediate spatial smoothing step, and using either w_1 or w_2 as the weight ("SL0 + w_1", "SL0 + w_2", "SL +

w_1", and "SL + w_2"). As previously mentioned, either the predicted probabilities for the first two ordinal categories or the predicted cancer category obtained from the base learners in stage one could be used as covariates for the stage-two model. Here we chose the predicted probabilities for the first two categories (i.e. $G = 1$ or 2) as the covariates to illustrate model performance. As in Section 4.3.1, we considered either moderate or strong regional heterogeneity. For each scenario, we simulated 100 data sets, with 40 simulated prostate images generated in each data set. The spatial Gaussian kernel smoothing bandwidths for models "SL+w_1" and "SL+w_2" as well as other parameters were selected by 5-fold Cross-Validation to minimize the error rate. We summarized classification results averaged across the 100 simulations using the aforementioned $K \times K$ classification table, FPR, FDR for each category, and the overall error rate.

Table 4.5 and Table 4.6 present simulation results assuming moderate and strong regional heterogeneity, respectively. Here, GLM was used as the base learner for the proposed algorithm. Comparing the baseline model to SL0 + w_1, the multi-resolution modeling strategy provides improvement in the classification of categories 1 and 3, with greater improvement when there is stronger regional heterogeneity. Comparing model SL0 + w_1 to model SL + w_1, the spatial smoothing step gives further improvement in classification of the two extreme categories, which increases with stronger spatial correlation. Overall, the number of correctly identified clinically significant cancer voxels has substantially increased from the baseline model to the four models that use the proposed multi-resolution modeling approach.

Results using QDA or RF as the base learner, or combining all three are summarized in Appendix C.2. Results are similar to those observed with GLM being the base learner. QDA and RF identify more clinically insignificant cancer voxels than GLM, but performance is worse for the other two categories. Regarding models SL0 + w_1, SL + w_1, SL0 + w_2 and SL + w_2: using GLM, QDA, RF or their combination as the base learner provide similar classification accuracy.

Table 4.5: Simulation results assuming moderate regional heterogeneity, and GLM (Ordinal probit regression) as the base learner.

Regional Heterogeneity	Spatial Correlation	Method	True Category	Classification Results					
				Classification Table			FPR	FDR	Overall Error Rate
				1	2	3			
			1	58625	0	10474	0.15	0.39	
		Baseline	2	17059	0	10580	1.00	NA	0.43
			3	20628	0	20832	0.50	0.50	
				1	2	3			
			1	59228	0	9871	0.14	0.38	
		SL0 + w_1	2	16789	0	10850	1.00	NA	0.41
			3	19744	0	21715	0.48	0.49	
				1	2	3			
	Moderate		1	60186	0	8914	0.13	0.36	
	$\sigma^2 = 5$,	SL + w_1	2	17644	0	9995	1.00	NA	0.39
	$\phi = 0.5$,		3	17022	0	24437	0.41	0.44	
	$\nu = 1$.			1	2	3			
			1	44811	12454	11834	0.35	0.29	
		SL0 + w_2	2	8802	6414	12424	0.77	0.76	0.46
			3	9709	7685	24066	0.42	0.50	
				1	2	3			
			1	45016	14998	9085	0.35	0.26	
		SL + w_2	2	9581	7824	10234	0.72	0.76	0.44
			3	6354	10021	25085	0.39	0.43	
				1	2	3			
Moderate			1	58222	0	10813	0.16	0.39	
		Baseline	2	16927	0	10686	1.00	NA	0.43
			3	20629	0	20792	0.50	0.51	
				1	2	3			
			1	59191	0	9844	0.14	0.38	
		SL0 + w_1	2	16981	0	10632	1.00	NA	0.42
			3	20384	0	21037	0.49	0.50	
				1	2	3			
	Strong		1	60984	180	7871	0.12	0.33	
	$\sigma^2 = 20$,	SL + w_1	2	16657	118	10838	1.00	NA	0.36
	$\phi = 2$,		3	14613	205	26603	0.36	0.42	
	$\nu = 1.5$.			1	2	3			
			1	43920	12189	11926	0.35	0.29	
		SL0 + w_2	2	8769	6347	12397	0.77	0.76	0.45
			3	9556	7663	24201	0.42	0.50	
				1	2	3			
			1	46422	15759	6854	0.33	0.22	
		SL + w_2	2	8522	9055	10036	0.67	0.75	0.41
			3	4838	10646	25936	0.37	0.39	

Table 4.6: Simulation results assuming strong regional heterogeneity, and GLM as the base learner.

Regional Heterogeneity	Spatial Correlation	Method	True Category	Classification Results					
				Classification Table			FPR	FDR	Overall Error Rate
				1	2	3			
			1	58823	0	10545	0.15	0.40	
		Baseline	2	17636	0	10111	1.00	NA	0.44
			3	22182	0	19439	0.53	0.52	
				1	2	3			
			1	59642	0	9726	0.14	0.38	
		SL0 + w_1	2	16839	0	10908	1.00	NA	0.41
			3	20126	0	21495	0.48	0.49	
				1	2	3			
	Moderate		1	60428	0	8940	0.13	0.36	
	$\sigma^2 = 5$,	SL + w_1	2	17653	0	10094	1.00	NA	0.39
	$\phi = 0.5$,		3	17137	0	24484	0.41	0.44	
	$\nu = 1$.			1	2	3			
			1	45195	12415	11757	0.35	0.29	
		SL0 + w_2	2	8800	6412	12536	0.77	0.76	0.46
			3	9850	7819	23953	0.42	0.50	
				1	2	3			
			1	45084	15301	8983	0.35	0.26	
		SL + w_2	2	9554	7936	10257	0.71	0.76	0.44
			3	6258	10312	25050	0.40	0.43	
				1	2	3			
Strong			1	58222	0	10813	0.16	0.39	
		Baseline	2	16927	0	10686	1.00	NA	0.43
			3	20629	0	20792	0.50	0.51	
				1	2	3			
			1	59191	0	9844	0.14	0.38	
		SL0 + w_1	2	16981	0	10632	1.00	NA	0.42
			3	20384	0	21037	0.49	0.50	
				1	2	3			
	Strong		1	60984	180	7871	0.12	0.33	
	$\sigma^2 = 20$,	SL + w_1	2	16657	118	10838	1.00	NA	0.36
	$\phi = 2$,		3	14613	205	26603	0.36	0.42	
	$\nu = 1.5$.			1	2	3			
			1	43920	12189	11926	0.35	0.29	
		SL0 + w_2	2	8769	6347	12397	0.77	0.76	0.45
			3	9556	7663	24201	0.42	0.50	
				1	2	3			
			1	46422	15759	6854	0.33	0.22	
		SL + w_2	2	8522	9055	10036	0.67	0.75	0.41
			3	4838	10646	25936	0.37	0.39	

One noticeable finding, which was mentioned previously in Section 4.4.1, is that correctly identifying the intermediate category is difficult with equal weights w_1. When using GLM as the base learner, both the baseline model and the proposed super learner algorithm with equal weights w_1 cannot identify any clinically insignificant cancer voxel. Although QDA and RF can correctly identify a small proportion of the clinically insignificant cancer voxels in the baseline model, the corresponding super learners with equal weights w_1 fail to identify any clinically insignificant cancer voxels, although better overall performance is achieved. Distinguishing the intermediate cancer category from the other two is challenging, due to the small differences in sample mean that are hard to detect given the large sample variance, as well as limited amounts of information provided by the less frequently occurring category in the data. Given the low prevalence of clinically insignificant cancer voxels, the proposed super learner algorithm using equal weights sacrifices classification of the intermediate cancer category to improve the overall classification. The small differences in sample mean between categories and relatively large sample variances are features/qualities of the mpMRI data, and cannot be addressed by the multi-resolution modeling strategy. To upweight the categories that have lower prevalence, we consider the set of weights, w_2, where for each observation/voxel, the weight is inversely proportional to the prevalence of the corresponding category. With weight w_2, detection of clinically insignificant cancer and clinically significant cancer, both of which have lower prevalence than noncancer, has been greatly improved. However, the improvement comes with the price of lower FPR and FDR for noncancer and slightly lower overall error rate. Although the overall ordinal classification is not satisfying, the improved capability of the weighted model for identifying intermediate and clinically significant cancer voxels could be important in clinic practice.

4.4.3 Application to in Vivo Data

We now discuss the performance of the proposed super learner algorithm with the motivating data set. Table 4.7 reports the ordinal classification results of the baseline model ("baseline"), model SL0 + w_1, SL + w_1, SL0 + w_2, and SL + w_2, using GLM (ordinal probit regression) as the base learner. Results were obtained using 4-fold Cross-Validation. For each voxel, the covariates for the stage-two model, $\tilde{\boldsymbol{x}}_{\boldsymbol{ij}}$, were set to be either the predicted probabilities for the first two categories or the classified categories obtained from the multi-resolution base learners in stage one.

Comparing the baseline model to SL0 + w_1, the proposed multi-resolution modeling approach with equal weights leads to lower overall error rates, with decreased FPRs and FDRs for the clinically significant cancer category, slightly increased FPRs and FDRs for noncancer category, and no identification of the clinically insignificant cancer voxels. Comparing SL0 + w_1 to SL + w_1, spatial smoothing of the predictions obtained from the base learners further improves classification. Comparing SL0 + w_1 to SL0 + w_2, using unequal weights enables the proposed multi-resolution modeling approach to correctly identify some clinically insignificant cancer voxels, if using the predicted probabilities as the covariates for stage-two model. However, it inevitably comes with a price: a higher FPR for the noncancer category. Comparing SL0 + w_2 to SL + w_2, the spatial smoothing further reduces the FPRs and FDRs for clinically insignificant and significant cancer categories, but also results in increased overall error rates and higher FPRs for the noncancer category. With equal weight w_1, using predicted probabilities and using classified cancer categories as the covariates for the stage-two model give similar ordinal classification results. However, using w_2 and the predicted probabilities as the covariates tends to identify more cancer voxels, leading to lower FPRs for the clinically insignificant and significant cancer categories, but also higher overall error rates, higher FPRs for the noncancer category, and higher FDRs for the clinically significant and insignificant cancer categories.

Table 4.7: Ordinal classification results on in vivo data assuming GLM as the base learner.

Covariates for stage-two model	Method	True Category	Classification Results: Classification Table			FPR	FDR	Overall Error Rate
			1	2	3			
Classified Categories	Baseline	1	88492	0	495	0.006	0.162	0.165
		2	6151	0	92	1.000	NA	
		3	10930	0	777	0.934	0.430	
			1	2	3			
	SL0 + w_1	1	87850	0	1137	0.013	0.151	0.159
		2	6083	0	160	1.000	NA	
		3	9583	0	2124	0.819	0.379	
			1	2	3			
	SL + w_1	1	87880	0	1107	0.012	0.139	0.144
		2	6118	0	125	1.000	NA	
		3	8038	0	3669	0.687	0.251	
			1	2	3			
	SL0 + w_2	1	86039	0	2948	0.033	0.131	0.155
		2	5524	0	719	1.000	NA	
		3	7437	0	4270	0.635	0.462	
			1	2	3			
	SL + w_2	1	78557	7202	3228	0.117	0.084	0.198
		2	4219	1271	753	0.796	0.887	
		3	3018	2765	5924	0.494	0.402	

We also considered using QDA or RF as the base learner, or combining the multi-resolution GLM, QDA and RF. Results are summarized in Appendix C.2. Using GLM, QDA or RF as the base learner provides similar classification results, while combining GLM, QDA and RF together identifies more clinically significant and insignificant cancer voxels, leading to lower FPRs for the clinically significant and insignificant cancer categories, but also potentially higher FPRs for the noncancer category and higher overall error rates.

4.5 Discussion

In this chapter, we propose a multi-resolution super learner algorithm for voxel-wise classification of prostate cancer. Our major contribution is to combine base learners trained under different resolutions (i.e. global v.s. local classifiers) to account for regional heterogeneity in the prostate. In addition, we use spatial Gaussian kernel smoothing to reduce noise in the presence of strong spatial correlation in the data, which is computationally fast compared to the spatial modeling approaches proposed in Chapter 3. Different from the Bayesian hierarchical models proposed in Chapters 2 and 3, the proposed super learner algorithm can implement any classifiers, including black-box machine learning algorithms, as the base learners, thus providing a flexible model framework that could be continuously enhanced and updated. Simulation studies and application to in vivo data both illustrated the promising performance for classifying binary cancer status, including high classification accuracy and scalability.

An important feature of the proposed algorithm is that it can be easily extended to more general classification/prediction problems. As an illustration, we extended the proposed classifier to the clinical significance of prostate cancer. Improved classification was confirmed by simulation studies and application to in vivo data, especially for the category of clinically significant cancer, which could be a meaningful feature that helps guide prostate cancer diagnosis and treatment in clinical practice.

We proposed weighted ordinal regression as the stage-two model to improve classification. Depending on the primary goal, the weights can be adjusted to improve the detection of specific categories. For example, when the primary goal is to improve classification of less prevalent categories, the weight for each voxel could be set equal to the inverse prevalence of the corresponding cancer category. Overall, among the three base learners considered in this chapter: GLM, QDA and RF, GLM (i.e. probit regression for binary outcome, and ordered probit regression for ordinal outcome) is recommended for future

application, given its high classification accuracy and computational efficiency.

