QUANTITATIVE INTEGRATION OF IMAGING AND NON-IMAGING DATA: APPLICATION TO INTEGRATING MULTI-PARAMETRIC MRI FOR PROSTATE CANCER DIAGNOSIS, GRADING AND TREATMENT EVALUATION

BY PALLAVI TIWARI

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

and

The Graduate School of Biomedical Sciences

University of Medicine and Dentistry of New Jersey

in partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

Graduate Program in Biomedical Engineering

Written under the direction of

Anant Madabhushi

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2012

ABSTRACT OF THE DISSERTATION

Quantitative Integration of Imaging and Non-imaging Data: Application to Integrating Multi-parametric MRI for Prostate Cancer Diagnosis, Grading and Treatment Evaluation

by Pallavi Tiwari

Dissertation Director: Anant Madabhushi

The problem of data integration involving imaging and non-imaging modalities is largely unexplored in the biomedical field, mainly due to the challenges in quantitatively combining such heterogeneous modalities existing in different dimensions and scales. Although several methods have been proposed in the literature involving quantitative integration of multi-protocol imaging, there has been a paucity of similar biomedical tools for quantitative integration of imaging and non-imaging data. In this work, we present novel data integration schemes to overcome the aforementioned challenges limiting the integration of imaging and non-imaging modalities, and hence improve disease characterization. Our novel data integration methods are applied to integration of multi-parametric Magnetic Resonance (MR) imaging (MP-MRI)-structural MR imaging with metabolic spectroscopic information (non-imaging) for improved prostate cancer (CaP) diagnosis, grading, and treatment evaluation post-radiation therapy (RT).

To this end, we have developed novel data integration schemes such as, Multimodal Wavelet Embedding Representation for data Combination (MaWERiC), and

Semi-Supervised Multi-Kernel (SeSMiK) Graph Embedding, which first uniformly represent individual data modalities into a common framework using dimensionality reduction and kernel embedding techniques, followed by a seamless integration of imaging and non-imaging data in the common framework. The integrated quantitative signatures thus obtained are shown to be significantly more diagnostically informative as compared to any single modality. Similar improvement in results was observed using integrated MP-MRI signatures for evaluating radiation therapy related changes in CaP patients, with an aim to identify (a) pre-RT disease extent along with extra capsule spread (if any) and (b) residual disease on post-RT MP-MRI.

Preface

This thesis represents the collective published and unpublished works of the author. Chapters 3-8 are primarily composed of the contents of conference papers [1, 2, 3, 4, 5], and peer-reviewed journal articles [6, 7, 8], written by the author of this dissertation over the course of her thesis work.

Acknowledgements

Foremost, I would like to thank Dr. Anant Madabhushi, my advisor and mentor who shared with me a lot of his expertise and research insight. His perpetual energy and enthusiasm in research has always motivated me to do my best. A big thanks also to all the members at the Laboratory of Computational Imaging and Bioinformatics who made it a convivial place to work. I would like to express my gratitude to the committee members and collaborators, Dr. John Kurhanewicz, Dr. Mark Rosen, Dr. John Tomaszewski, Dr. Michael Feldman, and Dr. Nicholoas Bloch for their insightful opinions and research discussions that made this research possible.

This work was made possible by grants from the Wallace H. Coulter Foundation, The Cancer Institute of New Jersey, National Cancer Institute (R01CA136535-01, R01CA140772-01, R21CA127186-01, R03CA143991-01), and Department of Defense (Predoctoral fellowship, W81XWH-09).

I cannot finish without saying how grateful I am to my family. Thanks to my brother Anmol, sister Ruchi and brother-in-law, Anand, for their everlasting love and support. Words cannot express how grateful I am to my parents, Swati and Suresh Tiwari, who have always motivated me to achieve excellence in all matters of life. Without them this would have never been possible. Lastly, a big hug to my biggest motivator and fan, my husband Vikas, for always being beside me as a friend, confidant, supporter, mentor and so much more. This is from me to you.

Dedication

To, Abha, the sunshine that left me too soon,

and

To, my parents, for holding my hand and pointing it to the sky.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

In the biomedical domain, there is a real need for developing data integration strategies for combining discriminatory features from multiple modalities to develop multi-modal meta-classifiers for improved disease detection, diagnosis, and prognosis [9]. While data integration methods have been proposed for quantitatively combining multiple imaging modalities [10, 11, 12], these tools are not readily extensible to integration of imaging and non-imaging data on account of differences in scale and feature dimensionality. For instance, consider the difficulties in quantitatively combining, at the voxel-level, T2-weighted (T2-w) magnetic resonance (MR) imaging (reflecting structural attributes) acquired as scalar intensity values, with MR spectroscopy (MRS) acquired as a vector (or spectrum) of metabolite concentrations; each modality encoding a different type (structural or metabolic) and dimensionality of information. Nonetheless, both modalities reflect information regarding the same region of interest they are captured from, and consequently, examining them in conjunction is crucial.

In this dissertation, we present computerized decision support (CDS) strategies that leverage combination of features from different imaging and non-imaging protocols and quantitatively integrate these features across different imaging and non-imaging protocols for improved disease diagnosis. We demonstrate the utility of our methods for quantitative integration of multi-parametric Magnetic Resonance Imaging (MP-MRI) (T2-w MRI, MRS, and diffusion weighted (DWI) MRI) for early prostate cancer diagnosis, grading, and evaluating radiation-therapy related changes in the prostate.

1.2 Significance of Multi-parametric Magnetic Resonance Imaging (MP-MRI) for prostate cancer

1.2.1 MP-MRI for prostate cancer diagnosis

Prostate cancer (CaP) is the second leading cause of cancer related deaths amongst men with 217,730 new cases and 32,050 estimated deaths in United States in 2010 (Source: American Cancer Society). The current gold standard for CaP detection is a blinded sextant trans-rectal ultrasound examination which is known to have a low detection sensitivity (20-25%) due to the poor image resolution of ultrasound. In the last decade, MR Imaging (MRI) has shown great potential for characterizing disease presence as well as staging of CaP [13]. T2-w MRI provides high resolution structural details of the gland. CaP typically appears as a hypo-intense region within the gland showing less inherent structure as compared to non-CaP regions (Figure 1.1). MRS has recently emerged as a complement to traditional T2w MRI [14]. MRS quantifies the metabolic concentrations of specific molecular markers such as choline (A_{ch}), creatine (A_{cr}), and citrate (A_{cit}) in the prostate [15]. The relative concentrations of these metabolites are recorded by calculating area under the metabolic peak and relative changes in metabolite concentrations ($A_{ch+cr}/A_{cit} > 1$) and are used to assess presence of CaP at different spatial locations in the image [15, 16]. However, the utility of MRS metabolic features for detecting, localizing, and characterizing disease is highly dependent on the quality of spectral examinations obtained, automated spectral peak detection algorithms are challenged in their ability to resolve overlapping peaks (for instance the choline peak overlaps with the creatine peak in case of CaP spectra) [17]. Figure 1.1 demonstrates one such noisy spectrum with poor signal to noise ratio and with the baseline affected by a tail of broad upfield lipid resonance (2.0-2.5 ppm). Manual diagnosis of CaP on T2w MRI and MRS involves first visually identifying hypo-intense regions on T2w MRI, followed by inspection of spectra at those corresponding spatial locations for changes in metabolite (choline, creatine, citrate) ratios.

While clinical studies have shown that the use of structural and metabolic MR information yields greater CaP detection accuracy compared to diagnosis based off any

individual modality [18] few attempts have been made to quantitatively combine the different information channels [19, 20]. Additionally, some studies have shown that manual interpretation and visual integration of multi-modal data is subjective and thus prone to inter- and intra-observer variability [21, 22]. It is hence desirable to build a data integration based computerized decision support system (DI-CDS) that can accurately extract and combine relevant information from both imaging and non-imaging data channels for improved disease diagnosis and detection [11, 23]. Such a CDS could then be integrated into a clinical setting to assist radiologists in accurate characterization, staging, as well as directing and evaluating treatment of the disease. While a few groups have proposed CDS classifiers for combining multiple MR imaging protocols (T2w, dynamic contrast enhanced (DCE), line-scan diffusion, T2-mapping MRI) [11, 23, 24], to the best of our knowledge no method for quantitative integration of T2w MRI and MRS for CaP detection has been presented thus far in the literature.

1.2.2 MP-MRI for Gleason grading of prostate cancer

Gleason grading is the most widely used grading scheme for CaP [25], with low Gleason scores being associated with better patient outcome and high Gleason score tending to be correlated with more biologically aggressive disease and worse prognoses for long-term, metastasis-free survival [26, 27]. Currently, over one million prostate biopsies are performed annually in the US, of which approximately 60-70% are negative for CaP presence [28]. The overdiagnosis and associated overtreatment due to these false positives causes severe health implications such as risk of bleeding and infection of the prostate gland or urinary tract [29]. Most of the otherwise correctly diagnosed CaP cases are identified as low grade disease, and are not destined to metastasize [30]. Such patients are now opting for a "wait and watch policy" involving active surveillance, as opposed to opting for immediate aggressive therapy [31].

MP-MRI has begun to be routinely used in several centers for staging of disease in patients previously identified with CaP. Over the last decade several researchers have been investigating MRI and MRS for staging and possible screening of CaP [14, 32, 33] with a view to reduce unnecessary biopsies in men, with elevated PSA but without

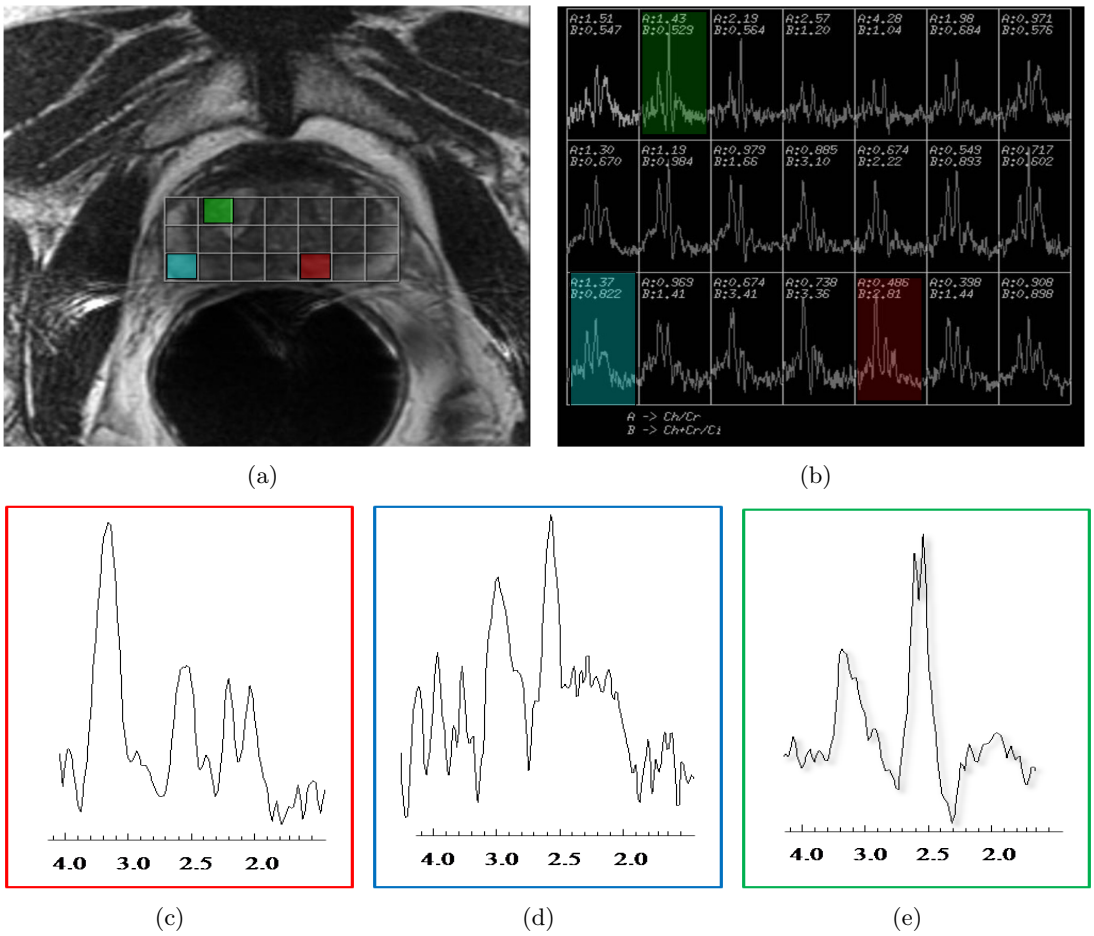


Figure 1.1: (a) shows a T2-w MRI slice of the prostate with a user-selected 3x7 voxel grid overlaid on the prostate, MRS spectral grid corresponding to the T2 slice is shown in 1.1(b). An abnormal appearing spectra (red voxel in (a) and (b)) is shown in (c), while (e) shows a normal appearing spectra (green voxel in (a) and (b)). Figure 1.1(d) shows an additional spectra (blue voxel in (a) and (b)), with poor signal to noise ratio and with the baseline affected by a tail of broad upfield lipid resonance (2.0-2.5 ppm).

CaP, who might otherwise have a significant risk of sexual, urinary, and bowel related symptoms caused due to biopsy [29].

Recently, some investigators have begun to explore the correlation between MP MRS and T2w MRI features and corresponding low and high Gleason grades of CaP [34, 35, 36]. It has been qualitatively demonstrated in clinical studies that high Gleason grade is associated with elevated ratios of A_{ch+cr}/A_{cit} [37]. Hypo-intense signal intensities on T2w MRI are also found to be significantly correlated with CaP aggressiveness [38]. In [35], qualitatively combining T2w MRI-MRS allowed for accurately predicting the presence of low grade CaP. In a similar related multi-protocol study, Shukla-Dave *et*

al. studied the correlation of T2w MRI and MRS along with expression levels of three molecular markers: Ki-67, phospho-Akt, and androgen receptor obtained via immunohistochemical analysis, to successfully differentiate clinically insignificant and significant CaP [36]. Biologically significant disease was defined based on pathologic examination of surgical specimens. Correlation of the three molecular markers with respect to combined MRI-MRS signatures was observed. Additionally, a high area under the receiver operating characteristic curve (ROC) of 0.91 was obtained for identifying significant high grade CaP using combined MRI, MRS parameters.

1.2.3 MP-MRI for evaluating treatment related changes post- Radiation Therapy (RT)

Upto 25% of all CaP patients undergo some form of radiation therapy (RT) (e.g. intensity-modulated radiation therapy (IMRT), proton beam therapy, brachytherapy) as treatment for clinically localized disease[1]. Currently, differentiation between local or systemic recurrence of CaP (which have radically different prognoses and treatment regimens) is only appreciable on trans-rectal ultrasound, that too at a relatively advanced stage [39]. Early identification of non-responders via the use of imaging will allow for modification of the therapy [39], as well as provide clues about long-term patient outcome.

Recently, *in-vivo* MP-MRI (T2-w, MRS, and DWI) has shown great potential in early identification of RT related changes in the prostate [40, 41] (Figure 1.2). Pucar et al. [39] showed that MP-MRI significantly improves identification of CaP regions pre-, post-RT, compared to sextant biopsy and digital rectal examination. Similarly, in [42] an area under the receiver operating curve (AUC) of 0.79 was obtained via qualitative examination of MP-MRI (T2w, MRS) in accurately identifying new and recurrent disease post-RT. In another similar study [41], DWI when combined with T2-w MRI was shown to significantly outperform T2w MRI alone, for accurately predicting locally recurrent CaP post-RT. Successful treatment of CaP on T2w MRI, post RT, is

[1]American Cancer Society

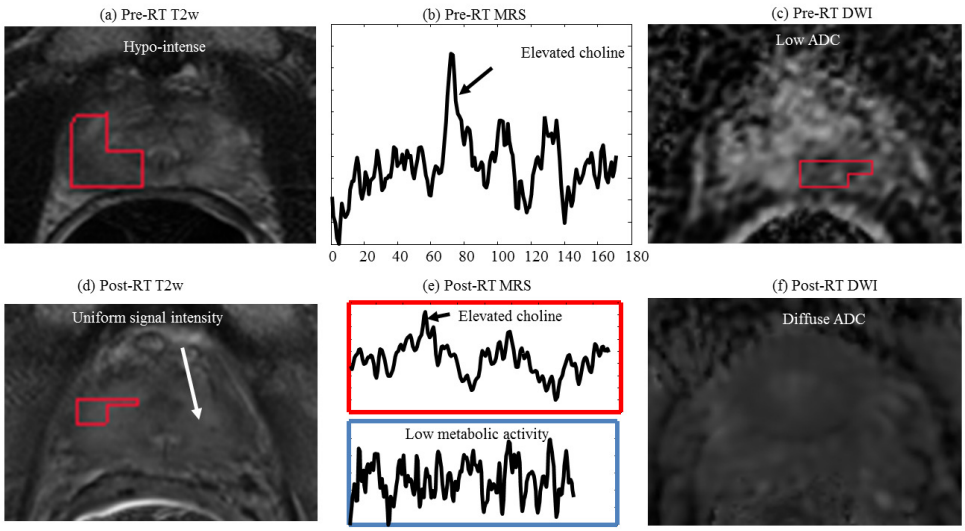


Figure 1.2: (a), (b), (c) show the pre-treatment imaging markers for T2w MRI, MRS and DWI respectively, where (a) shows hypointense T2w MRI intensity regions (outlined in red) on a single T2w MRI slice, (b) shows a typical MRS CaP spectrum characterized with elevated choline peak, and (c) shows a DWI image with CaP outlined in red, defined by low ADC values. Figures 1.2(d), (e), and (f) show the T2w MRI, MRS and DWI scenes post-RT where, (d) shows the corresponding post RT T2w MRI slice with uniform signal intensities characterizing successful treatment, while hypo intense CaP region is outlined in red, reflecting local disease recurrence, Similarly in (e) the elevated choline peak appears to suggest local CaP recurrence within a single MR spectrum (outlined in red), along with a spectrum showing low metabolic activity (outlined in blue) reflecting disappearance of disease. (f) shows CaP recurrence defined by low ADC and successful treatment characterized by diffuse ADC values.

characterized by uniform T2w signal intensity without focal abnormalities, while new or locally recurrent CaP is characterized by hypo-intense regions of smooth texture [43] (Figure 1.2(d)). MRS shows an absence of citrate, as well as low metabolic activity in cases of successful treatment (Figure 1.2(e), outlined in blue). New foci and locally recurrent CaP on post-RT MRS is characterized by elevated levels of choline [43] (Figure 1.2(e), outlined in red). Similarly, post-RT, DWI shows an overall increase in apparent diffusion coefficient (ADC) values within the entire prostate when CaP is successfully treated. Unchanged or decreased ADC values correspond to locally recurrent CaP [44] (Figure 1.2(f)).

1.3 Challenges with integrating MP-MRI protocols for CaP diagnosis and treatment evaluation

1.3.1 Challenges with combining MP-MRI for CaP diagnosis and grading

The relatively recent efforts at attempting to combine multiple MRI markers (qualitatively) to identify high grade CaP *in-vivo* are limited in that (a) the different MRI protocols are not quantitatively combined, and (b) qualitative evaluation is often subjective and prone to inter-observer variability [21, 22]. It is hence desirable to build a CDS that can (a) quantitatively integrate relevant MP MRI data to create meta-classifiers that can identify disease presence *in-vivo* [45, 24, 46, 47] (b) can subsequently characterize the Gleason grade of areas ascertained by the meta-classifier, most likely to be CaP, and (c) assist in early identification of treatment related changes (such as recurrence, new CaP foci) in CaP patients. Such a CDS could then be used in a clinical setting to assist a radiologist in making a more informed diagnosis of the presence, extent and aggressiveness of the disease. However, one of the major challenges in constructing a meta-classifier that can quantitatively combine heterogeneous imaging and non-imaging modalities such as T2w MRI and MRS, is to overcome the differences in dimensionality and resolution associated with each of the heterogeneous imaging protocols; each modality encoding a different type (structural or metabolic) and dimensionality of information. Nonetheless, both modalities reflect information regarding the same region of interest they are captured from, and consequently, examining them in conjunction is crucial (Figure 1.3). One of the major challenges in quantitatively integrating imaging and non-imaging data is to represent them in a unified framework prior to integration.

1.3.2 Challenges with MP-MRI for evaluating treatment related changes post-RT

Visual examination of post-RT MRI for evaluating treatment related changes and residual disease is usually associated with poor detection rates due to (a) diffuse T2w signal

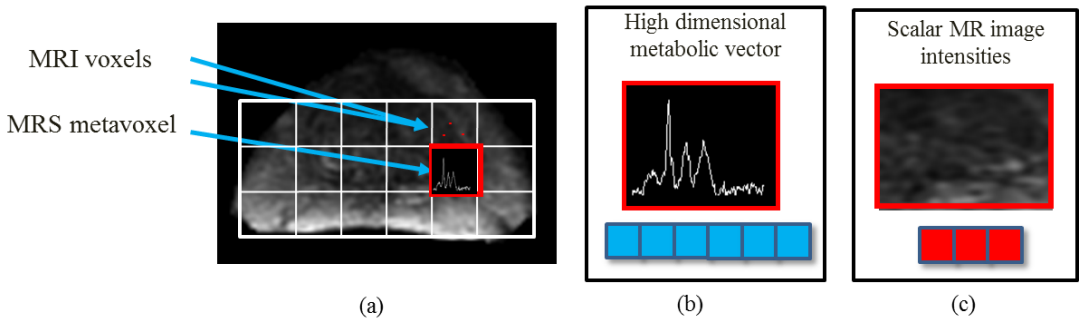


Figure 1.3: (a) shows a MRS grid superposed on a T2w MRI image. (b) shows a high dimensional spectral signal from a spatial location on the MRI grid, and (c) shows the scalar T2w MRI intensities corresponding to the spatial location shown in (b). Note the heterogeneity in the data, which lead to difficulties in quantitative integration of these imaging and non-imaging channels of information.

intensity and indistinct zonal anatomy on T2w MRI [48], (b) adverse effects of post-biopsy hemorrhage and hormonal therapy which in turn adversely affects identification of metabolic peaks [49], and (c) significant gland shrinkage and distortion post-RT [39]. Automated quantitative assessment of RT changes on a per voxel basis may thus allow accurate and precise identification of (a) residual disease, and (b) new foci of cancer (local recurrence) within the prostate. Additionally, since each of the individual MR imaging markers provide orthogonal information (structural, functional, metabolic), a MP-MRI approach that can leverage multiple imaging channels could significantly improve detection specificity and sensitivity.

A few major challenges associated with developing an automated quantitative scheme for pre-, post-RT evaluation include, (a) developing elastic registration tools to deal with changes in shape and size of the prostate gland pre-, post-RT, (b) accurate alignment of various MP imaging protocols for computing voxel level absolute difference of the imaging markers (reflective of treatment changes), and (c) optimized weighted quantitative integration of imaging marker changes across MP-MRI.

1.4 Summary of the major goals of this thesis

In this work, we develop novel CDS strategies that seamlessly integrate imaging and non-imaging features by overcoming the aforementioned challenges with registration, representation and integration of heterogeneous protocols. We demonstrate the utility

of our approaches by quantitatively integrating MP-MRI data for (a) prostate cancer diagnosis, (b) Gleason grading, and (c) evaluating treatment related changes in the prostate. The CDS strategies developed in this dissertation are described in the following chapters,

In Chapter 3 and 4, feature extraction, and data integration strategies developed in this work are respectively described. In Chapters 5 and 6, we employ the techniques described in Chapters 3 and 4 for integrating MP-MRI for CaP diagnosis; while in Chapter 7, these methods are employed for identifying high-grade aggressive prostate cancer using integrated MP-MRI signatures. In Chapter 8, MP-MRI signatures corresponding to treatment-related changes in CaP patients are developed and evaluated for early identification of CaP recurrence in the prostate. The remainder of this thesis is presented based on examining each of following topics in turn.