Application to the motivating mpMRI data set demonstrates the promising performance of our proposed super learner algorithm. However, while the proposed algorithm outperforms the base learner for classifying cancer significance, classification accuracy is still relatively low, especially for the two cancerous categories. One of the reasons is that the current method distinguishes between different cancer categories by the voxel-level, category-specific distributions of the mpMRI parameters, which, based on the exploratory analysis of the motivating data set, have similar means across categories and large sample variances. This issue cannot be solved by our multi-resolution modeling strategy. To achieve higher classification accuracy for the two cancerous categories, novel methods should be developed for the modeling of mpMRI parameters, to better distinguish between cancer categories.

The proposed super learner model structure can be modified to incorporate more complex model assumptions. In the current super learner, we assume that the coefficients for the stage-two model are fixed across the whole prostate image. Alternately, the weights for each resolution could be allowed to vary by the location in a prostate image. Another modification could be to incorporate the spatial smoothing step in the modeling process rather than as a post-hoc procedure. However, this could diminish the scalability of the resulting classifier.

Chapter 5

Conclusion

5.1 Summary

This dissertation aimed to develop novel statistical methods for voxel-wise prostate cancer detection using mpMRI by exploiting specific features of the data, including regional heterogeneity, spatial correlation, and between-patient heterogeneity.

In Chapter 2, we developed a general Bayesian hierarchical modeling framework that jointly models the voxel-wise mpMRI parameters and cancer status conditional on location information. Two models, Mregion and Mcoord, were proposed that utilize different approaches to modeling regional heterogeneity. In addition, we proposed post-hoc spatial Gaussian kernel smoothing on the voxel-wise posterior predictive cancer probabilities to reduce the effect of residual spatial correlation and random noise. Two advantages of the proposed classifiers are: (1) the available closed form solution for the posterior predictive cancer probabilities, which ensures high computational efficiency; and (2) the ability to handle voxels with missing mpMRI parameters without the need for imputation.

Building on the Bayesian hierarchical modeling framework developed in Chapter 2, Chapter 3 proposed to model the spatial correlation between voxels and the variability in the distribution of mpMRI parameters between patients. Bayesian modeling of the spatial correlation is challenging and even infeasible in the presence of high-dimensional mpMRI

data. As a solution, we considered three computationally feasible spatial modeling strategies, based on the NNGP, a reduced-rank approximation approach, and the CAR model. To account for the between-patient variability in mpMRI, we proposed to add a multivariate normally distributed random shift on the mpMRI parameters for each patient. Our results show that modeling the spatial correlation between voxels results in a substantial improvement in the classification accuracy, but modeling the between-patient variability does not improve classification in our motivating data set.

Chapter 4 proposed a machine learning approach to accounting for regional heterogeneity. Specifically, we propose to separately train base learners at different resolutions, and use a super learner algorithm to combine the base learners into a single multi-resolution classifier. A spatial Gaussian kernel smoother is applied to the classification results obtained from the multi-resolution base learners to reduce the variance from spatial correlation and random noise. Important features of our proposed algorithm include: (1) straightforward extension to ordinal cancer outcomes; (2) flexible modeling framework, where any algorithm, including black-box machine learning algorithms, can be used as the base learner in the super learner algorithm, which gives the algorithm the potential to be continuously upgraded, with more data and better base learners available in the future.

Contributions have been made to the statistical methodology for the diagnosis of prostate cancer using mpMRI through developing both a Bayesian modeling framework and a machine-learning approach. These proposed classifiers have shown a number of advantages over existing methods, and could be a helpful tool for guiding pathologists, urologists and clinicians during cancer diagnosis and treatment decision-making.

5.2 Future Work

In the following paragraphs, we present several potential topics/directions for future research, which are motivated by either the limitations of our proposed methods or extensions

that could improve clinical application.

Regarding spatial modeling, we currently assume stationarity for the spatial correlation structure, which may not be true for our motivating data. A natural extension would be to allow nonstationary patterns in the spatial correlation structure. For example, the NNGP prior in Chapter 3 can be modified so that the spatial variance, range and smoothness parameters are functions of the standardized 2-D coordinates. However, assuming that the spatial parameters vary simply by the location in a prostate may not be realistic, and additional variability between patients/images should be considered as well.

In Chapter 3, we proposed to model the between-patient heterogeneity in the mpMRI parameters by adding a multivariate normally distributed random shift from the mean. Although this approach improved classification for simulated data, it did not improve classification when applied to our motivating data set. As mentioned in Chapter 3, more complex patterns have been observed for the subject specific effects than a multivariate normal distribution; for example, our exploratory data analysis indicated the potential for a bimodal pattern in the random effects, and the random effects distribution may also differ for cancer and non-cancer voxels. We have considered more complex models for the random effects distributions, but these additional model assumptions did not improve classification. Regardless, the between-patient heterogeneity remains an important feature of the data, and should be further investigated once more data are available.

As discussed in Chapter 4, both the baseline model and the proposed super learner algorithm show limitations in distinguishing between clinically significant and insignificant prostate cancer. This is because the voxel-level distribution of mpMRI parameters has small difference in the sample mean between categories that cannot be easily detected given the large sample variance. This issue cannot be simply addressed by our proposed multi-resolution modeling technique. In fact, the difficulty in differentiating the voxel-level distribution of mpMRI parameters between categories leads us to consider modeling the region-level distribution of mpMRI parameters instead. For instance, when conduct-

ing classification for a new prostate image, we can iteratively update our classification by first identifying candidate lesions using a binary classifier of cancer v.s. non-cancer and, then, assess the similarity between the joint distribution of the voxels within candidate lesions and that of the voxels in each cancer category, to determine the aggressiveness of the candidate lesion.

An alternate approach to voxel-wise classification is lesion segmentation/detection using voxel-wise mpMRI data, which has received little attention in the statistical literature. Two recently published papers proposed to first delineate potential lesion candidates, then classify each candidate using feature detection or pattern recognition algorithms (Litjens et al., 2011, 2012). Novel methods for this type of cancer detection would represent an important contribution given the importance of lesion detection in clinical practice.

Regarding the proposed super learner algorithm in Chapter 4, a future extension could be to allow the model coefficients to vary spatially in a prostate image. Assume that we segment the prostate gland, $(-1,1) \times (-1,1)$, into M regular grids, and denote the standardized coordinates for the centers of the grids as $\{\boldsymbol{u_i}, i = 1,2,\ldots,M\}$. The model coefficients, $\boldsymbol{\beta}(\boldsymbol{u_i})$, could be allowed to vary by location $\boldsymbol{u_i}$, and specify the following prior covariance structure:

$$\text{vec}([\boldsymbol{\beta}(\boldsymbol{u_1})|\boldsymbol{\beta}(\boldsymbol{u_2})|\ldots|\boldsymbol{\beta}(\boldsymbol{u_M})]) \sim \mathcal{MVN}(\boldsymbol{0}, \boldsymbol{U} \otimes \boldsymbol{V}), \tag{5.1}$$

where $\boldsymbol{V}$ denotes the nonspatial prior covariance matrix for $\boldsymbol{\beta}(\boldsymbol{u_i})$, and $\boldsymbol{U}$ denotes the $M \times M$ prior spatial covariance matrix between grids. The proposed super learner algorithm with spatially-varying coefficients can be trained in the Bayesian modeling framework. We considered this approach when considering our methodological development in Chapter 4. However, results show that allowing the model coefficients to be spatially-varying does not further improve classification. This is probably because the spatially-varying coefficients, $\boldsymbol{\beta}(\boldsymbol{u})$, may also vary between patients, but the current classification

method cannot incorporate this variability. Novel methods should be developed to implement spatially-varying coefficients for our proposed super learner algorithm.

Chapter 6

Bibliography

Claudio Agostinelli and Luca Greco. Weighted likelihood in Bayesian inference. In *Proceedings of the 46th Scientific Meeting of the Italian Statistical Society*, 2012.

American Cancer Society. Key Statistics for Prostate Cancer. `https://www.cancer.org/cancer/prostate-cancer/about/key-statistics.html`, 2019. Accessed: 2019-04-15.

Yusuf Artan, Masoom A Haider, Deanna L Langer, Theodorus H Van der Kwast, Andrew J Evans, Yongyi Yang, Miles N Wernick, John Trachtenberg, and Imam Samil Yetik. Prostate cancer localization with multispectral MRI using cost-sensitive support vector machines and conditional random fields. *IEEE Transactions on Image Processing*, 19(9):2444–2455, 2010.

Adrian Baddeley, Rolf Turner, et al. Spatstat: an R package for analyzing spatial point patterns. *Journal of statistical software*, 12(6):1–42, 2005.

Sudipto Banerjee, Alan E Gelfand, Andrew O Finley, and Huiyan Sang. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848, 2008.

Jelle O Barentsz, Jonathan Richenberg, Richard Clements, Peter Choyke, Sadhna Verma,

Geert Villeirs, Olivier Rouviere, Vibeke Logager, and Jurgen J Fütterer. ESUR prostate MR guidelines 2012. *European radiology*, 22(4):746–757, 2012.

Smarajit Bose, T Hastie, R Tibshirani, and J Friedman. The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2003.

Andrew Cameron, Farzad Khalvati, Masoom A Haider, and Alexander Wong. MAPS: a quantitative radiomics approach for prostate cancer detection. *IEEE Transactions on Biomedical Engineering*, 63(6):1145–1156, 2016.

Bradley P Carlin and Thomas A Louis. *Bayesian methods for data analysis*. CRC Press, 2008.

Noel Cressie and Gardar Johannesson. Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):209–226, 2008.

Abhirup Datta, Sudipto Banerjee, Andrew O Finley, and Alan E Gelfand. Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812, 2016.

Carl De Boor, Carl De Boor, Etats-Unis Mathématicien, Carl De Boor, and Carl De Boor. *A practical guide to splines*, volume 27. Springer-Verlag New York, 1978.

Brett Delahunt, Rose J Miller, John R Srigley, Andrew J Evans, and Hemamali Samaratunga. Gleason grading: past, present and future. *Histopathology*, 60(1):75–86, 2012.

Louise Dickinson, Hashim U Ahmed, Clare Allen, Jelle O Barentsz, Brendan Carey, Jurgen J Futterer, Stijn W Heijmink, Peter J Hoskin, Alex Kirkham, Anwar R Padhani, et al. Magnetic resonance imaging for the detection, localisation, and characterisation of prostate cancer: recommendations from a European consensus meeting. *European urology*, 59(4):477–494, 2011.

Dirk Eddelbuettel, Romain François, J Allaire, Kevin Ushey, Qiang Kou, N Russel, John Chambers, and D Bates. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18, 2011.

Jo Eidsvik, Benjamin A Shaby, Brian J Reich, Matthew Wheeler, and Jarad Niemi. Estimation and prediction in spatial models with block composite likelihoods. *Journal of Computational and Graphical Statistics*, 23(2):295–315, 2014.

Jonathan I Epstein, William C Allsbrook Jr, Mahul B Amin, Lars L Egevad, ISUP Grading Committee, et al. The 2005 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. *The American journal of surgical pathology*, 29(9):1228–1242, 2005.

Duc Fehr, Harini Veeraraghavan, Andreas Wibmer, Tatsuo Gondo, Kazuhiro Matsumoto, Herbert Alberto Vargas, Evis Sala, Hedvig Hricak, and Joseph O Deasy. Automatic classification of prostate cancer Gleason Scores from multiparametric magnetic resonance images. *Proceedings of the National Academy of Sciences*, 112(46):E6265–E6273, 2015.

Andrew O Finley, Sudipto Banerjee, and Ronald E McRoberts. Hierarchical spatial models for predicting tree species assemblages across large domains. *The annals of applied statistics*, 3(3):1052, 2009.

Chris Fraley and Adrian E Raftery. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association*, 97(458):611–631, 2002.

Felix Franke, Rodrigo Quian Quiroga, Andreas Hierlemann, and Klaus Obermayer. Bayes optimal template matching for spike sorting–combining fisher discriminant analysis with optimal filtering. *Journal of computational neuroscience*, 38(3):439–459, 2015.

Reinhard Furrer, Marc G Genton, and Douglas Nychka. Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3): 502–523, 2006.

Kirema Garcia-Reyes, Niccolò M Passoni, Mark L Palmeri, Christopher R Kauffman, Kingshuk Roy Choudhury, Thomas J Polascik, and Rajan T Gupta. Detection of prostate cancer with multiparametric MRI (mpMRI): effect of dedicated reader education on accuracy and confidence of index and anterior cancer diagnosis. *Abdominal imaging*, 40 (1):134–142, 2015.

Valentina Giannini, Anna Vignati, Simone Mazzetti, Massimo De Luca, Christian Bracco, Michele Stasi, Filippo Russo, Enrico Armando, and Daniele Regge. A prostate CAD system based on multiparametric analysis of DCE T1-w, and DW automatically registered images. In *Medical Imaging 2013: Computer-Aided Diagnosis*, volume 8670, page 86703E. International Society for Optics and Photonics, 2013.

Peter Gibbs, Daniel J Tozer, Gary P Liney, and Lindsay W Turnbull. Comparison of quantitative T2 mapping and diffusion-weighted imaging in the normal and pathologic prostate. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 46(6):1054–1058, 2001.