- Developing novel feature extraction strategies for extracting unique features across imaging and non imaging protocols for common data representation.
- Developing novel data integration strategies for integrating imaging and non-imaging protocols.
- Demonstrating an application of the feature extraction and data integration strategies developed in this work for CaP diagnosis via MP-MRI.
- Demonstrating an application of the feature extraction and data integration strategies developed in this work for CaP grading via MP-MRI.
- Developing novel MP-MRI signatures for evaluating radiation-treatment related changes in CaP patients.

Chapter 2

Previous Work and Novel Contributions

2.1 Previous Work in multi-modal data integration

Broadly speaking, MP data fusion strategies, from a classification perspective, may be categorized as combination of data (COD) (where the information from each channel is combined prior to classification), and combination of interpretations (COI) (where independent classifications based on the individual channels are combined) [58]. It has been suggested that COI approaches are less than optimal since binary classifier outputs from individual classifiers are combined without accounting for inter-channel dependencies [59]. Consequently, several COD strategies with the express purpose of building integrated quantitative meta-classifiers have been proposed, including feature-based [60], kernel-based [61], and dimensionality reduction (DR)-based [62] strategies.

2.1.1 Feature-based COD schemes

In COD strategies, MP data is combined by direct concatenation of raw features in their original acquired form (such as unfiltered image intensities). In [60], image intensities obtained via five structural MRI sequences (B0, diffusion weighted images, FLAIR, T1-w and gadolinium-enhanced T1-w) and two scalar maps obtained via diffusion tensor imaging (fractional anisotropy and apparent diffusion coefficient) were concatenated into a single feature vector which was then used to train a classifier for brain tumor diagnosis. However, directly aggregating heterogeneous data channels without accounting for differences in the dimensionality (number of features) and their relative scaling, can adversely impact classifier performance [63]. This is especially true in the case of combining imaging and non-imaging data.

2.1.2 Multi-kernel learning (MKL)

Another recent COD scheme is multi-kernel learning (MKL) [61], which attempts to overcome differences in dimensions and resolutions across imaging protocols, by first representing the data in a common kernel space, prior to data combination and meta-classification. Kernels [64] are positive definite functions which transform the input data to a dot product similarity space such that $K(\mathbf{F}(c_i), \mathbf{F}(c_j)) = \langle \Phi((\mathbf{F}(c_i)), \Phi(\mathbf{F}(c_j))) \rangle$, where Φ is the implicit pairwise embedding between input feature vectors $\mathbf{F}(c_i)$ and $\mathbf{F}(c_j)$ associated with points c_i, c_j, and $\langle . \rangle$ denotes the dot product operation. MKL involves computing similarity matrices for kernels derived from the individual modalities being combined, so that fused classifiers (within the fused kernel space) can be trained in order to make integrated predictions. In [61], Lanckriet *et al.* transformed data from amino acid sequences, protein complex data, gene expression data, and protein interactions into a common kernel space. The kernels were linearly combined and used to train a support vector machine (SVM) classifier for classifying functions of yeast proteins. However, when a large amount of information is present in each input source, most COD methods, including MKL, suffer from the curse of dimensionality problem [63].

2.1.3 Previous work in developing CDS strategies for CaP diagnosis

Unimodal classifiers for T2-w MRI

Since image intensity on T2-w MRI, is susceptible to artifacts such as bias field inhomogeneity [65] and intensity non-standardness [66], researchers have explored alternate representations of T2-w image intensities (e.g. Gabor or wavelet based texture features [67]) to build classifiers for predicting CaP presence on MRI. In [67], Madabhushi et al. presented a supervised CDS system for detection of CaP from 4 Tesla (T) ex vivo prostate T2-w MRI where 33 3D texture features (statistical, gradient, and Gabor) were quantitatively extracted at each voxel (T2-w MRI spatial resolution). These extracted features were then used to train a number of supervised classifiers (Adaboost, Bayes, and Decision Trees) which were employed to assign a probability of CaP presence at

each image voxel.

Unimodal classifiers for MRS

Previous CDS approaches that have been developed in the context of MRS data can be broadly divided into two main categories: (a) signal quantification (model dependent) [68, 69, 70], and (b) statistical pattern recognition based (model independent) approaches [67, 71, 72, 73]. Commonly used MRS quantification methods include VARPRO [69], AMARES [70], and QUEST [68], which are software utilities where the objective is to minimize the squared distance between the acquired data and a model basis function built on prior knowledge about the metabolic profiles of a typical MR spectrum. Pattern recognition based features on the other hand, try to capture the underlying variance in the data using regression analysis. Kelm et al. [71] presented a comparative study of classification techniques for prostate MRS data based on pattern recognition methods such as PCA [74] and Independent Component Analysis (ICA) [75] against quantification based feature extraction methods using SVM, Random Forest (RF) and Gaussian processes classifiers. They showed that pattern recognition based classifiers provided better classification results for CaP detection compared to MRS quantification schemes. In [73] we presented a CDS for CaP detection using 1.5 Tesla in vivo prostate MRS where each prostate spectrum was classified, on a per voxel basis, as either belonging to cancerous or non-cancerous classes using a hierarchical, clustering scheme in conjunction with non-linear dimensionality reduction (NLDR) methods. NLDR schemes were employed to obtain a low dimensional representation of high dimensional MR spectra, followed by hierarchical k-means clustering to identify CaP signatures in the prostate. A sensitivity of 89.33% and a specificity of 79.79%, on a per voxel basis, were obtained across a total of 18 1.5 T prostate MRS studies. Luts et al. [72] presented a method which leveraged ICA and Relief-F in conjunction with SVM and linear discriminant analysis classifiers for brain tumor classification using MRS.

Combining imaging-imaging MRI channels

While DSS schemes for CaP detection using individual MRS [76, 73], and T2w MRI [23] protocols have been proposed, we are not aware of any works that have attempted quantitative integration of imaging and non-imaging protocols. In [11] Chan et al. presented a statistical classifier which integrated texture features from multi-protocol 1.5 T in vivo MRI to generate a statistical probability map representing likelihoods of cancer for different regions within the prostate. Liu et al. [77] examined multi parametric in vivo MRI maps (T2-w, DCE, DWI) within a fuzzy Markov Random Fields framework. The maps were generated via curve fitting of data from each of the protocols with the ROI limited to the peripheral zone of the prostate, while the evaluation of the results was done against manually delineated CaP regions on MRI (with corresponding whole-mount histology and ex vivo MRI data used for reference). Ampeliotis et al. [78] explored the use of image intensity features from both DCE and T2-w MRI data for the classification of CaP. A statistically significant improvement in classifier performance when fusing modalities, over the use of individual modalities, was reported. Another multi-protocol (DCE and T2-w) MRI based CSS was presented in [24] which combined pharmacokinetic features from DCE along with T2-w image intensities. A comparison of different supervised and unsupervised methods for CaP segmentation using MP MRI was presented in [79]. However, for none of these methods is it immediately obvious how one might extend existing frameworks to combining imaging and non-imaging data with widely differing dimensionalities.

Decision Integration strategies for integrating MRI, MRS

To the best of our knowledge, no data or decision integration methods for combining imaging and spectroscopy in the context of prostate cancer have been proposed. Jesneck et al. [80] proposed a decision integration scheme where probabilities for breast cancer presence obtained from classifiers built individually from features extracted from different imaging modalities (sonogram, mammogram) and patient history (non-imaging) were combined to obtain an integrated classifier for improved breast cancer diagnosis.

Another COI scheme was presented in [81], which combined classifier outputs from three heterogeneous modalities: face recognition, voice recognition, and hand geometry within a Bayesian framework for improved biometric based personnel identification. However, one of the major disadvantages of such decision integration based approaches is that all inter-source dependencies between modalities may be lost, given that each modality is being treated independently [82].

Data Integration strategies for integrating MRI, MRS

A data integration method involving integration of multi-protocol MR image intensities (T1-w, T2-w, proton density-weighted, and gadolinium-DTPA) with the areas under spectral peaks of specific metabolites (myo-inositol, glucose, choline, creatine, glutamate/glutamine, N-acetyl aspartate, lactate/fatty acids and fatty acids) from MRS was presented in [83] for classifying four brain tumor types (Tumor II, III, IV, meningioma), healthy tissues and cerebrospinal fluid (CSF). All MRS and MRI features were directly concatenated into a single joint feature vector and employed in conjunction with a Mahalanobis distance based classifier. The classifier results showed that the voxel-level classification obtained via this multi-modal feature combination was significantly superior compared to the results obtained using unimodal classifiers.

2.2 Novel Contributions

2.2.1 Developing novel quantitative dimensionality reduction based feature extraction strategies

The first goal of this work is to evaluate and develop novel quantitative signatures on individual imaging and non-imaging protocols, in order to develop a common data representation platform across different imaging and non-imaging protocols. We will first investigate dimensionality reduction techniques commonly used in the literature for the purposes of data representation of high dimensional biomedical data. We will then discuss the limitations of some of the parameter (κ) based non-linear DR techniques, such as Locally linear embedding (LLE).

We will then present a novel Consensus-LLE (C-LLE) algorithm for creating a stable low dimensional representation of the data, in a manner analogous to building classifier ensembles such as Breiman's Bagging scheme [84]. Instead of attempting to estimate a single globally optimal κ value as in [62], [85] to be applied to the entire dataset, our scheme aims to estimate the true pairwise object adjacency $\mathcal{D}(c,d)$ in the low dimensional embedding between two objects $c,d \in C$. We formulate the problem of estimating object distances $\mathcal{D}(c,d)$ as a Maximum Likelihood Estimation problem (MLE) from multiple approximations $\mathcal{D}_\kappa(c,d)$ obtained by varying κ, which we assume are unstable and uncorrelated. Our scheme thus differs from related work in two fundamental ways: (a) C-LLE attempts to reconstruct the true low dimensional data manifold by learning pairwise object distance across the entire data space and avoids the κ estimation problem, and (b) C-LLE learns the low dimensional manifold in a locally adaptive fashion, compared to [62], [85] that attempt to learn an optimal κ value which is then uniformly applied to learning the manifold across the entire data space.

2.2.2 Developing integrated MP-MRI signatures for prostate cancer diagnosis and grading

In the second goal, we will present two novel data integration strategies, Multimodal Wavelet Embedding Representation for data Combination, MaWERiC, and Semi supervised Multi-Kernel Graph Embedding (SeSMiK-GE) that will leverage both imaging and non-imaging features to obtain an integrated quantitative signature.

The first data integration strategy, MaWERiC, is specifically geared towards quantitative integration of imaging and non-imaging data. MaWERiC comprises of two transformation modules, (i) wavelet transformation and (ii) principal component analysis which together provide a platform for uniform and homogeneous data integration across modalities. The homogeneous, low-dimensional representation of disparate data sources obtained via MaWERiC is then combined in the Eigen space. The MaWERiC data integration framework provides a general framework for potentially integrating any combination of heterogeneous data modalities, independent of scales and dimensions.

It considers equal contributions from each protocol while computing the integrated low dimensional representation.

The second data integration strategy, SeSMiK-GE, employs MKL to represent each data channel in a common kernel framework, followed by a linear weighted combination of individual data kernels. To avoid the curse of dimensionality, the combined kernel is then reduced to a lower dimensional space using GE [52], which employs partial label information to maximize class separation using SSL thereby allowing for construction of a more accurate low dimensional representation of the different data sources. While MKL has previously been employed for a variety of multi-modal data integration strategies for biomedical applications [61, 86, 87], none of these methods have been employed in conjunction with DR.

2.2.3 Developing integrated MP-MRI signatures for evaluating radiation therapy (RT) related changes in the prostate

In the third and final goal, we will present a novel data integration strategy which optimally weights contributions from differences of individual imaging markers in accurately evaluating pre-, post-RT prostate cancer via MP-MRI. Different MRI protocols from pre- and post-RT MRI scans will be first registered on a per-voxel basis. Functional, structural, and metabolic difference maps will then be obtained individually using DWI, T2w, and MRS respectively, by taking a scaled absolute difference of the imaging markers pre-, post- RT. A combined weighted MP-MRI map is then created by leveraging differences across multiple imaging markers. We believe that such an accurate per-voxel based quantitative evaluation of treatment changes pre-, post-RT will have a high clinical impact in monitoring treatment effectiveness, and could be used to modify treatment regimen early, in cases of studies with new foci or recurrence of CaP.

Chapter 3

Feature Extraction and Classification Strategies

3.1 Notation

Let $\mathcal{F} = [\mathbf{F}(c_1), \mathbf{F}(c_2), ..., \mathbf{F}(c_n)] \in \mathbb{R}^D$ be a data matrix of n objects, c_i, $i \in \{1, ..., n\}$, with dimensionality D. The corresponding class labels for these objects are given as $\omega_i \in \{0, 1\}$. Let $G = \{\mathcal{F}, W\}$ be an undirected weighted graph with vertex set $\mathcal{F}$ and similarity matrix $W \in \mathbb{R}^{n \times n}$. $W = [w_{ij}]$ assigns edge weight similarities in a pairwise fashion between objects c_i and c_j, $i, j \in \{1, ..., n\}$. The diagonal matrix $\mathcal{D}$ and Laplacian matrix L of a graph G is defined as: $\mathcal{D}_{ii} = \sum_i w_{ij}$, where $L = \mathcal{D} - W$. A kernel gram matrix defining similarities between n points is given as K_m for protocol m, $m \in \{1, ..., M\}$, where M is the total number of protocols (or data channels).

The 3D prostate T2w MRI scene is represented by $\hat{\mathcal{C}} = (\hat{C}, \hat{f})$, where $\hat{C}$ is a 3D grid of voxels $\hat{c} \in \hat{C}$ and $\hat{f}(\hat{c})$ is a function that assigns an intensity value to every $\hat{c} \in \hat{C}$. Hence at every metavoxel $c \in C$, the corresponding intensity feature vector is denoted as $\boldsymbol{F}^{T2}(c)$ while the corresponding mean Gabor wavelet feature vector (details in the next section) is denoted as $\boldsymbol{F}^{T2}_{w}(c)$. We define a metavoxel in the MRS grid as $c \in C$, where C is a 3D grid of MRS metavoxels. For each $c \in C$, $\boldsymbol{F}(c) = [f_\alpha(c)], \alpha = \{1, ..., M\}$, represents the MR spectral vector, reflecting the frequency component of each of the metabolites being measured [73]. For MRS, the feature vector comprised of ratios of concentrations of metabolites is denoted as $\boldsymbol{F}^{MRS}(c)$, while the corresponding Haar wavelet feature vector for each $c \in C$ is denoted as $\boldsymbol{F}^{MRSw}(c)$. Note that on account of differences in resolution of MRI/MRS, a single spectral metavoxel is several times larger compared to the size of a corresponding T2w MRI voxel. We define a CaP classifier output as $\mathbf{h}(c)$ $\mathbf{h} \in \{\mathbf{RF}, \mathbf{SVM}, \mathbf{PBT}\}$ where **RF** is a random forest, **SVM** is a support vector machine, and **PBT** is a probabilistic boosting tree classifier (described in the

subsequent Sections). The corresponding pairwise probabilistic markov model (PPMM) classifier is defined as $\tilde{\mathbf{h}}(c)$, and a high grade CaP classifier output is defined as $\hat{\mathbf{h}}(c)$. Similarly, notation for a classifier trained in conjunction with different feature vectors is identical to the corresponding notation for the feature vectors and involves replacing the $\mathbf{F}$ with $\mathbf{h}$ (e.g. a CaP classifier that leverages T2w MRI features in $\mathbf{F}^{T2}(c)$ is denoted as $\mathbf{h}^{T2}(c)$, corresponding MRF CaP classifier as $\tilde{\mathbf{h}}^{T2}(c)$ and the corresponding high grade CaP classifier as $\hat{\mathbf{h}}^{T2}(c)$). The common notations used in this dissertation are listed in Table 11.2.

3.2 Wavelet feature extraction of spectral signals

In mathematical terms, the wavelet transform is defined as a convolution of the wavelet function η with the signal $\boldsymbol{F}(c)$. The spectral (non-imaging) signal $\boldsymbol{F}(c)$ is convolved simultaneously with a high pass (η_h) and a low pass filter (η_l) to obtain the corresponding high ($\mathcal{H}_h$) and low filter ($\mathcal{L}_l$) coefficients as,

$$\mathcal{H}_h = \eta_h * \boldsymbol{F}(c) \tag{3.1}$$

$$\mathcal{L}_l = \eta_l * \boldsymbol{F}(c) \tag{3.2}$$

where $*$ is the convolution operator and dimensionality of coefficients $\mathcal{H}_h$ and $\mathcal{L}_l$ is $M/2$ (M is the dimensionality of each spectrum). Downsampling or decimation by a factor 2 is performed on coefficients after each pass through filters during the multi-level decomposition. In this work we considered the wavelet packet decomposition (WPD) scheme for extracting spectral wavelet features as suggested in [88], in which both $\mathcal{H}_h$ and $\mathcal{L}_l$ coefficients are iteratively decomposed using high (η_h) and low (η_l) pass filters, into a full tree like structure of a pre-defined length K, producing a total of 2^K coefficients. The reconstruction of the signal in WPD is then performed by using the best basis algorithm [89] which combines the coefficients that minimize the entropy at each level of the tree. Hence, for each spectrum $\boldsymbol{F}(c)$, at each $c \in C$, an M dimensional wavelet feature vector $\boldsymbol{F}^{MRSw}(c)$ is extracted using a Haar wavelet basis function. M varies as a function of the number of coefficients retained by the best basis algorithm, which in turn aims to minimize the entropy for each spectrum.

3.3 Metabolic ratio quantification of spectral signals

In the clinic, radiologists typically assess presence of CaP on MRS based on the choline (A_{ch}), creatine(A_{cr}), citrate peaks (A_{cit}) and the A_{ch+cr}/A_{cit} ratio. Variations in these values from predefined normal ranges ($A_{ch+cr}/A_{cit} < 1$) is highly indicative of the presence of the disease. However, peak detection algorithms in the context of MRS are often limited in their ability to deal with the poor signal to noise ratio (54). In this work, we have developed a simple, yet accurate peak detection algorithm to calculate A^{ch}, A^{cr}, A^{cit} peaks and the corresponding ratios($A_{ch}/A_{cr}, A_{ch+cr}/A_{cit}$). The algorithm is initialized with an approximate range of parts per million (ppm) for each of choline (v^{ch}), creatine (v^{cr}), and citrate (v^{cit}) peaks. For each spectral voxel, $c \in C$, the peak value $f^{\beta}_{\max}(c) = \max[v^{\beta}]$ is obtained over the range v^{β} for each specific metabolite $\beta \in \{ch, cr, cit\}$. The first minima on both left and right sides of the peak value $f^{\beta}_{\max}(c)$ are then calculated as $f^{\beta}_{l}, f^{\beta}_{r}$. If either of f^{β}_{l} or f^{β}_{r} is found to be greater than a predefined threshold (defined as 25% of $f^{\beta}_{\max}(c)$), the algorithm iteratively finds the next minima on either side that does not satisfy this condition. The area (A^{β}) under the metabolite peak is calculated between the final range $\beta \in \{ch, cr, cit\}$, using the composite trapezoidal rule.

3.4 z-score

z-score is a statistical measure defined as the ratio of the difference between the population mean and individual score to the population standard deviation. For a set of voxels, Φ^{tr} of c, $\Phi^{tr} \subset C$, the mean spectral vector $\boldsymbol{F}^{\mu} = [f^{\mu}_{u} | u \in \{1, ..M\}]$ is obtained and the corresponding standard deviation vector $\boldsymbol{F}^{\sigma} = [f^{\sigma}_{u} | u \in \{1, ...M\}]$, where $f^{\mu}_{u} = \frac{1}{|\Phi^{tr}|} \sum_{c \in \Phi^{tr}} f_u(c)$ and $f^{\sigma}_{u} = \sqrt{\frac{1}{|\Phi^{tr}|} \sum_{c \in \Phi^{tr}} [f_u(c) - f^{\mu}(c)]^2}$. The z-score at each $c \in C$ is given as $z(c) = \frac{||\boldsymbol{F}(c) - \boldsymbol{F}^{\mu}||_2}{||\boldsymbol{F}^{\sigma}||_2}$, where $|\Phi^{tr}|$ is the cardinality of Φ^{tr}. A predefined threshold θ_z is then used to identify each $c \in C$ as cancerous or not based on whether $z(c) \geq \theta_z$.

3.5 Dimensionality Reduction (DR) Techniques

The analysis and classification of high-dimensional biomedical data has been significantly facilitated via the use of dimensionality reduction techniques, which allow classifier schemes to overcome issues such as the *curse of dimensionality*. This is an issue where the number of variables (features) is disproportionately large compared to the number of training instances (objects) [90]. Dimensionality reduction (DR) involves the projection of data originally represented in a N-dimensional (N-D) space into a lower n-dimensional (n-D) space (known as an *embedding*) such that $n << N$. DR techniques are broadly categorized as linear or non-linear, based on the type of projection method used.

3.5.1 Principal Component Analysis (PCA)

PCA is a linear DR method widely used to visualize high-dimensional data and discern object relationships in the data by finding orthogonal axes that contain the greatest amount of variance in the data [74]. These orthogonal Eigen vectors corresponding to the largest Eigenvalues are called '*principal components*'. To obtain these principal components each data point c in set C is first centered by subtracting the mean of all the features for each observation c from its original M dimensional feature value $\boldsymbol{f}_u(c)$, $u \in \{1, ..., M\}$ as shown in Equation 3.3.

$$\bar{\boldsymbol{f}_u}(c) = \boldsymbol{f}_u(c) - \frac{1}{|C|} \sum_{c \in C} \boldsymbol{f}_u(c), u \in \{1, .., M\} \tag{3.3}$$

From feature values $\bar{\boldsymbol{f}}(c)$ for each $c \in C$, a new $|C| \times M$ matrix $\mathcal{Y}$ is constructed, where $|C|$ is the cardinality of set C. The matrix $\mathcal{Y}$ is then decomposed into corresponding singular values as shown in Equation 3.4.

$$\mathcal{Y} = UW_{PCA}V^T, \tag{3.4}$$

where via singular value decomposition a $|C| \times |C|$ diagonal matrix W_{PCA} containing the Eigenvalues of the principal components, a $m \times |C|$ left singular matrix U, and a

$M \times |C|$ matrix V, are obtained. The Eigenvalues in W_{PCA} represent the amount of variance for each Eigen vector $\boldsymbol{S}_v^{PCA}$, $v \in \{1, 2, ..., m\}$, in matrix V^T and are used to rank the corresponding Eigen vectors in the order of greatest variance. Thus the first m Eigen vectors that represent a pre-specified percentage of the variance in the data are extracted while the remaining Eigen vectors are discarded. Thus each data sample $c \in C$ is now described by an m-dimensional embedding vector $\boldsymbol{S}^{PCA}(c)$. In spite of the fact that PCA assumes that the data lies on a linear manifold, it allows for specification of the number of Eigen vectors required to explain a pre-specified percentage of the variance in the data.

3.5.2 Non-linear Dimensionality Reduction

Due to inherent non-linearities in biomedical data, non-linear dimensionality reduction (NLDR) schemes such as Locally Linear Embedding (LLE) and Graph Embedding (GE) have begun to be employed for data analysis and visualization. LLE [91] attempts to preserve geodesic distances between objects, while projecting the data from the high to the low dimensional feature spaces unlike linear DR schemes such as PCA which preserve the Euclidean distances between objects. GE [52] preserves object adjacencies in the data by non-linearly projecting the data from high to low dimensional space.