Donald F Gleason. Classification of prostatic carcinomas. *Cancer Chemother. Rep.*, 50: 125–128, 1966.

Donald F Gleason. The Veterans Administration Cooperative Urological Research Group. Histological grading and clinical staging of prostatic carcinoma. *Urologic pathology: the prostate*, pages 171–198, 1977.

Onur C Hamsici and Aleix M Martinez. Bayes optimality in linear discriminant analysis. *IEEE transactions on pattern analysis and machine intelligence*, 30(4):647–657, 2008.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The elements of statistical learning: data mining, inference, and prediction, Springer Series in Statistics, 2009.

John V Hegde, Robert V Mulkern, Lawrence P Panych, Fiona M Fennessy, Andriy Fedorov, Stephan E Maier, and Clare MC Tempany. Multiparametric MRI of prostate cancer: an update on state-of-the-art techniques and their performance in detecting and localizing prostate cancer. *Journal of magnetic resonance imaging*, 37(5):1035–1054, 2013.

Dave Higdon. Space and space-time modeling using process convolutions. In *Quantitative methods for current environmental issues*, pages 37–56. Springer, 2002.

Feifang Hu and James V Zidek. The weighted likelihood. *Canadian Journal of Statistics*, 30(3):347–371, 2002.

Carina Jensen, Jesper Carl, Lars Boesen, Niels Christian Langkilde, and Lasse Riis Østergaard. Assessment of prostate cancer prognostic Gleason grade group using zonal-specific features extracted from biparametric MRI using a KNN classifier. *Journal of applied clinical medical physics*, 20(2):146–153, 2019.

Jin Jin, Lin Zhang, Ethan Leng, Gregory J Metzger, and Joseph S Koopmeiners. Detection of prostate cancer with multiparametric MRI utilizing the anatomic structure of the prostate. *Statistics in Medicine*, 2018.

Chaitanya Kalavagunta, Xiangmin Zhou, Stephen C Schmechel, and Gregory J Metzger. Registration of in vivo prostate MRI and pseudo-whole mount histology using Local Affine Transformations guided by Internal Structures (LATIS). *Journal of Magnetic Resonance Imaging*, 41(4):1104–1114, 2015.

Cari G Kaufman, Mark J Schervish, and Douglas W Nychka. Covariance tapering for

likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103(484):1545–1555, 2008.

B Michael Kelm, Bjoern H Menze, Christian M Zechmann, Klaus T Baudendistel, and Fred A Hamprecht. Automated estimation of tumor probability in prostate magnetic resonance spectroscopic imaging: pattern recognition vs quantification. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 57(1):150–159, 2007.

Farzad Khalvati, Alexander Wong, and Masoom A Haider. Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature models. *BMC medical imaging*, 15(1):27, 2015.

Charles Kooperberg and Maintainer Charles Kooperberg. Package 'polspline', 2019.

Charles Kooperberg, Smarajit Bose, and Charles J Stone. Polychotomous regression. *Journal of the American Statistical Association*, 92(437):117–127, 1997.

John Kurhanewicz, Daniel Vigneron, Peter Carroll, and Fergus Coakley. Multiparametric magnetic resonance imaging in prostate cancer: present and future. *Current opinion in urology*, 18(1):71, 2008.

Deanna L Langer, Theodorus H Van der Kwast, Andrew J Evans, John Trachtenberg, Brian C Wilson, and Masoom A Haider. Prostate cancer detection with multi-parametric MRI: Logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 30(2): 327–334, 2009.

Junghoon Lee, Jonghye Woo, Fangxu Xing, Emi Z Murano, Maureen Stone, and Jerry L

Prince. Semi-automatic segmentation for 3D motion analysis of the tongue with dynamic MRI. *Computerized Medical Imaging and Graphics*, 38(8):714–724, 2014.

Guillaume Lemaître, Robert Martí, Jordi Freixenet, Joan C Vilanova, Paul M Walker, and Fabrice Meriaudeau. Computer-Aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review. *Computers in biology and medicine*, 60: 8–31, 2015.

Ethan Leng, Benjamin Spilseth, Lin Zhang, Jin Jin, Joseph S Koopmeiners, and Gregory J Metzger. Development of a measure for evaluating lesion-wise performance of CAD algorithms in the context of mpMRI detection of prostate cancer. *Medical physics*, 45 (5):2076–2088, 2018.

Gary P Liney, Adrian J Knowles, David J Manton, Lindsay W Turnbull, Stephen J Blackband, and Anthony Horsman. Comparison of conventional single echo and multi-echo sequences with a fast spin-echo sequence for quantitative T2 mapping: Application to the prostate. *Journal of Magnetic Resonance Imaging*, 6(4):603–607, 1996.

Geert Litjens, Oscar Debats, Jelle Barentsz, Nico Karssemeijer, and Henkjan Huisman. Computer-aided detection of prostate cancer in MRI. *IEEE transactions on medical imaging*, 33(5):1083–1092, 2014.

Geert JS Litjens, Pieter C Vos, Jelle O Barentsz, Nico Karssemeijer, and Henkjan J Huisman. Automatic computer aided detection of abnormalities in multi-parametric prostate MRI. In *Medical Imaging 2011: Computer-Aided Diagnosis*, volume 7963, page 79630T. International Society for Optics and Photonics, 2011.

Geert JS Litjens, Jelle O Barentsz, Nico Karssemeijer, and Henkjan J Huisman. Automated computer-aided detection of prostate cancer in MR images: from a whole-organ to a zone-based approach. In *Medical Imaging 2012: Computer-Aided Diagnosis*, volume 8315, page 83150G. International Society for Optics and Photonics, 2012.

Roderick JA Little and Donald B Rubin. *Statistical Analysis with Missing Data, Second Edition*, pages 200–220, 2002.

Xin Liu and Imam Samil Yetik. Automated prostate cancer localization without the need for peripheral zone extraction using multiparametric MRI. *Medical physics*, 38(6):2986–2994, 2011.

Xin Liu, Deanna L Langer, Masoom A Haider, Yongyi Yang, Miles N Wernick, and Imam Samil Yetik. Prostate cancer segmentation with simultaneous estimation of Markov random field parameters and class. *IEEE Transactions on Medical Imaging*, 28(6):906–915, 2009.

R Lopes, A Ayache, N Makni, P Puech, A Villers, S Mordon, and N Betrouni. Prostate cancer characterization on MR images using fractal features. *Medical physics*, 38(1): 83–95, 2011.

Dongjiao Lv, Xuemei Guo, Xiaoying Wang, Jue Zhang, and Jing Fang. Computerized characterization of prostate cancer by fractal analysis in MR images. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 30(1):161–168, 2009.

Lukasz Matulewicz, Jacobus FA Jansen, Louisa Bokacheva, Hebert Alberto Vargas, Oguz Akin, Samson W Fine, Amita Shukla-Dave, James A Eastham, Hedvig Hricak, Jason A Koutcher, et al. Anatomic segmentation improves prostate cancer detection with artificial neural networks analysis of 1H magnetic resonance spectroscopic imaging. *Journal of Magnetic Resonance Imaging*, 40(6):1414–1421, 2014.

Simone Mazzetti, Massimo De Luca, Christian Bracco, Anna Vignati, Valentina Giannini, Michele Stasi, Filippo Russo, Enrico Armando, Silvano Agliozzo, and Daniele Regge. A CAD system based on multi-parametric analysis for cancer prostate detection on DCE-MRI. *Medical Imaging 2011: Computer-Aided Diagnosis*, 7963:79633Q, 2011.

Geoffrey McLachlan. *Discriminant analysis and statistical pattern recognition*, volume 544. John Wiley & Sons, 2004.

Gregory J Metzger, Chaitanya Kalavagunta, Benjamin Spilseth, Patrick J Bolan, Xiufeng Li, Diane Hutter, Jung W Nam, Andrew D Johnson, Jonathan C Henriksen, Laura Moench, et al. Detection of Prostate Cancer: Quantitative Multiparametric MR Imaging Models Developed Using Registered Correlative Histopathology. *Radiology*, 279(3): 805–816, 2016.

Robert P Myers. Structure of the adult prostate from a clinician's standpoint. *Clinical anatomy*, 13(3):214–215, 2000.

Elizbar A Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.

Emilie Niaf, Olivier Rouvière, Florence Mège-Lechevallier, Flavie Bratan, and Carole Lartizien. Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. *Physics in Medicine & Biology*, 57(12):3833, 2012.

Sedat Ozer, Deanna L Langer, Xin Liu, Masoom A Haider, Theodorus H van der Kwast, Andrew J Evans, Yongyi Yang, Miles N Wernick, and Imam S Yetik. Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI. *Medical physics*, 37(4):1873–1883, 2010.

Sébastien Parfait, Paul Michael Walker, Gilles Créhange, Xavier Tizon, and Johel Miteran. Classification of prostate magnetic resonance spectra using Support Vector Machine. *Biomedical Signal Processing and Control*, 7(5):499–508, 2012.

Yahui Peng, Yulei Jiang, Tatjana Antic, Maryellen L Giger, Scott Eggener, and Aytekin Oto. A study of T2-weighted MR image texture features and diffusion-weighted MR

image features for computer-aided diagnosis of prostate cancer. In *Proc. SPIE*, volume 8670, page 86701H, 2013.

Martyn Plummer. rjags: Bayesian graphical models using MCMC. *R package version*, 3 (10), 2013.

Prostate Conditions Education Council. Gleason score: prostate cancer grading & prognostic scoring. `https://www.prostateconditions.org/about-prostate-conditions/prostate-cancer/newly-diagnosed/gleason-score`, 2019. Accessed: 2019-05-01.

Philippe Puech, Nacim Betrouni, Nasr Makni, Anne-Sophie Dewalle, Arnauld Villers, and Laurent Lemaitre. Computer-assisted diagnosis of prostate cancer using dce-mri data: design, implementation and preliminary results. *International journal of computer assisted radiology and surgery*, 4(1):1–10, 2009.

Andrew B Rosenkrantz, Sooah Kim, Ruth P Lim, Nicole Hindman, Fang-Ming Deng, James S Babb, and Samir S Taneja. Prostate cancer localization using multiparametric MR imaging: comparison of Prostate Imaging Reporting and Data System (PI-RADS) and Likert scales. *Radiology*, 269(2):482–492, 2013.

HÅÅVARD RUE and Hååkon Tjelmeland. Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian journal of Statistics*, 29(1):31–49, 2002.

Havard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. CRC press, 2005.

David Ruppert, Matt P Wand, and Raymond J Carroll. Semiparametric regression during 2003–2007. *Electronic journal of statistics*, 3:1193, 2009.

Vijay Shah, Baris Turkbey, Haresh Mani, Yuxi Pang, Thomas Pohida, Maria J Merino, Peter A Pinto, Peter L Choyke, and Marcelino Bernardo. Decision support system for

localizing prostate cancer based on multiparametric magnetic resonance imaging. *Medical physics*, 39(7):4093–4103, 2012.

Luciano Spinello, Kai Oliver Arras, Rudolph Triebel, and Roland Siegwart. A Layered Approach to People Detection in 3D Range Data. In *AAAI*, volume 10, pages 1–1, 2010.

Michael L Stein. A modeling approach for large spatial datasets. *Journal of the Korean Statistical Society*, 37(1):3–10, 2008.

Michael L Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 2012.

Michael L Stein. Limitations on low rank approximations for covariance matrices of spatial data. *Spatial Statistics*, 8:1–19, 2014.

Michael L Stein, Zhiyi Chi, and Leah J Welty. Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66 (2):275–296, 2004.

Pallavi Tiwari, Anant Madabhushi, and Mark Rosen. A hierarchical unsupervised spectral clustering scheme for detection of prostate cancer from magnetic resonance spectroscopy (MRS). *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*, pages 278–286, 2007.

Pallavi Tiwari, Mark Rosen, and Anant Madabhushi. A hierarchical spectral clustering and nonlinear dimensionality reduction scheme for detection of prostate cancer from magnetic resonance spectroscopy (MRS). *Medical physics*, 36(9Part1):3927–3939, 2009.

Pallavi Tiwari, John Kurhanewicz, and Anant Madabhushi. Multi-kernel graph embedding for detection, Gleason grading of prostate cancer via MRI/MRS. *Medical image analysis*, 17(2):219–235, 2013.

Paul S Tofts. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. *Journal of Magnetic Resonance Imaging*, 7(1):91–101, 1997.

Robert Toth, Justin Ribault, John Gentile, Dan Sperling, and Anant Madabhushi. Simultaneous segmentation of prostatic zones using active appearance models with multiple coupled levelsets. *Computer Vision and Image Understanding*, 117(9):1051–1060, 2013.

Mark J van der Laan, Sandrine Dudoit, and Aad W van der Vaart. The cross-validated adaptive epsilon-net estimator. *Statistics & Decisions*, 24(3):373–395, 2006.

Mark J Van der Laan, Eric C Polley, and Alan E Hubbard. Super learner. *Statistical applications in genetics and molecular biology*, 6(1), 2007.

Aldo V Vecchia. Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2):297–312, 1988.

Satish Viswanath, B Nicholas Bloch, Jonathan Chappelow, Pratik Patel, Neil Rofsky, Robert Lenkinski, Elizabeth Genega, and Anant Madabhushi. Enhanced multi-protocol analysis via intelligent supervised embedding (EMPrAvISE): detecting prostate cancer on multi-parametric MRI. In *Medical Imaging 2011: Computer-Aided Diagnosis*, volume 7963, page 79630U. International Society for Optics and Photonics, 2011.

Satish E Viswanath, Nicholas B Bloch, Jonathan C Chappelow, Robert Toth, Neil M Rofsky, Elizabeth M Genega, Robert E Lenkinski, and Anant Madabhushi. Central gland and peripheral zone prostate tumors have significantly different quantitative imaging signatures on 3 tesla endorectal, in vivo T2-weighted MR imagery. *Journal of Magnetic Resonance Imaging*, 36(1):213–224, 2012.

PC Vos, JO Barentsz, N Karssemeijer, and HJ Huisman. Automatic computer-aided de-

tection of prostate cancer based on multiparametric magnetic resonance image analysis. *Physics in Medicine & Biology*, 57(6):1527, 2012.

Steven Xiaogang Wang. *Maximum weighted likelihood estimation*. PhD thesis, University of British Columbia, 2001.

Yuan Wang, Brian P Hobbs, Jianhua Hu, Chaan S Ng, and Kim-Anh Do. Predictive classification of correlated targets with application to detection of metastatic cancer using functional CT imaging. *Biometrics*, 71(3):792–802, 2015.

Neil F Wasserman, Benjamin Spilseth, Jafar Golzarian, and Gregory J Metzger. Use of MRI for lobar classification of benign prostatic hyperplasia: potential phenotypic biomarkers for research on treatment strategies. *American Journal of Roentgenology*, 205(3):564–571, 2015.

Geoffrey S Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26:359–372, 1964.

Jeffrey C Weinreb, Jelle O Barentsz, Peter L Choyke, Francois Cornud, Masoom A Haider, Katarzyna J Macura, Daniel Margolis, Mitchell D Schnall, Faina Shtern, Clare M Tempany, et al. PI-RADS prostate imaging–reporting and data system: 2015, version 2. *European urology*, 69(1):16–40, 2016.

Weidong Zhou, Yinxia Liu, Qi Yuan, and Xueli Li. Epileptic seizure detection using lacunarity and Bayesian linear discriminant analysis in intracranial EEG. *IEEE Transactions on Biomedical Engineering*, 60(12):3375–3381, 2013.

Appendix A

Supplementary Materials for Chapter 2

A.1 Posterior Predictive Distribution for Mbase

For Mbase, the posterior distribution of the model parameter $(\boldsymbol{\mu_c}, \boldsymbol{\Sigma_c} | c \in \{0, 1\})$ is:

$$\begin{aligned}\boldsymbol{\mu_c}, \boldsymbol{\Sigma_c} | \boldsymbol{Y^c} \sim & \mathcal{MVN}(\boldsymbol{\mu_c} | \tilde{\boldsymbol{y}}^{\boldsymbol{c}}, \frac{\boldsymbol{\Sigma_c}}{\sum_{i=1}^{N} n_i^c}) \\ & \times \mathcal{W}^{-1}(\boldsymbol{\Sigma_c} | \sum_{i=1}^{N} n_i^c + \delta, \tilde{\boldsymbol{S}}^{\boldsymbol{c}} + \hat{\boldsymbol{\Omega}}_{\boldsymbol{c}}),\end{aligned} \tag{A.1}$$

where $c \in \{0, 1\}$, $\boldsymbol{Y^c}$ denotes the set of mpMRI parameters for the voxels that have cancer status c, n_i^c denotes the number of voxels in the i^{th} prostate slice that have cancer status c, and the two additional terms, $\tilde{\boldsymbol{y}}^{\boldsymbol{c}}$ and $\tilde{\boldsymbol{S}}^{\boldsymbol{c}}$, are defined as:

$$\begin{aligned}\tilde{\boldsymbol{y}}^{\boldsymbol{c}} &= \frac{\sum_{i=1}^{N} \sum_{j=1}^{n_i^c} \boldsymbol{y_{ij}^c}}{\sum_{i=1}^{N} n_i^c}, \\ \tilde{\boldsymbol{S}}^{\boldsymbol{c}} &= \sum_{i=1}^{N} \boldsymbol{Y_i^c} (\boldsymbol{Y_i^c})^T - \sum_{i=1}^{N} \sum_{j=1}^{n_i^c} \bar{\boldsymbol{y}}^{\boldsymbol{c}} \bar{\boldsymbol{y}}^{\boldsymbol{c}T}.\end{aligned} \tag{A.2}$$

The posterior predictive density of the mpMRI parameters for a new voxel given its

cancer status can be expressed as:

$$
\begin{aligned}
& f(\boldsymbol{y^*}|c^*, \boldsymbol{Y}, \boldsymbol{C}) \\
=& \int\int f(\boldsymbol{y^*}|c^*, \boldsymbol{\mu_{c^*}}, \boldsymbol{\Sigma_{c^*}}) f(\boldsymbol{\mu_{c^*}}, \boldsymbol{\Sigma_{c^*}}|\boldsymbol{Y}) d\boldsymbol{\mu_{c^*}} d\boldsymbol{\Sigma_{c^*}} \\
=& (2\pi)^{-\frac{m}{2}} \Big(\frac{\sum_{i=1}^N n_i^{c^*} + 1}{\sum_{i=1}^N n_i^{c^*}} \Big)^{-\frac{m}{2}} \frac{h(\sum_{i=1}^N n_i^{c^*} + \delta, \tilde{\boldsymbol{S}}^{\boldsymbol{c^*}} + \hat{\boldsymbol{\Omega}}_{\boldsymbol{c^*}})}{h(\sum_{i=1}^N n_i^{c^*} + \delta + 1, \tilde{\boldsymbol{S}}_*^{\boldsymbol{c^*}} + \hat{\boldsymbol{\Omega}}_{\boldsymbol{c}})},
\end{aligned} \tag{A.3}
$$

where $\tilde{\boldsymbol{S}}_*^{\boldsymbol{c^*}}$ is the updated value of $\tilde{\boldsymbol{S}}^{\boldsymbol{c}}$ in (A.2) with the new voxel included.

The posterior predictive cancer probability, which also serves as the voxel-wise classifier of cancer can then be written as:

$$
\begin{aligned}
& p(c^* = 1|\boldsymbol{y^*}, \boldsymbol{Y}, \boldsymbol{C}) \\
=& \frac{p(c^* = 1) f(\boldsymbol{y^*}|c^* = 1, \boldsymbol{Y}, \boldsymbol{C})}{\sum_{c=0}^1 p(c^* = c|r^*) f(\boldsymbol{y^*}|c^* = c, \boldsymbol{Y}, \boldsymbol{C})},
\end{aligned} \tag{A.4}
$$

where $p(c^* = 1)$ is the estimated prior cancer risk as the cancer prevalence across all slices and voxels in the data, and $f(\boldsymbol{y^*}|c^* = 1, \boldsymbol{Y}, \boldsymbol{C})$ follows (A.3).

A.2 Additional Simulation Results

A.2.1 Stationary Spatial Correlation Structure That Is Consistent across Slices

Table A.1 reports the average and 95% confidence intervals of the AUC for Mbase, Mregion, Mcoord and Msmooth, under different settings for the regional heterogeneity and spatial pattern assuming stationary spatial correlation structure. Results were obtained from 100 simulation per setting.

A.2.2 Stationary Spatial Correlation Structure, with ϕ Varying by Slice

When assuming that the spatial correlation structure varies by prostate slice, we set the spatial correlation range parameter, ϕ, to vary by slice, and simulated ϕ_i from a uniform distribution with support (l, u): $\phi_i \sim \text{Unif}(l, u)$. Different settings for (l, u), the spatial variance σ^2, and the magnitude of regional heterogeneity, were considered.

Figure A.1 presents the AUC of Mbase, Mregion, Mcoord and Msmooth, under different settings for the regional heterogeneity, spatial correlation range parameters, ϕ_i's, and spatial variance, σ^2. In the figure, the error bars represent the 95% confidence intervals for the AUC, and the dots at the center of the bars show the average AUC. Table A.2 reports details of the average and 95% confidence intervals for the AUC.

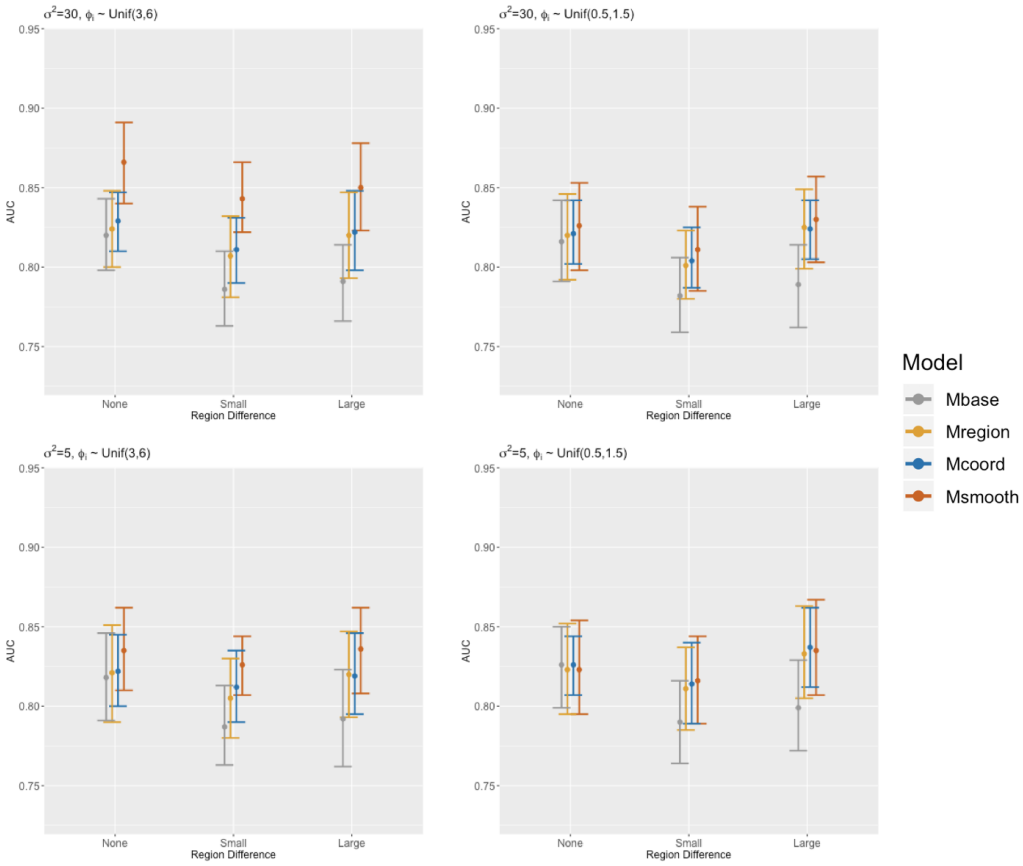

Figure A.1: Comparison of AUC between Mbase, Mregion, Mcoord and Msmooth, under different settings for the regional heterogeneity, spatial correlation range parameters, ϕ_i's, and spatial variance, σ^2, assuming that the spatial correlation range parameter varies by prostate slice.

Parameters			AUC			
Regional Difference	ϕ	σ^2	Mbase	Mregion	Mcoord	Msmooth
Large	5	30	.767 (.735, .800)	.823 (.786, .861)	.829 (.802, .857)	.898 (.855, .938)
		15	.783 (.757, .808)	.846 (.814, .877)	.851 (.823, .875)	.892 (.862, .921)
		5	.794 (.770, .816)	.850 (.823, .877)	.853 (.832, .874)	.881 (.853, .909)
	2	30	.769 (.750, .787)	.813 (.783, .843)	.814 (.800, .828)	.846 (.818, .874)
		15	.771 (.745, .798)	.821 (.796, .845)	.825 (.806, .843)	.858 (.832, .884)
		5	.782 (.763, .799)	.833 (.809, .856)	.830 (.808, .851)	.851 (.826, .874)
	1	30	.797 (.778, .816)	.841 (.823, .863)	.839 (.817, .858)	.840 (.819, .860)
		15	.797 (.777, .818)	.843 (.822, .866)	.847 (.827, .868)	.843 (.818, .864)
		5	.795 (.772, .815)	.849 (.826, .874)	.854 (.836, .872)	.849 (.828, .873)
Small	5	30	.785 (.752, .818)	.817 (.778, .856)	.822 (.797, .848)	.891 (.858, .925)
		15	.779 (.751, .807)	.819 (.786, .849)	.825 (.798, .853)	.886 (.857, .915)
		5	.784 (.760, .808)	.823 (.795, .851)	.829 (.809, .850)	.852 (.820, .884)
	2	30	.770 (.751, .791)	.806 (.789, .822)	.810 (.797, .823)	.849 (.825, .873)
		15	.779 (.750, .808)	.803 (.777, .827)	.808 (.788, .830)	.840 (.816, .864)
		5	.783 (.762, .801)	.813 (.791, .835)	.817 (.800, .834)	.833 (.807, .857)
	1	30	.774 (.751, .795)	.811 (.787, .834)	.813 (.793, .831)	.810 (.784, .837)
		15	.775 (.751, .797)	.816 (.789, .840)	.820 (.798, .841)	.816 (.788, .845)
		5	.773 (.752, .792)	.823 (.798, .848)	.826 (.806, .844)	.823 (.801, .843)
None	5	30	.820 (.791, .849)	.817 (.779, .857)	.829 (.797, .860)	.898 (.863, .935)
		15	.821 (.792, .850)	.816 (.780, .852)	.829 (.804, .854)	.890 (.858, .920)
		5	.829 (.808, .850)	.826 (.796, .856)	.835 (.815, .855)	.859 (.830, .887)
	2	30	.820 (.803, .837)	.818 (.795, .841)	.830 (.817, .841)	.859 (.837, .881)
		15	.819 (.791, .849)	.817 (.794, .838)	.825 (.807, .844)	.855 (.830, .880)
		5	.826 (.805, .848)	.824 (.797, .849)	.830 (.812, .847)	.844 (.821, .864)
	1	30	.823 (.802, .844)	.820 (.795, .845)	.818 (.803, .843)	.823 (.797, .850)
		15	.820 (.790, .848)	.825 (.802, .845)	.830 (.813, .848)	.826 (.798, .857)
		5	.822 (.798, .845)	.827 (.804, .854)	.831 (.808, .855)	.830 (.807, .856)

Table A.1: Comparison of AUC between Mbase, Mregion, Mcoord and Msmooth, under different settings for the regional heterogeneity, spatial correlation range parameter, ϕ, and spatial variance, σ^2, assuming stationary spatial correlation structure.