Graph Embedding (GE)

The aim of GE [52] is to reduce the data matrix $\mathcal{F} \in \mathbb{R}^D$ into a low-dimensional space $\mathbf{y} \in \mathbb{R}^d$ ($D >> d$), such that object adjacencies are preserved from $\mathbb{R}^D$ to $\mathbb{R}^d$. GE attempts to find the optimal low dimensional vector representations among the vertices of G that best characterize the similarity relationship between the vertex pairs in G. The low dimensional representation $\mathbf{y} = [y_1, y_2, ..y_n]$ can be obtained by solving,

The aim of GE [52] is to reduce the data matrix $\mathcal{F} \in \mathbb{R}^D$ into a low-dimensional space $\mathbf{y} \in \mathbb{R}^d$ ($D >> d$), such that object adjacencies are preserved from $\mathbb{R}^D$ to $\mathbb{R}^d$. Let $\mathcal{F} = [\mathbf{F}(c_1), \mathbf{F}(c_2), ..., \mathbf{F}(c_n)] \in \mathbb{R}^D$ be a data matrix of n objects, $i \in \{1, ..., n\}$, with dimensionality D, and $\mathbf{y} = [y_1, y_2, ..., y_n]$ be the corresponding optimal low dimensional projection matrix. $\mathbf{y}$ can be obtained by solving,

$$\mathbf{y}^* = \arg\min_y (\sum_{i,j=1}^{n} ||y_i - y_j||^2 w_{ij}), \tag{3.5}$$

where $W = [w_{ij}]$ is a similarity matrix which assigns edge weights to characterize similarities between pairwise points c_i and c_j, $i, j \in \{1, ..., n\}$. Solving equation (3.5),

$$\mathbf{y}^* = \arg\min_y [\sum_{i=1}^{n} ||y_i||^2 - 2\sum_{i,j}^{n} y_i y_j + \sum_{j=1}^{n} ||y_j||^2] w_{ij},$$

Since i and j are notations and can be interchanged, it is reasonable to assume that,

$$\sum_{i=1}^{n} ||y_i||^2 w_{ij} = \sum_{j=1}^{n} ||y_j||^2 w_{ij} = \sum_{i=1}^{n} ||y_i||^2] w_{ii}$$

Hence equation (3.5) reduces to,

$$\mathbf{y}^* = \arg\min_y 2[\sum_{i=1}^{n} ||y_i||^2 w_{ii} - \sum_{i,j}^{n} y_i y_j w_{ij}],$$

which in matrix notation form is written as

$$\boldsymbol{y}^* = \arg\min_y (\boldsymbol{y} D \boldsymbol{y}^T - \boldsymbol{y} W \boldsymbol{y}^T) = \arg\min_y [\boldsymbol{y}(D - W)\boldsymbol{y}^T] = \arg\min_{\boldsymbol{y}} [\boldsymbol{y} L \boldsymbol{y}^T] \tag{3.6}$$

where D is a diagonal matrix, $D_{ii} = \sum_i W_{ii}$. Now to minimize Equation (3.6), $\mathbf{y}^*$ needs to be differentiated, such that $\frac{\partial \mathbf{y}^*}{d\boldsymbol{y}} = 2\boldsymbol{y}L = 0$ which makes either $\boldsymbol{y} = 0$ (all output embeddings converge to a single point at $y = 0$), or $L = 0$, (all elements of L are 0). To avoid this, a constraint is defined such that $\boldsymbol{y}^T D \boldsymbol{y} = I$. Now according to Lagrange multipliers, optimization equation becomes,

$$\frac{\partial \boldsymbol{y}^*}{d\boldsymbol{y}} = \arg\min_y [\boldsymbol{y} L \boldsymbol{y}^T - \lambda(\boldsymbol{y} D \boldsymbol{y}^T - I)]$$

$$2L\boldsymbol{y} - 2\lambda D\boldsymbol{y} = 0$$

The minimization hence reduces to an Eigenvalue decomposition problem,

$$L\boldsymbol{y} = \lambda D\boldsymbol{y}$$

$L = D - W$, hence the equation reduces to,

The solution to this objective function can be obtained using an Eigenvalue decomposition problem as,

$$(D - W)\boldsymbol{y} = \lambda D\boldsymbol{y},$$

$$(1 - \lambda)D\boldsymbol{y} = W\boldsymbol{y},$$

$$W\boldsymbol{y} = \lambda D\boldsymbol{y}$$

with the constraint $\boldsymbol{y}^T \mathcal{D}\boldsymbol{y} = 1$. $W = [w_{ij}]$ is a similarity matrix which assigns edge weights to characterize pairwise similarities between points c_i and c_j, $i, j \in \{1, ..., n\}$, $w_{ij} = \frac{-||F(c_i) - F(c_j)||}{\sigma}$, where σ is the scaling parameter.

Locally Linear Embedding (LLE)

LLE [91] operates by assuming that objects within a local neighborhood in a high dimensional feature space are linearly related. Consider the set of high dimensional feature vectors $\mathcal{F} = \{\boldsymbol{F}(c_1), \boldsymbol{F}(c_2), \ldots, \boldsymbol{F}(c_n)\}$, $\forall c_i \in C, i \in \{1, ..., n\}$. LLE aims to map the set $\mathcal{F}$ to the corresponding set $\mathcal{X}_{LLE} = \{\boldsymbol{S}^{LLE}(c_1), \boldsymbol{S}^{LLE}(c_2), \ldots, \boldsymbol{S}^{LLE}(c_n)\}$ of embedding co-ordinates. Let $d^{(1)}, ..., d^{(K)}$ be the K nearest neighbors of c_i and let $\eta^K(c_i)$ be the indices of the location of the K-nearest neighbors (K-NN) of $c_i \in C$. The feature vector $\boldsymbol{F}(c_i)$ and its K-NN's $\{\boldsymbol{F}\left(d^{(1)}\right), \boldsymbol{F}\left(d^{(2)}\right), \ldots, \boldsymbol{F}\left(d^{(K)}\right)\}$ are assumed to lie on a patch of the manifold that is locally linear, allowing us to use the Euclidean metric to determine distance between neighbors. Each $\boldsymbol{F}(c_i)$ can then be approximated by a weighted sum of its K-NN. The optimal reconstruction weights are given by the sparse matrix W_{LLE} (subject to the constraint $\sum_j W_{LLE}(i, j) = 1$) that minimizes

$$E_1(W_{LLE}) = \sum_{i=1}^{n} \left\| \boldsymbol{F}(c_i) - \sum_{r=1}^{K} W_{LLE}(i, \eta^r(c_i)) \boldsymbol{F}\left(d^{(r)}\right) \right\|_2 . \tag{3.7}$$

Having determined the weighting matrix W_{LLE}, the next step is to find a low-dimensional representation of the points in $\mathcal{F}$ that preserves this weighting. Thus, for each $\boldsymbol{F}(c_i)$ approximated as the weighted combination of its K-NN, its projection $\boldsymbol{S}^{LLE}(c_i)$ will be the weighted combination of the projections of these same K-NN. The optimal $\mathcal{X}_{LLE}$ in the least squares sense minimizes

$$E_2\left(\mathcal{X}_{LLE}\right) = \sum_{i=1}^{n} \left\| \boldsymbol{S}^{LLE}\left(c_i\right) - \sum_{j=1}^{n} W_{LLE}\left(i,j\right) \boldsymbol{S}^{LLE}\left(c_j\right) \right\|_2 = trace\left(\mathcal{X}_{LLE} L \mathcal{X}_{LLE}^{\mathsf{T}}\right), \tag{3.8}$$

where $\mathcal{X}_{LLE} = \left[\boldsymbol{S}^{LLE}\left(c_1\right), \boldsymbol{S}^{LLE}\left(c_2\right), \ldots, \boldsymbol{S}^{LLE}\left(c_n\right)\right]$, $L = \left(I - W_{LLE}\right)\left(I - W_{LLE}^{\mathsf{T}}\right)$, and I is the identity matrix. The minimization of Equation 3.8 subject to the constraint $\mathcal{X}_{LLE}\mathcal{X}_{LLE}^{\mathsf{T}} = I$ (a normalization constraint that prevents the solution $\mathcal{X}_{LLE} \equiv \mathbf{0}$) is an Eigenvalue problem whose solutions are the Eigen vectors of the Laplacian matrix L. Since the rank of L is $n-1$, the first Eigen vector is ignored and the second smallest Eigen vector represents the best one-dimensional projection of all the samples. The best two-dimensional projection is given by the Eigen vectors with the second and third smallest eigenvalues, and so forth.

3.6 Consensus-Locally Linear Embedding (C-LLE)

3.6.1 Issues with LLE

The low dimensional data representations obtained via LLE are a function of κ, a parameter controlling the size of the neighborhood within which local linearity is assumed and used to approximate geodesic distances. The objective behind LLE is to non-linearly map objects $c, d \in C$ that are adjacent in the M dimensional ambient space ($\boldsymbol{F}(c)$, $\boldsymbol{F}(d)$) to adjacent locations in the low dimensional embedding $(S(c), S(d))$, where $(S(c), S(d))$ represent the m-dimensional dominant eigen vectors corresponding to c, d ($m << M$). If d is in the κ neighborhood of $c \in C$, then $c, d \in C$ are assumed to be linearly related. LLE attempts to non-linearly project each $\boldsymbol{F}(c)$ to $S(c)$ so that the κ neighborhood of $c \in C$ is preserved. LLE is sensitive to the choice of κ since different values of κ will result in different low dimensional data representations. Since LLE is typically used in an unsupervised context for visualizing and identifying object clusters, *a priori*, the optimal value of κ is not-obvious. In [91], Roweis and Saul suggest that varying κ over a wide range of values, still yields stable, consistent low dimensional embeddings for dense synthetic datasets. Our own experiments on real biomedical data, suggest otherwise [54]. Further, for sparsely populated datasets, the most common

failure of LLE is to map faraway points to adjacent locations in the embedding space depending on the choice of κ [91].

Automatically estimating κ is largely an open problem, though some researchers have attempted to adaptively determine a globally optimal κ value [62], [85]. However, these experiments were limited to dense synthetic datasets. We argue that in general no single global optimal value of κ can be applied to learning the low dimensional manifold over the entire data space. We further contend that different values of κ are required in different regions of the data space to optimally reconstruct locally linear neighborhood.

3.6.2 Relationship between C-LLE and Bagging Classifiers

The aim behind constructing ensemble classifiers such as Bagging [84] is to reduce the variance and bias across weak classifiers. In Bagging [84], for an input object $c \in C$, a sequence of weak predictors $\phi(c, \mathcal{S}_k)$ are generated from K bootstrapped training sets $\mathcal{S}_k$ where $1 \leq k \leq K$. A strong Bagged classifier $\phi^{Bag}(c)$ is obtained by averaging or voting over the multiple weak classifiers $\phi(c, \mathcal{S}_k)$, $k \in \{1, ...K\}$. An analogous idea is used for C-LLE whereby we combine several weak embeddings, $S_\kappa(c)$ across different values of $\kappa \in \{1, ...K\}$ to obtain a comprehensive stable low dimensional data embedding, with lower variance and bias compared to individual weak embeddings. Our hypothesis is that for any $c, d \in C$, the pairwise object distance in the low dimensional space is faithfully represented in the stable consensus embedding $\hat{S}(c)$, for $c \in C$.

3.6.3 Maximum Likelihood Estimation (MLE) of Object Adjacency

The spirit behind C-LLE is the direct determination of pairwise object adjacencies in the low dimensional embedding space as opposed to κ estimation. For each c, d the aim is to find the true distance $\hat{\mathcal{D}}^{\psi}(c, d)$ between $c, d \in C$ in some lower dimensional embedding space, where ψ is an appropriately defined distance metric. Given multiple lower dimensional embeddings, the distance between c, d can be expressed as a distribution $\mathcal{D}_\kappa(c, d)$ where for brevity the metric notation has been dropped. The problem of determining $\hat{\mathcal{D}}(c, d)$ can be posed as a MLE problem. Thus we can rewrite this problem

as,

$$p(\varphi_{\mathcal{D}}|\hat{\mathcal{D}}) = \prod_{\kappa=1}^{K} p(\mathcal{D}_\kappa|\hat{\mathcal{D}}); \forall c, d \in C, \tag{3.9}$$

where $\varphi_{\mathcal{D}}$ is a set of low dimensional distance estimates between $c, d \in C$, and based on the assumption that the lower dimensional embeddings obtained for $\kappa \in \{1, ...K\}$ are independent. We endeavor to find the MLE of $\mathcal{D}$, $\tilde{\mathcal{D}}$ that maximizes ln $p(\varphi_{\mathcal{D}}|\mathcal{D})$ for $c, d \in C$. Intuitively this corresponds to computing the peak (mode) of the distribution $p(\varphi_{\mathcal{D}}|\hat{\mathcal{D}})$.

3.6.4 Algorithm for C-LLE

Step 1. Generate Multiple lower dimensional embeddings

Multiple lower dimensional embeddings are generated by varying $\kappa \in \{1, ..K\}$ using LLE. Each embedding $S_\kappa(c)$ will hence represent adjacencies between objects $c_i, c_j \in C, i, j \in \{1, ...|C|\}$, where $|C|$ is the cardinality of C. Thus $||S_\kappa(c_i) - S_\kappa(c_j)||_\psi$ will vary as a function of κ.

Step 2. Obtain MLE of pairwise object adjacency:

A confusion matrix $W_\kappa^{LLE} \in \Re^{|C|\times|C|}$ representing the adjacency between any two objects $c_i, c_j \in C, i, j \in \{1, ..., |C|\}$ in the lower dimensional embedding representation $S_\kappa(c)$ is calculated as:

$$W_\kappa^{LLE}(i, j) = \mathcal{D}_\kappa^{LLE}(c_i, c_j) = \|S_\kappa(c_i) - S_\kappa(c_j)\|_\psi , \tag{3.10}$$

where $c_i, c_j \in C$, for $i, j \in \{1, \ldots, |C|\}, \kappa \in \{1, \ldots, K\}$, and ψ in our case is the L2 norm. MLE of $\mathcal{D}_\kappa^{LLE}(c_i, c_j)$ is estimated as the mode of all adjacency values in $W_\kappa^{LLE}(i, j)$ over all κ. This $\hat{\mathcal{D}}^{LLE}$ for all $c \in C$ is then used to obtain the new confusion matrix $\hat{W}^{LLE}$.

Step 3. Multidimensional scaling (MDS):

MDS [92] is applied to $\hat{W}^{LLE}$ to achieve the final combined embedding $\tilde{S}(c)$ for $c \in C$. MDS is implemented as a linear method that preserves the Euclidean geometry between each pair of objects $c_i, c_j \in C, i, j \in \{1, ..., |C|\}$. This is done by finding optimal

positions for the data points c_i, c_j in lower-dimensional space through minimization of the least squares error in the input pairwise distances in $\hat{W}^{LLE}$.

3.6.5 Semi-supervised Dimensionality Reduction

Most NLDR schemes are unsupervised and often lead to overlapping embeddings with poor class discriminability. Recently several semi-supervised DR (SSDR) schemes have been proposed where limited labeled information is employed in the construction of a pairwise similarity matrix. In [55], Sugiyama et al. applied semi supervised (SS)-learning to Fisher's discriminant analysis in order to find projections that maximize class separation. [56] implemented a SS version of PCA by exploiting between-class and within-class scatter matrices. Semi-supervised graph embedding [57] is another semi-supervised method based on GE which constructs a weight matrix (W) by leveraging the known class labels such that higher weights are given to within-class points and lower weights to points from different classes. The proximity of labeled and unlabeled data is then used to construct the low dimensional manifold. Assuming the first l of n samples are labeled $\omega_l \in \{0, 1\}$, we can incorporate the partial known labels into the similarity matrix $W = [w_{ij}]$. A $\mathcal{N}$ nearest neighbor graph, $\mathcal{N} > 0$, is created to obtain W such that pairwise points in $\mathcal{N}$ neighborhood with same labels are given high weights and points with different class labels are given low weights [57]. If the points are not in $\mathcal{N}$, the corresponding edges are not connected. Thus the weight matrix may be expressed as,

$$\tilde{w}_{ij} = \begin{cases} \gamma(1+\gamma), & \text{if } c_i \in \mathcal{N}_j \text{ or } c_j \in \mathcal{N}_i \text{ and } \omega_i = \omega_j, \\ \gamma(1-\gamma), & \text{if } c_i \in \mathcal{N}_j \text{ or } c_j \in \mathcal{N}_i \text{ and } \omega_i \neq \omega_j, \\ \gamma, & \text{if } c_i \in \mathcal{N}_j \text{ or } c_j \in \mathcal{N}_i,\ i > l \text{ or } j > l, \\ 0, & \text{otherwise.} \end{cases} \tag{3.11}$$

where $\gamma = e^{\frac{-||F(c)_i - F(c)_j||^2}{\sigma}}$.

3.7 Texture Feature extraction

The textural operators Φ_β, $\beta \in \{1, \ldots, n\}$ include both linear and non-linear image operations, and are drawn from 3 general classes of texture features including (i) gradient, (ii) first order statistical, and (iii) second order statistical texture features. Table 3.1 details the texture features that were employed in this work.

3.8 Classifier Strategies

3.8.1 Random Forest (RF) Classifier

The RF classifier uses the majority voting rule for class assignment by combining decisions from an ensemble of bagged (bootstrapped aggregated) [97] decision trees. The C4.5 decision tree [98] is a multistage classifier which creates a tree like structure by breaking down a complex decision process into a collection of simpler decisions for predicting the best possible outcome solution by combining the simple decisions. RF further combines these decisions obtained from different decision trees to provide a more optimal and stable solution. For a given training set, N bootstrapped subsets are created with replacement of the training data. Based on each training subset, a C4.5 decision tree [98] classifier , $\mathbf{h}_j, j \in \{1, ..., N\}$ is constructed. The class label (CaP or normal) $\mathbf{h}_j(c)$ for each metavoxel $c \in C$, based on the feature vector $\boldsymbol{F}_{PCA}^{T2w}(c)$, is then obtained using the decision trees $\mathbf{h}_j, j \in \{1, ..., N\}$; $\mathbf{h}_j(c) = 1$ if c is classified as CaP (scale 4, 5) and otherwise $\mathbf{h}_j(c) = 0$. The final class likelihood that c belongs to the CaP class, via the RF classifier, is obtained by aggregating the decisions of individual weak learners as $\frac{1}{N}\sum_{j=1}^{N} \mathbf{h}_j(c)$. The higher the value of this class likelihood, the more likely c belongs to the CaP class.

3.8.2 Probabilistic Boosting Tree (PBT) classifier

The PBT algorithm [99] is a combination of the decision tree [98] and Adaboost [100] classifiers. Adaboost is an ensemble classifier obtained by combining classifier predictions from several weak classifiers. PBT combines decision tree and Adaboost by

iteratively generating a tree structure of a predefined length in the training stage where each node of the tree is boosted with weak classifiers. The hierarchical tree is obtained by dividing the training samples in two left and right subsets and recursively training the left and right sub-trees using Adaboost [100]. During testing, the conditional probability that any object c belongs to the target class, given the feature vector, $\boldsymbol{F}(c)$, is calculated at each node based on the learned hierarchical tree.

3.8.3 Support Vector Machines (SVM)

SVM aims at identifying the best possible hyper plane that can accurately separate the data into two classes. SVM classifier [101] is constructed by using a kernel function which projects the training data into a higher-dimensional space via an implicit feature mapping in the dot product space. In our implementation, the radial basis function (RBF) kernel was employed to project the training data into a higher dimensional space. In contrast to PBTs and RFs where a probability (or likelihood) is generated for each voxel belonging to a class, SVM classifiers are typically used to generate a hard decision; $\mathbf{h}(c) = 1$ if metavoxel c is identified as CaP and $\mathbf{h}(c) = 0$ otherwise. However, a pseudo-likelihood that any meta-voxel c belongs to a class can be generated by calculating how far or close each c is from the SVM decision hyperplane during classification and converting this distance in terms of likelihood of each c belonging to a class. Thus the greater the distance of from the hyperplane, the higher the likelihood that it belongs to a particular class; the proximity of an object to the hyperplane reflects greater ambiguity with respect to class membership.

Feature	Implementation	Purpose	Significance for quantifying CaP appearance
Gabor wavelet transform (48)	Modulation of a complex sinusoid by a Gaussian function	Attempt to match localized frequency characteristics at multiple scales and orientations[93]	Quantify visual processing features used by radiologists when examining appearance of the carcinoma
Haar wavelet transform (12)	Decomposition coefficients via wavelet decomposition at multiple scales	Attempt decomposition of a signal in the discrete space while offering localization in the time and frequency domains[94]	Differentiate the amorphous nature of the non-CaP regions within foci of low SI
Haralick texture feature (36)	Construct joint probability distribution of the occurrence of greylevel intensities in an image (spatial relationship between pixels used to restrict counting of greylevel co-occurrences). Statistical features are then calculated from this distribution	Differentiate between different types of texture excellently due to calculation of 2nd order statistics (which quantify perceptual appearance of image)[95]	Useful in differentiating homogeneous low SI regions (CaP) from more hyper-intense appearance of normal prostate
Greylevel statistical features (14)	Mean, standard deviation as well as derivative features such as via convolution with the Sobel and Kirsch operators are calculated	Provide 1st order information, quantifying macroscopic appearance of image e.g. variation of intensities within image[96] etc.	May help localize regions of significant differences on T2w MR image, accurately detect region boundaries

Table 3.1: Summary of texture features used in this study as well as their significance for localization of CaP on T2w MRI (numbers in brackets signify how many features of each texture category were computed)

Chapter 4

Data Integration Strategies

4.1 Multi-wavelet Embedding Representation for Data Combination (MaWERiC)

In [102], it has been suggested that data level integration can be achieved by aggregating features from two disparate sources ($\boldsymbol{F}_1(c)$ and $\boldsymbol{F}_2(c)$) into a single feature vector $\boldsymbol{E}(c)$ before classification. While this may be a reasonable strategy when $|\boldsymbol{F}_1(c)|=|\boldsymbol{F}_2(c)|$, it may not be optimal when the feature vectors are of very different dimensionalities. In this work, we have developed a novel data integration strategy, called MaWERiC, that incorporates wavelet transformation and dimensionality reduction to represent different data streams into a common framework, that is conducive for data integration. Figure 4.1 illustrates the organization of our MaWERiC strategy. Below we describe each of the modules associated with our MaWERiC strategy.

1. Feature Extraction: Our scheme is based on combining wavelet (Gabor and Haar filters) features [103] extracted from both the T2-w MRI and MRS modalities. The advantage of using wavelet transform [103] is that it can provide multi-resolution discriminatory information from different data modalities, including but not limited to signals and images [67, 104, 105]. The 2D-Gabor wavelet filter is defined as the convolution of a 2D Gaussian function with a sinusoid [106]. A Gabor filter bank is then generated by variation of the associated scale and orientation parameters. This filter bank provides a means for multi-scale, multi-orientation texture characterization and representation of an image. Haar wavelet decomposition is a commonly employed signal filtering technique [103]

which provides a way of extracting the class discriminating frequency components that may yield higher classifier accuracy compared to the original signal [107]. An advantage of the Haar wavelet [107] is that it preserves features that are representative of abrupt changes in signals, dominant spectral peaks such as those that correspond to the most significant metabolites on MRS, while simultaneously eliminating spectral noise. Both the Gabor and Haar wavelet filters have been previously used in conjunction with CDS classifiers to distinguish between different data classes for various biomedical applications [67, 104, 105, 108]. In the context of this work, multi-resolution features for T2-w MRI and MRS are obtained via use of the Gabor and Haar wavelet filters respectively.