Parameters			AUC			
Regional Difference	ϕ_i's	σ^2	Mbase	Mregion	Mcoord	Msmooth
Large	Unif(3,6)	30	.791 (.766, .814)	.820 (.793, .847)	.822 (.798, .848)	.850 (.823, .878)
		5	.792 (.762, .823)	.820 (.793, .847)	.819 (.795, .846)	.836 (.808, .862)
	Unif(0.5,1.5)	30	.789 (.762, .814)	.825 (.799, .849)	.824 (.805, .842)	.830 (.803, .857)
		5	.799 (.772, .829)	.833 (.805, .863)	.837 (.812, .862)	.835 (.807, .867)
Small	Unif(3,6)	30	.786 (.763, .810)	.807 (.781, .832)	.811 (.790, .831)	.843 (.822, .866)
		5	.787 (.763, .813)	.805 (.780, .830)	.812 (.790, .835)	.826 (.807, .844)
	Unif(0.5,1.5)	30	.782 (.759, .806)	.801 (.780, .823)	.804 (.787, .825)	.811 (.785, .838)
		5	.790 (.764, 816)	.811 (.785, .837)	.814 (.789, .840)	.816 (.789, .844)
None	Unif(3,6)	30	.820 (.798, .843)	.824 (.800, .848)	.829 (.810, .847)	.866 (.840, .891)
		5	.818 (.791, .846)	.821 (.790, .851)	.822 (.800, .845)	.835 (.810, .862)
	Unif(0.5,1.5)	30	.816 (.791, .842)	.820 (.792, .846)	.821 (.802, .842)	.826 (.798, .853)
		5	.826 (.799, .850)	.823 (.795, .852)	.826 (.807, .844)	.823 (.795, .854)

Note:Unif(l, u) means that the spatial correlation range parameters, ϕ's, were generated from a uniform distribution with support (l, u).

Table A.2: Comparison of AUC between Mbase, Mregion, Mcoord and Msmooth, under different settings for the regional heterogeneity, spatial correlation range parameters, ϕ_i's, and spatial variance σ^2, assuming that the spatial correlation range parameter varies by prostate slice.

A.2.3 Non-stationary Spatial Correlation Structure

For the scenario where there is potential local non-stationarity in the spatial pattern within a prostate slice, we assume that the spatial correlation range parameter varies by the location in a prostate slice. Specifically, we assume that $\phi(\boldsymbol{s_{ij}}) \sim \text{Unif}(l, u)$, and that the range parameter for the spatial correlation between the j^{th} and the k^{th} voxel in the i^{th} slice is $\frac{1}{2}(\phi(\boldsymbol{s_{ij}}) + \phi(\boldsymbol{s_{ik}}))$, where $\boldsymbol{s_{ij}}$ refers to the standardized 2-D coordinate of the j^{th} voxel in the i^{th} slice, as introduced in Section 2.2. In this way, we can construct the spatial covariance matrix, $\boldsymbol{C_i}(\phi, \sigma^2)$, with the (j, k)-th entry being:

$$\boldsymbol{C_i}(\phi, \sigma^2)_{j,k} = \sigma^2 e^{-\frac{1}{2}(\phi(\boldsymbol{s_{ij}})+\phi(\boldsymbol{s_{ij}}))||\boldsymbol{s_{ij}}-\boldsymbol{s_{ik}}||_2}. \tag{A.5}$$

We evaluate the model performance under different settings for the regional heterogeneity, the support, (l, u), for the uniform distribution used to generate $\phi(\boldsymbol{s_{ij}})$'s, and the spatial variance, σ^2. Figure 2 presents the AUC for the Mbase, Mregion, Mcoord and Msmooth models as a function of the magnitude of regional heterogeneity under different patterns of spatial variance σ^2 and correlation ϕ, assuming that there is local non-stationarity in the spatial pattern in each slice. In the figure, the error bars represent the 95% confidence intervals for the AUC, and the dots at the center of the bars show the average AUC. Table 3 shows the details of the estimated AUC and their corresponding 95% confidence intervals.

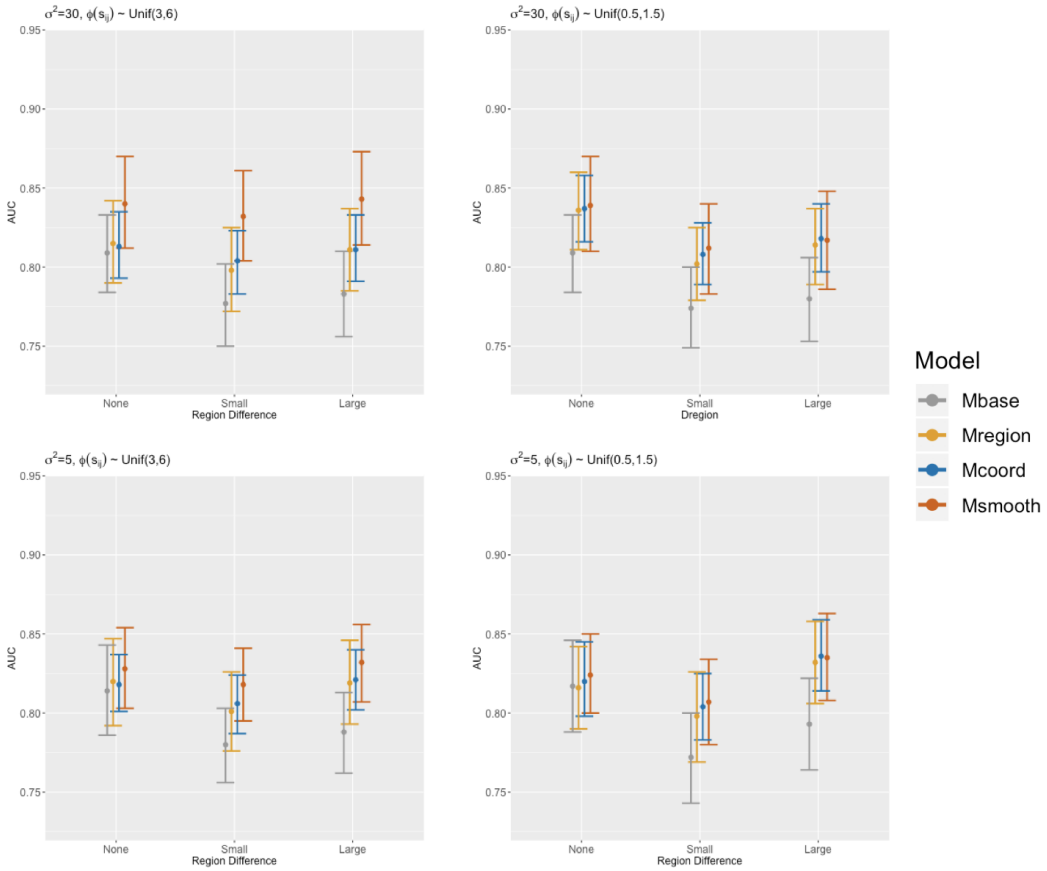

Figure A.2: Comparison of AUC between Mbase, Mregion, Mcoord and Msmooth, under different settings for the regional heterogeneity, spatial correlation range parameters, $\phi(\boldsymbol{s_{ij}})$'s, and spatial variance, σ^2, assuming that there is local non-stationary spatial pattern within slices.

Parameters			AUC			
Regional Difference	$\phi(\boldsymbol{s_{ij}})$'s	σ^2	Mbase	Mregion	Mcoord	Msmooth
Large	Unif(3,6)	30	.783 (.756, .810)	.811 (.785, .837)	.811 (.791, .833)	.843 (.814, .873)
		5	.788 (.762, .813)	.819 (.793, .846)	.821 (.802, .840)	.832 (.807, .856)
	Unif(0.5,1.5)	30	.780 (.753, .806)	.814 (.789, .837)	.818 (.797, .840)	.817 (.786, .848)
		5	.788 (.762, .813)	.819 (.793, .846)	.821 (.802, .840)	.832 (.807, .856)
Small	Unif(3,6)	30	.777 (.750, .802)	.798 (.772, .825)	.804 (.783, .823)	.832 (.804, .861)
		5	.780 (.756, .803)	.801 (.776, .826)	.806 (.787, .824)	.818 (.795, .841)
	Unif(0.5,1.5)	30	.774 (.749, .800)	.802 (.779, .825)	.808 (.789, .828)	.812 (.783, .840)
		5	.780 (.756, 803)	.801 (.776, .826)	.806 (.787, .824)	.818 (.795, .841)
None	Unif(3,6)	30	.809 (.784, .833)	.815 (.790, .842)	.813 (.793, .835)	.840 (.812, .870)
		5	.814 (.786, .843)	.820 (.792, .847)	.818 (.801, .837)	.828 (.803, .854)
	Unif(0.5,1.5)	30	.809 (.784, .833)	.836 (.811, .860)	.837 (.816, .858)	.839 (.810, .870)
		5	.814 (.786, .843)	.820 (.792, .847)	.818 (.801, .837)	.828 (.803, .854)

Note: Unif(l, u) means that the spatial correlation range parameters, $\phi(\boldsymbol{s_{ij}})$'s, were generated from a Uniform distribution with support, (l, u), for each location, $(\boldsymbol{s_{ij}})$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, n_i$.

Table A.3: Comparison of AUC between Mbase, Mregion, Mcoord and Msmooth, under different settings for the regional heterogeneity, spatial correlation range parameters, $\phi(\boldsymbol{s_{ij}})$'s, and spatial variance, σ^2, assuming that there is local non-stationary spatial pattern within slices.

Appendix B

Supplementary Materials for Chapter 3

B.1 MCMC Algorithm for Bayesian Inference and Classification

Bayesian inference and classification of the various models were implemented using MCMC algorithms via Gibbs samplers with a Metropolis-Hastings sampling step. Model parameters are estimated using data from all images, and the voxel-wise cancer status indicators for a new patient, $\{c_{N+1,j}|j = 1, \ldots, n_{N+1}\}$, are classified by treating them as unknown parameters and updating them along with the other model parameters. Derivation of the MCMC algorithm is based on the joint density:

$$\prod_{i=1}^{N+1}\prod_{j=1}^{n_i}\Big\{\mathcal{MVN}(\boldsymbol{y_{ij}}|\boldsymbol{\mu_{c_{ij},r_{ij}}}+\boldsymbol{\delta_i},\boldsymbol{\Gamma_{c_{ij},r_{ij}}})P(\kappa_{ij}>0)^{c_{ij}}P(\kappa_{ij}\leqslant 0)^{1-c_{ij}}\Big\}\times$$
$$\prod_{i=1}^{N+1}\mathcal{MVN}(\boldsymbol{\delta_i}|\mathbf{0},\boldsymbol{\Sigma})\times\prod_{i=1}^{N+1}\prod_{j=1}^{n_i}N(\kappa_{ij}|q_{r_{ij},0}+w_{ij},1)\times\prod_{i=1}^{N+1}\mathcal{MVN}(\boldsymbol{w_i}|\mathbf{0},\boldsymbol{C_i})$$
$$\times\prod_{c=0}^{1}\prod_{r=0}^{1}\mathcal{W}^{-1}(\boldsymbol{\Gamma_{c,r}}|d-1,\boldsymbol{I_d})\times\mathcal{W}^{-1}(\boldsymbol{\Sigma}|d-1,\boldsymbol{I_d})\times\pi(\boldsymbol{\theta}), \tag{B.1}$$

where $\pi(\boldsymbol{\theta})$ denotes the prior distribution of the spatial parameters $\boldsymbol{\theta}$, and $\boldsymbol{C_i}$ is the general notation of the spatial covariance matrix of $\boldsymbol{w_i}$ in different models, where which is replaced

by $\boldsymbol{C}_{\boldsymbol{S}_i}$ in the full model, $\tilde{\boldsymbol{C}}_{\boldsymbol{S}_i}$ in the NNGP model, $\tilde{\tilde{\boldsymbol{C}}}_{\boldsymbol{S}_i}$ in the reduced-rank model, and $(\boldsymbol{I}-\boldsymbol{B}_i)^{-1}\boldsymbol{D}_i$ in the CAR model, with $\boldsymbol{B}_i$ and $\boldsymbol{D}_i$ defined in Section 3.2.3.

The MCMC algorithm can be split into two parts: one for the update of parameters in the non-spatial model components, and the other for the update of the spatial parameters. We will discuss the two parts of MCMC separately in the following sub-sections.

B.1.1 MCMC Algorithms for the Update of Non-spatial Model Parameters

We first discuss the MCMC algorithm for the update of the non-spatial model parameters, which holds for all models in Section 3.2.2 and Section 3.2.3. Given a conjugate Inverse-Gamma prior $\mathcal{W}^{-1}(d+1, \boldsymbol{I_d})$, the full conditional distribution for $\boldsymbol{\Gamma_0}$ becomes $\mathcal{W}^{-1}(\sum_{i=1}^{N+1} n_i^0 + d + 1, \tilde{\boldsymbol{S}}^{\boldsymbol{0}} + \boldsymbol{I_d})$, where

$$\tilde{\boldsymbol{S}}^{\boldsymbol{0}} = \sum_{i=1}^{N+1}\sum_{j=1}^{n_i^0}\big[\boldsymbol{y_{ij}^0} - (\boldsymbol{\mu_0}+\boldsymbol{\delta_i})\big]\big[\boldsymbol{y_{ij}^0} - (\boldsymbol{\mu_0}+\boldsymbol{\delta_i})\big]^T, \tag{B.2}$$

and $\boldsymbol{y_{ij}^0}$ denotes the mpMRI parameters of the noncancer voxels, with n_i^c denoting the number of voxels in image i that has cancer status $c \in \{0,1\}$. Similar results can be derived for the update of $\boldsymbol{\Gamma_1}$. Given a flat prior, $\boldsymbol{\mu_0}$ has a multivariate normal full conditional distribution $\mathcal{MVN}(\frac{\sum_{i=1}^{N+1}\sum_{n=1}^{n_i}(\boldsymbol{y_{ij}^0}-\boldsymbol{\delta_i})}{\sum_{i=1}^{N+1} n_i^0}, \frac{\boldsymbol{\Gamma_0}}{\sum_{i=1}^{N+1} n_i^0})$, and similar results can be derived for the update of $\boldsymbol{\mu_1}$. With a conjugate prior $\mathcal{MVN}(\boldsymbol{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\delta_i}$ has a full conditional distribution $\mathcal{MVN}(\boldsymbol{\delta_i}|\boldsymbol{\mu_{\delta_i}}, \boldsymbol{V_{\delta_i}})$, where

$$\begin{aligned}
\boldsymbol{V_{\delta_i}} &= \big[\sum_{j=1}^{n_i}\boldsymbol{\Gamma}_{\boldsymbol{c_{ij}}}^{-1} + \boldsymbol{\Sigma^{-1}}\big]^{-1},\\
\boldsymbol{\mu_{\delta_i}} &= \boldsymbol{V_{\delta_i}}\big[\sum_{j=1}^{n_i}\Gamma_{c_{ij}}^{-1}\big(\boldsymbol{y_{ij}} - \boldsymbol{\mu_{c_{ij}}}\big)\big].
\end{aligned} \tag{B.3}$$

Assuming a prior $\mathcal{W}^{-1}(\boldsymbol{\Sigma}|d+1, \boldsymbol{I_d})$, $\boldsymbol{\Sigma}$ has a full conditional distribution $\mathcal{W}^{-1}(\boldsymbol{\Sigma}|N + d + 2, \sum_{i=1}^{N+1} \boldsymbol{\delta_i}\boldsymbol{\delta_i^T} + \boldsymbol{I_d})$. We also set $q_{r,0} = \Phi^{-1}(\sum_{i=1}^{N}\sum_{j:r_{ij}=r} c_{ij}/\sum_{i=1}^{N}\sum_{j:r_{ij}=r} 1)$, $r \in \{0, 1\}$.

B.1.2 MCMC Algorithm for the Update of Spatial Model Parameters and Unknown Voxel-wise Cancer Statuses

Full Spatial Model

For model (3.2) that has full spatial structure, $\boldsymbol{\theta}$ is updated using a random walk Metropolis step with target density $\pi(\boldsymbol{\theta}) \times \prod_{i=1}^{N+1} \mathcal{MVN}(\boldsymbol{w_i}|\mathbf{0}, \boldsymbol{C_{S_i}}(\boldsymbol{\theta}))$, where $\pi(\boldsymbol{\theta})$ is the prior of $\boldsymbol{\theta}$. In this dissertation, we assume a Matérn stationary correlation function, and $\boldsymbol{\theta} = \{\sigma^2, \phi, \nu\}$, where σ^2 denotes the stationary spatial variance, ϕ denotes the spatial range parameter, and ν denotes the smoothness of spatial correlation. We assume a non-informative prior on $\boldsymbol{\theta}$, with a IG$(1, 1)$ prior on σ^2, and a flat prior on both ϕ and ν. The update of κ_{ij} holds for all spatial models: for the j^{th} voxel in the i^{th} image, if the previous draw of c_{ij} is 0, then the full conditional distribution of κ_{ij} is a truncated normal $TrN(\kappa_{ij}|q_{r_{ij},0} + w_{ij}, 1, (-\infty, 0))$, and if the previous draw of c_{ij} is 1, then the full conditional of κ_{ij} is $TrN(\kappa_{ij}|q_{r_{ij},0} + w_{ij}, 1, (0, \infty))$, where if a random variable follows $TrN(a, b, (c, d))$, it means that it follows $N(a, b)$ but only lies within the interval (c, d), $a, b, c, d \in \mathbb{R}$. The update of w_{ij} proceeds as follows. First, notice that $w_{ij}|\boldsymbol{w_{i,-j}} \sim N(w_{ij}|\mu_{w_{ij}}, \sigma^2_{w_{ij}})$, where

$$\begin{aligned}
\mu_{w_{ij}} &= \boldsymbol{C_{S_i(j,-j)}}\boldsymbol{C^{-1}_{S_i(-j,-j)}}\boldsymbol{w_{i,-j}}, \\
\sigma^2_{w_{ij}} &= C_{\boldsymbol{S}_i(j,j)} - \boldsymbol{C_{S_i(j,-j)}}\boldsymbol{C^{-1}_{S_i(-j,-j)}}\boldsymbol{C_{S_i(-j,j)}},
\end{aligned} \tag{B.4}$$

with $\boldsymbol{C_{S_i(j,-j)}}$ denoting the j^{th} row of $\boldsymbol{C_{S_i}}$ after deleting its j^{th} column, $\boldsymbol{C_{S_i(-j,-j)}}$ denoting $\boldsymbol{C_{S_i}}$ after deleting its j^{th} row and column, and $C_{\boldsymbol{S}_i(j,j)}$ is the (j, j)-th entry of $\boldsymbol{C_{S_i}}$. The

full conditional distribution of w_{ij} is then:

$$N(w_{ij}|(1+\sigma_{w_{ij}}^{-2})^{-1}(\sigma_{w_{ij}}^{-2}\mu_{w_{ij}}+\kappa_{ij}-q_{r_{ij},0}),(1+\sigma_{w_{ij}}^{-2})^{-1}). \tag{B.5}$$

Finally, we update $\{c_{N+1,j}|j=1,...,n_{N+1}\}$. The full conditional probability of $c_{N+1,j}=1$ can be expressed as:

$$\frac{P(\kappa_{N+1,j}>0)f(\boldsymbol{y_{N+1,j}}|\boldsymbol{\mu_{1,r_{N+1,j}}}+\boldsymbol{\delta_{N+1}},\boldsymbol{\Gamma_{1,r_{N+1,j}}})}{\sum_{c=0}^{1}P(\kappa_{N+1,j}>0)^{c}P(\kappa_{N+1,j}\leq 0)^{1-c}f(\boldsymbol{y_{N+1,j}}|\boldsymbol{\mu_{c,r_{N+1,j}}}+\boldsymbol{\delta_{N+1}},\boldsymbol{\Gamma_{c,r_{N+1,j}}})}, \tag{B.6}$$

where $P(\kappa_{N+1,j}\leq 0)$ is the cdf of $N(q_{r_{ij},0}+w_{ij},1)$ at 0, and $f(\boldsymbol{y_{N+1,j}}|\boldsymbol{\mu_{c,r_{N+1,j}}}+\boldsymbol{\delta_{N+1}},\boldsymbol{\Gamma_{c,r_{N+1,j}}})$ is the pdf of $\mathcal{MVN}(\boldsymbol{\mu_{c,r_{N+1,j}}}+\boldsymbol{\delta_{N+1}},\boldsymbol{\Gamma_{c,r_{N+1,j}}})$, $c\in\{0,1\}$. This step holds for all models.

NNGP Model

The update of $\boldsymbol{\theta}$ still follows a random walk Metropolis step, but with a different target density $\pi(\boldsymbol{\theta})\times\prod_{i=1}^{N+1}\prod_{j=1}^{n_i}N(w_{s_{ij}}|B_{\boldsymbol{s_{ij}}}\boldsymbol{w_{N_{ij}}},F_{\boldsymbol{s_{ij}}})$, where $B_{\boldsymbol{s_{ij}}}$ and $F_{\boldsymbol{s_{ij}}}$ were previously defined in Section 3.2.3. The full conditional of w_{ij} is $N(w_{ij}|M_{ij},G_{ij})$, where M_{ij} and G_{ij} are defined as follows. Let $\boldsymbol{s}$ and $\boldsymbol{t}$ be any two locations within an image, if $\boldsymbol{s}\in N(\boldsymbol{t})$ and $\boldsymbol{s}$ is the l^{th} element of $N(\boldsymbol{t})$, then we denote $\boldsymbol{s}=N(\boldsymbol{t})(l)$. Let $B_{\boldsymbol{t},\boldsymbol{s}}$ be the l^{th} element of $\boldsymbol{B_t}$. Let $U(\boldsymbol{s_{ij}})=\{\boldsymbol{t}|\boldsymbol{s_{ij}}\in N(\boldsymbol{t})\}$ and for every $\boldsymbol{t}\in U(\boldsymbol{s_{ij}})$ we define $a_{\boldsymbol{t},\boldsymbol{s_{ij}}}=w_{\boldsymbol{t}}-\sum_{\boldsymbol{ss}\in N(\boldsymbol{t}),ss\neq\boldsymbol{s_{ij}}}B_{\boldsymbol{t},\boldsymbol{ss}}w_{\boldsymbol{ss}}$. Then $w_{ij}|N(w_{ij})\sim N(w_{ij}|M_{ij},G_{ij})$, where

$$\begin{aligned}
G_{ij}&=\big[F_{ij}^{-1}+\sum_{t\in U(\boldsymbol{s_{ij}})}\frac{B_{t,s_{ij}}^2}{F_t}\big]^{-1},\\
M_{ij}&=G_{ij}\cdot\Big\{F_{ij}^{-1}\boldsymbol{B_{s_{ij}}}\boldsymbol{w_{N(s_{ij})}}+\sum_{t\in U(s_{ij})}B_{t,s_{ij}}F_t^{-1}a_{t,s_{ij}}\Big\}.
\end{aligned} \tag{B.7}$$

Knot-bsed Reduced-rank Model

For the update of spatial parameters in the knot-based reduced-rank model, the MCMC algorithm becomes computationally intensive if we update $\{w_{ij}|i = 1, 2, \ldots, N+1, j = 1, 2, \ldots, n_i\}$. Instead, we marginalize over $\boldsymbol{w_{ij}}$'s and rewrite the joint density as:

$$\prod_{i=1}^{N+1}\prod_{j=1}^{n_i}\Big\{\mathcal{MVN}(\boldsymbol{y_{ij}}|\boldsymbol{\mu_{c_{ij},r_{ij}}}+\boldsymbol{\delta_i},\boldsymbol{\Gamma_{c_{ij},r_{ij}}})\,P(\kappa_{ij}>0)^{c_{ij}}P(\kappa_{ij}\leqslant 0)^{1-c_{ij}}\Big\}\times$$
$$\prod_{i=1}^{N+1}\mathcal{MVN}(\boldsymbol{\delta_i}|\mathbf{0},\boldsymbol{\Sigma})\times\prod_{i=1}^{N+1}\prod_{j=1}^{n_i}N(\kappa_{ij}|q_{r_{ij},0}+\boldsymbol{C_{s_{ij},S_i^*}}\boldsymbol{C_{S_i^*}^{-1}}\boldsymbol{w_i^*},1)\times\prod_{i=1}^{N+1}\mathcal{MVN}(\boldsymbol{w_i^*}|\mathbf{0},\boldsymbol{C_{S_i^*}})$$
$$\times\prod_{c=0}^{1}\prod_{r=0}^{1}\mathcal{W}^{-1}(\boldsymbol{\Gamma_{c,r}}|d-1,\boldsymbol{I_d})\times\mathcal{W}^{-1}(\boldsymbol{\Sigma}|d-1,\boldsymbol{I_d})\times\pi(\boldsymbol{\theta}). \quad \text{(B.8)}$$

Under this joint density, we only need to update $\{\boldsymbol{w_i^*}|i = 1, 2, \ldots, N+1\}$.