2. Data Representation: While the wavelet based representations of the MRS, T2-w MRI channels provide a uniform, feature representation of the data, the feature vectors obtained via the application of the Gabor and Haar wavelet filters are of a very high and dimensionality and hence subject to the curse of dimensionality [82]. Consequently a subsequent step is application of PCA [74] to the high dimensional feature vectors obtained (via wavelet decomposition) from each individual data modality (T2-w MRI, MRS) to obtain a reduced dimensional representation of the data and thus make the data representation amenable to the application of the classifier. The representation of the MRS and T2-w MRI data in terms of the Eigen vectors (obtained via PCA) also allows for the overcoming of the scale (resolution) and dimensionality differences between the two modalities, since the wavelet representations obtained from each individual modality are of varying dimensionality, and can be reduced to the same number of Eigen vectors. MaW-ERiC enables data fusion between different dimensional feature vectors $\boldsymbol{F}_1(c)$, $\boldsymbol{F}_2(c)$ by first decomposing them in a wavelet space, followed by independently reducing them to a low dimensional space ($|\boldsymbol{F}^1_{PCA}(c)| = |\boldsymbol{F}^2_{PCA}(c)|$).

3. Data Integration: The reduced low dimensional representations across the two protocols, $\boldsymbol{F}^1_{PCA}(c), \boldsymbol{F}^2_{PCA}(c)$, are then concatenated to obtain the fused Eigen vector representation as, $\boldsymbol{F}^{Int} = [\ \boldsymbol{F}^1_{PCA}(c), \boldsymbol{F}^2_{PCA}(c)]$.

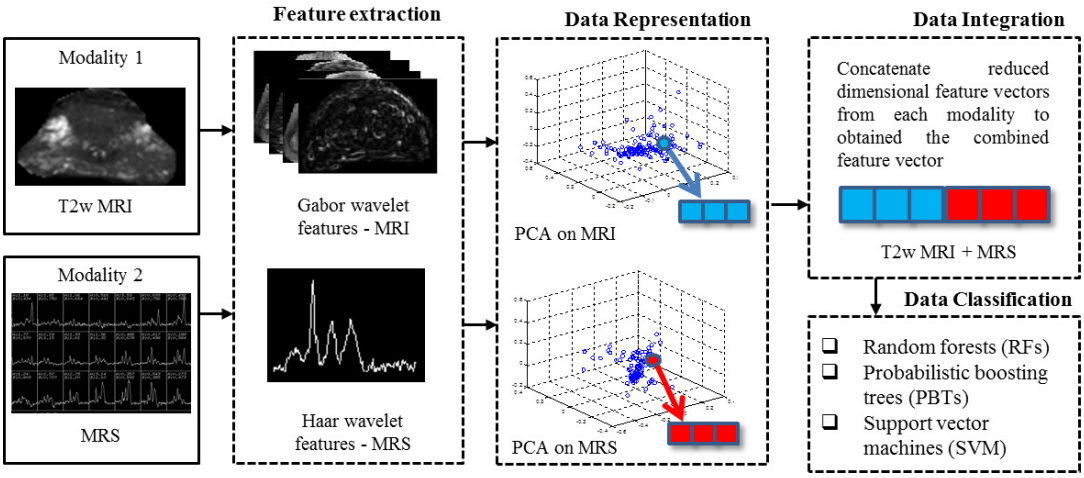


Figure 4.1: Flowchart showing various components and methodological overview of our data integration scheme. Wavelet features are first extracted individually from T2-w MRI and MRS, followed by dimensionality reduction using PCA. The reduced dimensional vectors, now existing in the same dimensions and scale, are concatenated during data combination followed by data classification.

4. Data Classification: This fused Eigen vector representation is then used to train a Random forest (RF) classifier. RF is a commonly used ensemble classifier that combines predictions from several weak classifiers to generate a more accurate and stable classifier [109]. The RF classifier has previously been successfully employed for various biomedical classification applications [67, 110, 71]. Advantages of RF include, (1) ability to integrate a large number of input variables, (2) robustness to noise in the data, and (3) relatively few tuning-parameters.

4.2 Semi-Supervised Multi-kernel Graph Embedding (SeSMiK)

4.2.1 Kernel Graph Embedding (KGE) Framework

KGE is a technique to extend linear projections of data to a non-linear dot product space using the kernel trick [111], which maps data from the original input space to an alternative higher dimensional space as $K(\mathbf{F}(c_i), \mathbf{F}(c_j)) = \langle(\Phi(\mathbf{F}(c_i)), \Phi(\mathbf{F}(c_j)))\rangle$, where Φ is the implicit pairwise embedding between $\mathbf{F}(c_i)$ and $\mathbf{F}(c_j)$. A kernel gram matrix K_m for each protocol m may be obtained as $K_m = [K(\mathbf{F}(c_i), \mathbf{F}(c_j))]$, $\forall i, j \in \{1, ..., n\}$,

where K_m may be expressed as

$$K_m = \begin{bmatrix} K(\mathbf{F}(c_1),\mathbf{F}(c_1)) & K(\mathbf{F}(c_1),\mathbf{F}(c_2)) & \dots & K(\mathbf{F}(c_1),\mathbf{F}(c_n)) \\ K(\mathbf{F}(c_2),\mathbf{F}(c_1)) & K(\mathbf{F}(c_2),\mathbf{F}(c_2)) & \dots & K(\mathbf{F}(c_2),\mathbf{F}(c_n)) \\ \vdots & \vdots & \ddots & \vdots \\ K(\mathbf{F}(c_n),\mathbf{F}(c_1)) & K(\mathbf{F}(c_n),\mathbf{F}(c_1)) & \dots & K(\mathbf{F}(c_n),\mathbf{F}(c_n)) \end{bmatrix}_m . \tag{4.1}$$

According to the Representer Theorem [112], to calculate the kernel representation $K(\mathbf{F}(c_i), \mathbf{F}(c_j))$ of input data, it is assumed that the optimal embedding $\mathbf{y}$ lies in the input space such that,

$$\mathbf{y} = \sum_{j=1}^{n} \alpha_j K(\mathcal{F}(.), \mathbf{F}(c_j)) = K(\mathcal{F})^T \boldsymbol{\alpha}, \tag{4.2}$$

where, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, ..., \alpha_n]^T$ is the low dimensional matrix representation for KGE and $K(\mathcal{F}) = [K(\mathcal{F}(.), \mathbf{F}(c_1)), K(\mathcal{F}(.), \mathbf{F}(c_2)), ..., K(\mathcal{F}(.), \mathbf{F}(c_n))]_m$. Hence,

$$\begin{aligned} \boldsymbol{y} &= [y_1, y_2, ..y_n]^T \\ &= \begin{bmatrix} K(\mathcal{F}_1) \\ \vdots \\ K(\mathcal{F}_n) \end{bmatrix}_m \boldsymbol{\alpha} \\ &= \begin{bmatrix} K(\mathbf{F}(c_1),\mathbf{F}(c_1)) & \dots & \dots & K(\mathbf{F}(c_1),\mathbf{F}(c_n)) \\ K(\mathbf{F}(c_2),\mathbf{F}(c_1)) & \dots & \dots & K(\mathbf{F}(c_2),\mathbf{F}(c_n)) \\ \vdots & \vdots & \ddots & \vdots \\ K(\mathbf{F}(c_n),\mathbf{F}(c_1)) & \dots & \dots & K(\mathbf{F}(c_n),\mathbf{F}(c_n)) \end{bmatrix}_m \boldsymbol{\alpha} \\ &= K_m \boldsymbol{\alpha} \end{aligned} \tag{4.3}$$

where K_m is the kernel matrix $K_{ij} = K(\mathbf{F}(c_i), \mathbf{F}(c_j))$. Using Equation (4.3), the objective function can be reduced to,

$$\boldsymbol{\alpha}^* = \arg\max_{\alpha} [\boldsymbol{\alpha}^T K \mathcal{D} K \boldsymbol{\alpha}], \tag{4.4}$$

where K is a valid positive semi-definite kernel and $\boldsymbol{\alpha}$ is the d dimensional Eigenvector of the objective kernel function in Equation (4.4). Similar to Equation (3.5.2), optimization function can again be solved by an Eigenvalue decomposition problem as,

$$KWK\boldsymbol{\alpha} = \lambda K \mathcal{D} K \boldsymbol{\alpha}, \tag{4.5}$$

with the constraint $\boldsymbol{\alpha}^T K \mathcal{D} K \boldsymbol{\alpha} = 1$.

4.2.2 Semi-Supervised Multi-Kernel Graph Embedding (SeSMiK-GE Algorithm)

Overview of SeSMiK-GE

Figure 4.2 shows a schematic flowchart of the proposed data integration strategy, SeSMiK-GE. Below we briefly describe each of the modules comprising SeSMiK-GE.

1. Module 1: Multi Kernel Learning: Due to the dimensionality and resolution differences between different data channels, each source needs to first be represented in a common framework prior to data integration. MKL is employed to transform each of m individual data channel to a kernel similarity matrix (K_m), in order to then derive a weighted combination of individual kernels as $\hat{K} = \sum_{m=1}^{M} \beta_m K_m$, where $K_m, m \in \{1, 2, ..., M\}$, is the kernel obtained from each data channel, and β_m is the weight assigned to each kernel.

2. Module 2: Semi-Supervised Learning: Employing SSDR in a DSS scheme has benefits in that, (a) SSDR is computationally inexpensive, and (b) yields a low dimensional representation with better discriminability between object classes with fewer training samples. SeSMiK-GE employs SSGE [57], a well known SSDR scheme, which modifies the similarity weight matrix ($\tilde{W}_m$) for each data channel, by incorporating partial label information. A modified weight matrix ($\tilde{W}_m$) is obtained from each of the M data channels, which are then averaged to obtain, $\hat{W} = \frac{1}{M} \sum_{i=1}^{M} \tilde{W}$.

3. Module 3: Dimensionality Reduction: $\hat{K}$ and $\hat{W}$ are employed in a generalized KGE framework to obtain the integrated low dimensional data representation. It is worth noting that although other NLDR schemes such as LLE [50] and Isomaps [51] are also popular for DR purposes, GE was our method of choice for SeSMiK framework since unlike LLE [50] and Isomaps [51], GE is not dependent on kappa (the parameter determining the size of the local neighborhood within which linearity of the manifold is assumed) which is known to significantly affect the quality of the low dimensional manifold [113, 85].

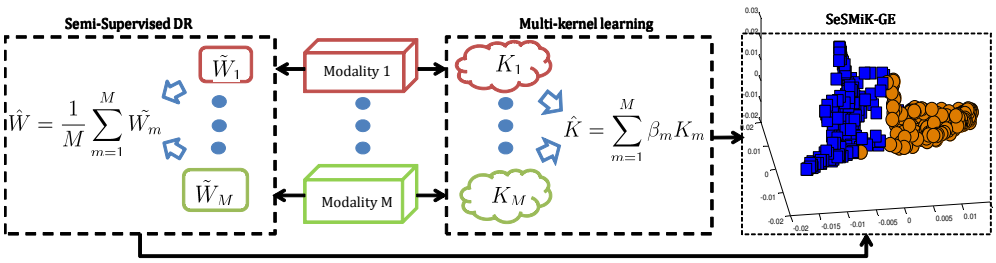


Figure 4.2: Flowchart showing various components of SeSMiK-GE. MKL and SSDR are performed simultaneously on the M individual data channels followed by DR on the combined kernel and weight matrix. A supervised classifier is subsequently trained in the integrated low dimensional space to discriminate the object classes (shown via different colors in the right most panel).

Module 1. Multi-Kernel Learning

A linear combination of different kernels has the advantage of also yielding a kernel which is at once a symmetric, positive definite matrix. Assuming we have M base kernel functions for M channels, K_m, $m \in \{1, ..., M\}$, with corresponding individual kernel weights β_m, the combined kernel function may be expressed as,

$$\begin{aligned} \hat{K}(\mathbf{F}(c_i), \mathbf{F}(c_j)) &= \sum_{m=1}^{M} \beta_m K_m(\mathbf{F}(c_i), \mathbf{F}(c_j)), \beta_m \geq 0, \forall i, j \in \{1, ..., n\}, \\ &= \sum_{m=1}^{M} \beta_m K_m, \beta_m \geq 0. \end{aligned} \tag{4.6}$$

$\hat{K}$ is the combined multi-channel kernel obtained by combining M protocols in a multi-kernel framework.