Specifically, for the update of $\boldsymbol{\theta}$, we use a random walk Metropolis step with a target density $\pi(\boldsymbol{\theta})\times\prod_{i=1}^{N+1}\prod_{j=1}^{n_i}N(\kappa_{ij}|q_{r_{ij},0}+\boldsymbol{C_{s_{ij},S_i^*}}\boldsymbol{C_{S_i^*}^{-1}}\boldsymbol{w_i^*},1)\times\prod_{i=1}^{N+1}\mathcal{MVN}(\boldsymbol{w_i^*}|\mathbf{0},\boldsymbol{C_{S_i^*}})$. For the update of $\boldsymbol{w_i^*}$, we use Gibbs sampling with the full conditional distribution $\boldsymbol{w_i^*}|\cdot\sim\mathcal{MVN}(\boldsymbol{\mu_{w_i^*}},\boldsymbol{V_{w_i^*}})$, where

$$\boldsymbol{V_{w_i^*}}=\Big[\sum_{j=1}^{n_i}\boldsymbol{C_{S_i^*}^{-1}}\boldsymbol{C_{s_{ij},S_i^*}^{T}}\boldsymbol{C_{s_{ij},S_i^*}}\boldsymbol{C_{S_i^*}^{-1}}+\boldsymbol{C_{S_i}^{*-1}}\Big]^{-1},$$
$$\boldsymbol{\mu_{w_i^*}}=\boldsymbol{V_{w_i^*}}\Big[\sum_{i=1}^{n_i}\boldsymbol{C_{S_i^*}^{-1}}\boldsymbol{C_{s_{ij},S_i^*}^{T}}\big(k_{ij}-q_{r_{ij},0}\big)\Big]. \quad \text{(B.9)}$$

CAR Model

CAR model does not assume any spatial correlation parameter, so there is no $\boldsymbol{\theta}$. For the update of σ^2, the variance of $w_{ij}|\boldsymbol{w_{i,-j}}$, we assume a conjugate $IG(a_{\sigma^2},b_{\sigma^2})$ prior, and the full conditional for σ^2 becomes $IG(a_{\sigma^2}+\frac{\sum_{i=1}^{N+1}n_i}{2},b_{\sigma^2}+\frac{1}{2}\sum_{i=1}^{N+1}\boldsymbol{w_i^T}(\boldsymbol{I}-\boldsymbol{B_i})\boldsymbol{w_i})$.The full conditional for $\boldsymbol{w_i}$ is $N(w_{ij}|\sum_{k\neq j}b_{ijk}w_{ik},\sigma^2)$, where $\boldsymbol{B_i}=[b_{ijk}]_{j,k=1}^{n_i}$ is fixed.

Appendix C

Supplementary Materials for Chapter 4

C.1 Additional Simulation Settings

As described in Section 4.3.1, the simulated data sets were generated according to model (C.1):

$$
\begin{aligned}
\boldsymbol{w_i} &\sim \mathcal{MVN}(\mathbf{0}, \boldsymbol{C}(\boldsymbol{S_i}, \boldsymbol{S_i}|\boldsymbol{\theta})),\\
G^*_{ij} &\sim N(q_{r_{ij},0} + w_{ij}, 1),\\
G_{ij} &= I(G^*_{ij} > 0),\\
\boldsymbol{e_{a_{k^2}}} &\sim \mathcal{MVN}(\mathbf{0}, \boldsymbol{\Lambda}),\\
\boldsymbol{\delta_i} &\sim \mathcal{MVN}(\mathbf{0}, \boldsymbol{\Sigma}),\\
\boldsymbol{y_{ij}} &\overset{ind}{\sim} \mathcal{MVN}(\boldsymbol{\mu_{c_{ij},r_{ij}}} + \sum_{k=1}^{K} \boldsymbol{e_{a^k_{ij}}} + \boldsymbol{\delta_i}, \boldsymbol{\Gamma_{c_{ij},r_{ij}}}).
\end{aligned}
\tag{C.1}
$$

The general model parameters, including the means, $\{\boldsymbol{\mu_{c,r}}, c = 0, 1, r = 0, 1\}$, within-patient covariance, $\{\boldsymbol{\Gamma_{c_{ij},r_{ij}}}, c = 0, 1, r = 0, 1\}$, and between-patient covariance, $\boldsymbol{\Sigma}$, of the mpMRI parameters, and the probit of the cancer prevalence in the PZ and CG, $q_{r_{ij},0}$, were set based on the estimates from the real data. $\boldsymbol{\mu_{c_{ij},r_{ij}}}$, $c = 0, 1$, $r = 0, 1$, was set to equal to the estimates from the motivating data set. $\boldsymbol{\Gamma_{c_{ij},r_{ij}}}$, $c = 0, 1$, $r = 0, 1$, was set to equal to

1/1.5 times the estimates from the motivating data set. $\boldsymbol{\Sigma}$ was set to equal to a $d \times d$ diagonal matrix (d=4), with the diagonal entries equal to 4, 0.24, 0.24, and 0.24, respectively. This was set based on the estimated between-patient covariance from the motivating data set, but was simplified as a diagonal matrix, and was modified so that the variances of the mpMRI parameters have a magnitude approximately 1/1.5 of that of the motivating data set. To simulate the regional heterogeneity of different magnitude, we varied $q_{r,0}$, $r = 0, 1$ and $\boldsymbol{\Lambda}$. To simulate different spatial correlation structures, we varied $\boldsymbol{\theta} = (\sigma^2, \phi, \nu)$. Details are described in Section 4.3.1.

C.2 Additional Classification Results on the Ordinal Cancer Outcome

C.2.1 Simulation Studies

Table C.1 and Table C.2 report simulation results of models SL0 + w_1, SL + w_1, SL0 + w_2, SL + w_2, assuming that QDA is the base learner. Table C.3 and Table C.4 report simulation results assuming that RF is the base learner. Table C.5 and Table C.6 report simulation results, assuming that the multi-resolution GLM, QDA and RF are combined in the super learner.

Table C.1: Simulation results assuming moderate regional heterogeneity, and QDA as the base learner.

Regional Heterogeneity	Spatial Correlation	Method	True Category	Classification Results					
				Classification Table			FPR	FDR	Overall Error Rate
				1	2	3			
			1	57476	1477	10147	0.17	0.38	
		Baseline	2	16310	1951	9378	0.93	0.60	0.42
			3	19742	1384	20333	0.51	0.49	
				1	2	3			
			1	59057	0	10042	0.15	0.38	
		SL0 + w_1	2	16927	0	10712	1.00	NA	0.42
			3	19789	0	21671	0.48	0.49	
				1	2	3			
	Moderate		1	59990	0	9109	0.13	0.36	
	$\sigma^2 = 10$,	SL + w_1	2	17549	0	10090	1.00	NA	0.39
	$\phi = 2/3$,		3	16927	0	24532	0.41	0.44	
	$\nu = 1$.			1	2	3			
			1	45622	11931	11546	0.34	0.30	
		SL0 + w_2	2	9496	6394	11749	0.77	0.75	0.45
			3	10305	7572	23583	0.43	0.50	
				1	2	3			
			1	44792	15159	9149	0.35	0.27	
		SL + w_2	2	9626	7771	10242	0.72	0.76	0.44
			3	6593	9839	25027	0.40	0.43	
				1	2	3			
Moderate			1	56648	1926	10461	0.18	0.38	
		Baseline	2	16010	2193	9410	0.92	0.64	0.43
			3	19395	1869	20156	0.51	0.50	
				1	2	3			
			1	58196	0	10119	0.15	0.39	
		SL0 + w_1	2	17097	0	10517	1.00	NA	0.42
			3	20657	0	20764	0.50	0.50	
				1	2	3			
	Strong		1	60700	200	8136	0.12	0.34	
	$\sigma^2 = 20$,	SL + w_1	2	16632	97	10085	1.00	NA	0.37
	$\phi = 2$,		3	14828	166	26427	0.36	0.43	
	$\nu = 1.5$.			1	2	3			
			1	45557	11461	12017	0.34	0.30	
		SL0 + w_2	2	9345	6149	12119	0.78	0.76	0.45
			3	10172	7309	23939	0.42	0.50	
				1	2	3			
			1	46323	15485	7227	0.33	0.23	
		SL + w_2	2	8621	8814	10179	0.68	0.75	0.41
			3	5209	10103	26108	0.37	0.40	

Table C.2: Simulation results assuming strong regional heterogeneity, and QDA as the base learner.

Regional Heterogeneity	Spatial Correlation	Method	True Category	Classification Results					
				Classification Table			FPR	FDR	Overall Error Rate
				1	2	3			
			1	57809	1092	10467	0.17	0.40	
		Baseline	2	16937	1449	9360	0.95	0.61	0.43
			3	21294	1050	19276	0.54	0.51	
				1	2	3			
			1	59291	0	10077	0.15	0.38	
		SL0 + w_1	2	16752	0	10995	1.00	NA	0.42
			3	19867	0	21754	0.48	0.49	
				1	2	3			
	Moderate		1	59919	0	9449	0.14	0.36	
	$\sigma^2 = 10$,	SL + w_1	2	17310	0	10437	1.00	NA	0.39
	$\phi = 2/3$,		3	16567	0	25054	0.40	0.44	
	$\nu = 1$.			1	2	3			
			1	45869	11834	11665	0.34	0.30	
		SL0 + w_2	2	9315	6338	12094	0.77	0.75	0.45
			3	9850	7819	23953	0.43	0.50	
				1	2	3			
			1	44805	15193	9370	0.35	0.26	
		SL + w_2	2	9431	7766	10549	0.72	0.76	0.44
			3	6196	9874	25551	0.39	0.44	
				1	2	3			
Strong			1	56830	1598	10792	0.18	0.40	
		Baseline	2	16905	1738	9044	0.94	0.66	0.44
			3	21453	1560	18519	0.55	0.52	
				1	2	3			
			1	58941	0	10279	0.15	0.39	
		SL0 + w_1	2	17202	0	10485	1.00	NA	0.43
			3	21251	0	20281	0.51	0.51	
				1	2	3			
	Strong		1	60620	78	8523	0.12	0.35	
	$\sigma^2 = 20$,	SL + w_1	2	16746	49	10893	1.00	NA	0.38
	$\phi = 2$,		3	15974	69	25488	0.39	0.44	
	$\nu = 1.5$.			1	2	3			
			1	45097	11363	12760	0.35	0.31	
		SL0 + w_2	2	8411	5842	12435	0.79	0.76	0.46
			3	10537	7086	23909	0.42	0.51	
				1	2	3			
			1	45644	15480	8096	0.34	0.24	
		SL + w_2	2	8770	8268	10650	0.70	0.76	0.42
			3	5705	9895	25932	0.38	0.42	

Table C.3: Simulation results assuming moderate regional heterogeneity, and RF as the base learner.

Regional Heterogeneity	Spatial Correlation	Method	True Category	Classification Results					
				Classification Table			FPR	FDR	Overall Error Rate
				1	2	3			
			1	53579	4562	10958	0.22	0.39	
		Baseline	2	15018	3937	8684	0.86	0.69	0.45
			3	18628	4253	18579	0.55	0.51	
				1	2	3			
			1	58748	0	10351	0.15	0.39	
		SL0 + w_1	2	17381	0	10258	1.00	NA	0.43
			3	20857	0	20602	0.50	0.50	
				1	2	3			
	Moderate		1	59738	0	9362	0.14	0.37	
	$\sigma^2 = 10$,	SL + w_1	2	17588	0	10051	1.00	NA	0.39
	$\phi = 2/3$,		3	17239	0	24221	0.42	0.45	
	$\nu = 1$.			1	2	3			
			1	44923	11129	13048	0.35	0.32	
		SL0 + w_2	2	9890	5693	12056	0.79	0.76	0.46
			3	11023	6900	23536	0.43	0.52	
				1	2	3			
			1	44612	14865	9623	0.35	0.27	
		SL + w_2	2	9704	7585	10350	0.73	0.77	0.44
			3	6766	9741	24952	0.40	0.44	
				1	2	3			
Moderate			1	52778	4997	11261	0.24	0.38	
		Baseline	2	14808	4041	8764	0.85	0.71	0.45
			3	18265	4661	18495	0.55	0.52	
				1	2	3			
			1	58610	0	10425	0.15	0.40	
		SL0 + w_1	2	17670	0	9944	1.00	NA	0.43
			3	21804	0	19617	0.53	0.52	
				1	2	3			
	Strong		1	60419	197	8420	0.12	0.34	
	$\sigma^2 = 20$,	SL + w_1	2	16705	91	10818	1.00	NA	0.37
	$\phi = 2$,		3	15113	184	26124	0.37	0.43	
	$\nu = 1.5$.			1	2	3			
			1	44916	10549	13571	0.35	0.32	
		SL0 + w_2	2	9889	5402	12322	0.80	0.76	0.46
			3	11039	6523	23860	0.42	0.52	
				1	2	3			
			1	45983	15416	7637	0.33	0.23	
		SL + w_2	2	8739	8609	10266	0.69	0.75	0.42
			3	5488	9964	25969	0.37	0.40	

Table C.4: Simulation results assuming strong regional heterogeneity, and RF as the base learner.

Regional Heterogeneity	Spatial Correlation	Method	True Category	Classification Results					
				Classification Table			FPR	FDR	Overall Error Rate
				1	2	3			
			1	53313	4545	11510	0.23	0.39	
		Baseline	2	15147	3902	8697	0.86	0.70	0.46
			3	19457	4314	17850	0.57	0.53	
				1	2	3			
			1	58889	0	10479	0.15	0.39	
		SL0 + w_1	2	17336	0	10410	1.00	NA	0.43
			3	21132	0	20489	0.51	0.51	
				1	2	3			
	Moderate		1	59754	0	9614	0.14	0.36	
	$\sigma^2 = 10$,	SL + w_1	2	17374	0	10372	1.00	NA	0.39
	$\phi = 2/3$,		3	16902	0	24719	0.41	0.45	
	$\nu = 1$.			1	2	3			
			1	45086	10837	13445	0.35	0.32	
		SL0 + w_2	2	9848	5512	12386	0.80	0.76	0.46
			3	11142	6783	23696	0.43	0.52	
				1	2	3			
			1	44683	14921	9764	0.36	0.26	
		SL + w_2	2	9529	7575	10643	0.73	0.76	0.44
			3	6503	9651	25467	0.39	0.44	
Strong				1	2	3			
			1	52310	5110	11800	0.24	0.40	
		Baseline	2	15097	4024	8567	0.85	0.71	0.47
			3	19457	4764	17311	0.58	0.54	
				1	2	3			
			1	58672	0	10548	0.15	0.41	
		SL0 + w_1	2	17834	0	9853	1.00	NA	0.44
			3	22609	0	18923	0.54	0.52	
				1	2	3			
	Strong		1	60532	61	8627	0.13	0.35	
	$\sigma^2 = 20$,	SL + w_1	2	16831	43	10814	1.00	NA	0.38
	$\phi = 2$,		3	16338	69	25125	0.39	0.44	
	$\nu = 1.5$.			1	2	3			
			1	44598	10081	14541	0.36	0.32	
		SL0 + w_2	2	9980	5000	12707	0.82	0.77	0.47
			3	11520	6163	23849	0.43	0.53	
				1	2	3			
			1	45632	15151	8437	0.34	0.24	
		SL + w_2	2	8846	8036	10806	0.71	0.76	0.43
			3	5987	9688	25857	0.38	0.42	

Table C.5: Simulation results for GLM + QDA + RF assuming moderate regional heterogeneity.

Regional Heterogeneity	Spatial Correlation	Method	True Category	Classification Results					
				Classification Table			FPR	FDR	Overall Error Rate
				1	2	3			
			1	58951	0	10148	0.15	0.38	
		SL0 + w_1	2	16896	0	10743	1.00	NA	0.42
			3	19615	0	21844	0.47	0.49	
				1	2	3			
			1	59638	0	9461	0.14	0.36	
		SL + w_1	2	17368	0	10271	1.00	NA	0.39
	Moderate		3	16725	0	24734	0.40	0.44	
	$\sigma^2 = 10$,			1	2	3			
	$\phi = 2/3$,		1	44469	12934	11696	0.36	0.30	
	$\nu = 1$.	SL0 + w_2	2	9160	6872	11607	0.75	0.75	0.46
			3	9656	8055	23748	0.43	0.49	
				1	2	3			
			1	44372	16068	8660	0.36	0.27	
		SL + w_2	2	9609	8294	9735	0.70	0.76	0.44
			3	6625	10711	24123	0.42	0.43	
				1	2	3			
Moderate			1	58611	0	10424	0.15	0.39	
		SL0 + w_1	2	16883	0	10730	1.00	NA	0.42
			3	20270	0	21151	0.49	0.50	
				1	2	3			
			1	59451	829	8755	0.14	0.33	
		SL + w_1	2	16010	469	11135	0.98	NA	0.37
	Strong		3	13852	770	26799	0.35	0.43	
	$\sigma^2 = 20$,			1	2	3			
	$\phi = 2$,		1	44363	13000	11672	0.36	0.30	
	$\nu = 1.5$.	SL0 + w_2	2	9006	6909	11699	0.75	0.76	0.46
			3	9612	8318	23491	0.43	0.50	
				1	2	3			
			1	45028	17682	6325	0.35	0.23	
		SL + w_2	2	8494	9964	9156	0.64	0.75	0.43
			3	5266	11918	24238	0.41	0.39	

Table C.6: Simulation results for GLM + QDA + RF assuming strong regional heterogeneity.

Regional Heterogeneity	Spatial Correlation	Method	True Category	Classification Results					
				Classification Table			FPR	FDR	Overall Error Rate
				1	2	3			
			1	59258	0	10110	0.15	0.38	
		SL0 + w_1	2	16773	0	10973	1.00	NA	0.42
			3	19807	0	21814	0.48	0.49	
				1	2	3			
			1	59526	0	9842	0.14	0.36	
		SL + w_1	2	17192	0	10555	1.00	NA	0.39
	Moderate		3	16517	0	25104	0.40	0.45	
	$\sigma^2 = 10$,			1	2	3			
	$\phi = 2/3$,		1	44795	12822	11750	0.35	0.29	
	$\nu = 1$.	SL0 + w_2	2	9049	6787	11910	0.76	0.76	0.46
			3	9703	8099	23818	0.43	0.50	
				1	2	3			
			1	44354	16161	8863	0.36	0.26	
		SL + w_2	2	9478	8375	9894	0.70	0.76	0.44
			3	6399	10893	24329	0.42	0.43	
				1	2	3			
			1	58798	0	10422	0.15	0.39	
		SL0 + w_1	2	17033	0	10654	1.00	NA	0.43
			3	21088	0	20444	0.51	0.51	
				1	2	3			
			1	59699	340	9181	0.14	0.35	
		SL + w_1	2	16243	276	11168	0.99	NA	0.38
Strong	Strong		3	15639	441	25451	0.39	0.45	
	$\sigma^2 = 20$,			1	2	3			
	$\phi = 2$,		1	44100	13002	12118	0.36	0.30	
	$\nu = 1.5$.	SL0 + w_2	2	9105	6695	11888	0.76	0.76	0.47
			3	10087	8258	23187	0.44	0.51	
				1	2	3			
			1	44573	17589	7059	0.36	0.25	
		SL + w_2	2	8626	9550	9511	0.65	0.76	0.44
			3	5933	12306	23293	0.44	0.41	

C.2.2 Application to In Vivo Data

Table C.7 and Table C.8 report the ordinal classification results using QDA and RF, respectively, as the base learner. Table C.9 reports the ordinal classification results after combining the multi-resolution GLM, QDA and RF. Each table reports the results using either the predicted probabilities for the first two categories from multi-resolution base learners, or the classified cancer categories from multi-resolution base learners, as the covariates for the stage-two model in the proposed super learner.

Table C.7: Ordinal classification results on in vivo data assuming QDA as the base learner.

Covariates for stage-two model	Method	True Category	Classification Results: Classification Table			FPR	FDR	Overall Error Rate
			1	2	3			
		1	87136	297	1554	0.021	0.152	
	Baseline	2	5891	10	342	0.998	0.975	0.168
		3	9782	90	1835	0.843	0.508	
			1	2	3			
		1	87515	0	1472	0.017	0.155	
	SL0 + w_1	2	5937	0	306	1.000	NA	0.167
		3	10136	0	1571	0.866	0.531	
			1	2	3			
Predicted Probabilities		1	87776	0	1211	0.014	0.144	
for the first	SL + w_1	2	5864	0	379	1.000	NA	0.152
two categories		3	8850	0	2857	0.756	0.358	
			1	2	3			
		1	65662	13795	9530	0.262	0.086	
	SL0 + w_2	2	2484	1312	2447	0.790	0.925	0.321
		3	3702	2344	5661	0.516	0.679	
			1	2	3			
		1	65310	16809	6868	0.266	0.081	
	SL + w_2	2	2518	1486	2239	0.762	0.929	0.321
		3	3231	2622	5854	0.500	0.609	
			1	2	3			
		1	87136	297	1554	0.021	0.152	
	Baseline	2	5891	10	342	0.998	0.975	0.168
		3	9782	90	1835	0.843	0.508	
			1	2	3			
		1	87431	0	1556	0.017	0.150	
	SL0 + w_1	2	5846	0	397	1.000	NA	0.163
		3	9582	0	2125	0.818	0.479	
			1	2	3			
		1	87883	0	1204	0.014	0.146	
Classified Categories	SL + w_1	2	5998	0	245	1.000	NA	0.154
		3	9018	0	2689	0.770	0.350	
			1	2	3			
		1	82439	1400	5148	0.074	0.118	
	SL0 + w_2	2	4345	527	1371	0.916	0.779	0.182
		3	6696	460	4551	0.611	0.589	
			1	2	3			
		1	77293	7567	4127	0.131	0.087	
	SL + w_2	2	3333	1638	1272	0.738	0.863	0.216
		3	4058	2733	4916	0.580	0.523	

Table C.8: Ordinal classification results on in vivo data assuming RF as the base learner.

Covariates for stage-two model	Method	True Category	Classification Results: Classification Table			FPR	FDR	Overall Error Rate
			1	2	3			
		1	84948	1048	2991	0.045	0.143	
	Baseline	2	5470	174	599	0.972	0.896	0.180
		3	8729	454	2524	0.784	0.587	
			1	2	3			
		1	87415	0	1572	0.018	0.146	
	SL0 + w_1	2	5821	0	422	1.000	NA	0.158
		3	9099	0	2608	0.777	0.433	
			1	2	3			
Predicted Probabilities for the first two categories		1	87382	0	1605	0.018	0.125	
	SL + w_1	2	5849	0	394	1.000	NA	0.135
		3	6590	0	5117	0.563	0.281	
			1	2	3			
		1	66000	12999	9988	0.258	0.086	
	SL0 + w_2	2	3218	1254	1771	0.799	0.925	0.311
		3	2957	2362	6388	0.454	0.648	
			1	2	3			
		1	65123	17997	5867	0.268	0.052	
	SL + w_2	2	2292	2300	1651	0.632	0.900	0.297
		3	1290	2617	7800	0.334	0.491	
			1	2	3			
		1	84927	1037	3023	0.046	0.143	
	Baseline	2	5475	172	596	0.972	0.898	0.180
		3	8671	481	2555	0.782	0.586	
			1	2	3			
		1	87495	0	1492	0.017	0.149	
	SL0 + w_1	2	5964	0	279	1.000	NA	0.160
		3	9363	0	2344	0.800	0.430	
			1	2	3			
		1	87491	0	1496	0.017	0.131	
Classified Categories	SL + w_1	2	5901	0	342	1.000	NA	0.141
		3	7316	0	4391	0.625	0.295	
			1	2	3			
		1	78988	2218	7781	0.112	0.115	
	SL0 + w_2	2	4558	305	1380	0.951	0.905	0.209
		3	5683	687	5337	0.544	0.632	
			1	2	3			
		1	71145	13255	4587	0.201	0.069	
	SL + w_2	2	3316	1543	1384	0.753	0.914	0.258
		3	1996	3044	6667	0.431	0.472	

Table C.9: Ordinal classification results for GLM + QDA + RF on in vivo data.

Covariates for stage-two model	Method	True Category	Classification Results: Classification Table			FPR	FDR	Overall Error Rate
			1	2	3			
		1	86674	0	2313	0.026	0.141	
	SL0 + w_1	2	5868	0	375	1.000	NA	0.158
		3	8371	0	3336	0.715	0.446	
			1	2	3			
		1	86153	0	2834	0.032	0.125	
	SL + w_1	2	5935	0	308	1.000	NA	0.145
Predicted Probabilities		3	6407	0	5300	0.547	0.372	
for the first			1	2	3			
two categories		1	65708	15409	7870	0.262	0.076	
	SL0 + w_2	2	2481	2064	1698	0.669	0.896	0.306
		3	2921	2336	6450	0.449	0.597	
			1	2	3			
		1	65709	17307	5971	0.262	0.062	
	SL + w_2	2	2720	2518	1005	0.597	0.891	0.298
		3	1604	3296	6807	0.419	0.506	
			1	2	3			
		1	87615	0	1372	0.015	0.146	
	SL0 + w_1	2	5943	0	300	1.000	NA	0.155
		3	9007	0	2700	0.769	0.382	
			1	2	3			
		1	87007	0	1980	0.022	0.131	
	SL + w_1	2	5969	0	274	1.000	NA	0.143
		3	7107	0	4600	0.607	0.329	
Classified Categories			1	2	3			
		1	78811	4527	5649	0.114	0.109	
	SL0 + w_2	2	4411	768	1064	0.877	0.883	0.207
		3	5252	1283	5172	0.558	0.565	
			1	2	3			
		1	71464	13560	3963	0.197	0.080	
	SL + w_2	2	3573	1909	761	0.694	0.894	0.254
		3	2664	2589	6454	0.449	0.423	