Module 2. Semi-supervised Graph Embedding (SSGE)

Equation 3.11 as described in Chapter 3 is used to obtain the Gaussian weight matrix $\tilde{W}_m = [\tilde{w}_{ij}]$. $\tilde{W}_m = [\tilde{w}_{ij}]$ is then normalized by σ such that $\sigma = \max(||F(c_i) - F(c_j)||^2)$ $\forall i, j$ for each individual data channel m, $m \in \{1, ..., M\}$. Hence, the range of normalized weight matrix, $\tilde{W}_m$, is between $e^{-1} = 0.333$ and $e^0 = 1$, which is subsequently scaled linearly between 0 and 1. Weight matrices across individual data channels, $\tilde{W}_m$, $m \in \{1, ..., M\}$ can then be averaged to obtain,

$$\hat{W} = \frac{1}{M} \sum_{m=1}^{M} \tilde{W}_m \tag{4.7}$$

where $\tilde{W}_m$ is the modified weight matrix obtained using Equation (3.11) for protocol m, and $\hat{W}$ is the combined weight matrix obtained by averaging the modified weight matrices from each of the M data channels.

Module 3. Dimensionality Reduction

The combined kernel $\hat{K}$ obtained from MKL, and the associated weight matrix $\hat{W}$ obtained from SSL can be used in a KGE framework to obtain the low dimensional fused representation of the multi-channel data. By substituting $\hat{K}$ from Equation (4.6) and $\hat{W}$ from Equation (4.7), Equation (4.5) is reduced to a multi-kernel Eigenvalue decomposition problem as,

$$\hat{K}\hat{W}\hat{K}^T\bar{\boldsymbol{\alpha}} = \lambda\hat{K}\hat{\mathcal{D}}\hat{K}^T\bar{\boldsymbol{\alpha}}, \tag{4.8}$$

where $\hat{\mathcal{D}} = \sum_j \hat{w}_{ji}$, and $\bar{\boldsymbol{\alpha}}$ is the combined low dimensional fused data representation obtained by combining M different channels.

4.2.3 SeSMiK-GE Optimization

In Equation 4.8, two variables, $\bar{\boldsymbol{\alpha}}$ and β_m (within $\hat{K}$) need to be optimized simultaneously. The low dimensional representation $\bar{\boldsymbol{\alpha}}$ is optimized using a kernel ridge regression function, while the weights β_m are optimized using a hierarchical brute force algorithm. Each individual optimization step is explained below.

Optimizing the low dimensional representation ($\bar{\boldsymbol{\alpha}}$)

The optimal d dimensional Eigenvectors $\bar{\boldsymbol{\alpha}} = \{\bar{\alpha}_1, ..., \bar{\alpha}_d\}$, $d << D$ are obtained from Equation (4.8) using standard kernel ridge regression optimization as described in [114]. Kernel ridge regression is a regularized least square linear regression in kernel space, and is used when the matrix $\hat{K}$ is invertible (*ill-conditioned*), or noisy to obtain the target output accurately. Hence, to solve $\bar{\boldsymbol{\alpha}}$, we make use of the regularized solution of Equation 4.3. The optimization of Equation 4.8 can then be solved via a two step process:

1.1) Solve the Eigenvalue decomposition problem as given in Equation 3.5.2 for the combined data matrix $\hat{W}\boldsymbol{y} = \lambda\hat{\mathcal{D}}\boldsymbol{y}$.

1.2) If $\hat{K}$ is non-singular for any given $\boldsymbol{y}$, unique Eigenvalues can be obtained as $\bar{\boldsymbol{\alpha}} = \hat{K}^{-1}\boldsymbol{y}$ (Equation 4.3). However, when $\hat{K}$ is singular, the equation may have no or infinite solutions. The solution is then obtained using regularized kernel ridge regression as: $\bar{\boldsymbol{\alpha}} = (\hat{K} + \delta I)^{-1}\boldsymbol{y}$, where I is the identity matrix and δ is the regularization parameter. In this work we used the regularization parameter, $\delta = 0.1$, as suggested in [114].

Optimizing weights (β) for MKL

To obtain $\bar{\boldsymbol{\alpha}}$, optimal set of weights $\hat{\boldsymbol{\beta}} = [\hat{\beta}_1, ..., \hat{\beta}_M]$, $\hat{\beta_m} \in \{0, 1\}$, have to be obtained for each modality m, $m \in \{1, ...M\}$ such that $\sum_{m=1}^{M} \hat{\beta}_m = 1$. A hierarchical brute force optimization strategy is employed to optimize weights, $\boldsymbol{\beta} = [\beta_1, ..., \beta_M]$, which iteratively optimizes $\boldsymbol{\beta}$ based on the classification accuracy of training data. Once initial values for $\hat{\boldsymbol{\beta}}$ $\forall m, m \in \{1, ..., M\}$, are estimated at a certain interval resolution that optimizes accuracy, the algorithm searches for a more accurate value only within the vicinity of $\hat{\boldsymbol{\beta}}$ estimated at the previous level of the hierarchy. At each level of the optimization strategy, the value of $\hat{\boldsymbol{\beta}}$ is estimated based on the ability of weights to create a low dimensional representation that maximizes classifier accuracy. The process is repeated until either a pre-defined interval resolution is reached or classification accuracy does not change significantly by reducing the step size.

4.2.4 SeSMiK-GE algorithm

Algorithm *SeSMiK-GE*

Input: $\mathcal{F}_m$, M, $\mathcal{N}$, d, $\hat{\boldsymbol{\beta}} = [\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_M]$

Output: $\bar{\boldsymbol{\alpha}}$

begin

0. *for* $m = 1 : M$
1. Obtain K_m for each data channel $\mathcal{F}_m$
2. Obtain $\tilde{W}_m$ using $\mathcal{N}$ from Equation (3.11)

3. *endfor*
4. Obtain $\hat{W}$ as $\hat{W} = \frac{1}{M}\sum_{m=1}^{M}\tilde{W}_m$ using each of the $\tilde{W}_m$ obtained in Step 2
5. Obtain $\hat{K} = \sum_{m=1}^{M}\hat{\beta}_m K_m$ using each of the K_m obtained in Step 1
6. Substitute $\hat{K}$ and $\hat{W}$ in Equation 4.5
7. Obtain d-dimensional $\bar{\boldsymbol{\alpha}}$ by solving Equation 4.8

end

K_m and $\tilde{W}_m$ are obtained for each of the data channels $\mathcal{F}_m$, $m \in \{1, ..., M\}$, in Steps 1 and 2 respectively. Combined weight matrix $\hat{W}$ is obtained by averaging the modified weight matrices (obtained in Step 2) across the M data channels (Step 4). Similarly, in Step 5, the contributions of each K_m are individually weighted using the optimal weights $\hat{\beta}_m$ across the M data channels. $\hat{W}$ and $\hat{K}$ are then substituted in the generalized KGE framework in Step 6 which is solved in Step 7 to obtain the d-dimensional fused representation of the multi-modal data. A supervised classifier can then be trained on the fused low dimensional representation $\bar{\boldsymbol{\alpha}}$ for subsequent object classification.

Chapter 5

Consensus Locally-Linear Embedding: Application to Prostate Cancer Diagnosis

5.1 Prostate cancer detection on MRS via C-LLE

5.1.1 Data Description

A total of 18 1.5 T *in vivo* endorectal T2-weighted MRI and MRS ACRIN studies[1] were obtained prior to prostatectomy. Partial ground truth for the CaP extent on MR studies is available in the form of approximate sextant locations and sizes for each study. The maximum diameter of the tumor is also recorded in each of the 6 prostate sextants (left base, left midgland, left apex, right base, right midgland, right apex). The tumor size and sextant locations were used to identify a *potential cancer space* used for performing a semi-quantitative evaluation of our CAD scheme. Additional details on identifying this cancer space are provided in [73].

5.1.2 C-LLE and consensus clustering for CaP detection on MRS

Figure 5.1 shows the flowchart demonstrating the different steps comprising our prostate MRS detection scheme. The adjacency matrix (W_{LLE}) is constructed across K embeddings of the MR spectral space so that for any $c_i, c_j \in C$, where $i, j \in \{1, ...|C|\}$, $W_\kappa(i,j)$ represents the distribution $\mathcal{D}_\kappa(c_i, c_j)$ of the low dimensional distance between MR spectra $\boldsymbol{F}(c_i), \boldsymbol{F}(c_j)$, for $k \in \{1, ...K\}$. As described in Chapter 3, the stable spectral distance matrix $\hat{W}_{LLE}$ is then obtained and MDS applied to obtain the stable embedding representation of the spectra, $\tilde{S}(c)$ for each $c \in C$.

To overcome the instability associated with centroid based clustering algorithms, we

[1] http : //www.acrin.org/6659 protocol.html

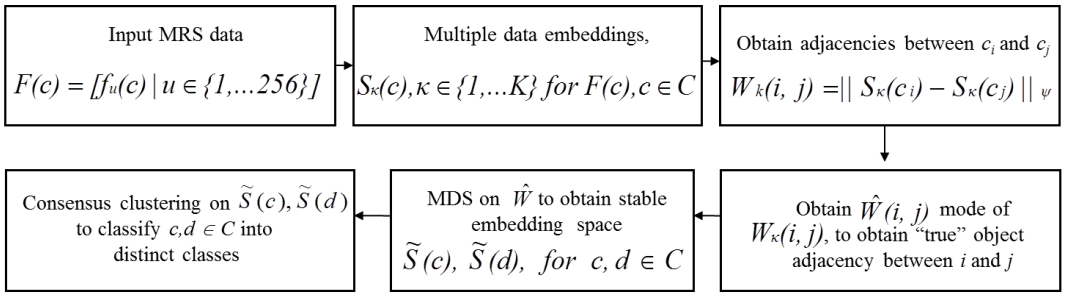


Figure 5.1: Flowchart showing the C-LLE algorithm and consensus clustering scheme for CaP detection on prostate MRS.

generate multiple weak clusterings $V_t^1, V_t^2, V_t^3, t \in \{0, \ldots, T\}$ by repeated application of k-means clustering on the combined low dimensional manifold $\tilde{S}(c)$, for all $c \in C$. We assume that each prostate spectra could be classified as one of the three classes: cancer, benign and other tissue classes (e.g. benign hyperplasia (BPH)). Each cluster, V_t is a set of objects which has been assigned the same class label by the k-means clustering algorithm. As the number of elements in each cluster tends to change for each such iteration of k-means, we calculate a co-association matrix H with the underlying assumption that objects belonging to a *natural* cluster are very likely to be co-located in the same cluster for each iteration. Co-occurrences of pairs of objects $c_i, c_j \in C$ in the same cluster V_t are hence taken as votes for their *association*. $H(i, j)$ thus represents the number of times $c_i, c_j \in C$, for $i, j \in \{1, ...|C|\}$, were found in the same cluster over T iterations. We apply MDS [92] to H followed by a final unsupervised classification using k-means, to obtain the final stable clusters $\hat{V}^1, \hat{V}^2, \hat{V}^3$.

5.1.3 Model based peak integration scheme via Independent Component Analysis (ICA)

Peak detection on prostate MRS is a difficult problem due to noise and spectral contributions from extra-prostatic regions. In this work, we have developed a model based approach to localize choline, creatine and citrate peaks based on Independent Component Analysis (ICA). ICA is a multivariate decomposition technique which linearly transforms the observed data into statistically maximally independent components (ICs). For a set of voxels identified offline as cancer, $\chi_{CaP} \subset C$, we obtain A independent

components $\mathcal{F}_{\alpha}^{IC}, \alpha \in \{1, ...A\}$ which represent spectral contributions of choline, creatine and citrate for prostate cancer. The parts per million (ppm) ranges (ν_{cc}, ν_{cr}) on the X-axis are then learnt for choline+creatine and citrate from $\mathcal{F}_{\alpha}^{IC}, \alpha \in \{1, ...A\}$. Peak detection is then performed on $\mathcal{C}$ to identify choline, creatine and citrate peaks within the ranges ν_{cc} and ν_{cr}. Area under the choline+creatine peak (A_{ch+cr}) and under citrate peak (A_{cit}) is obtained via integration for all voxels $c \in \mathcal{C}$ as described in Chapter 3, and a ratio of $A_{ch+cr}(c)/A_{cit}(c)$ is computed. A pre-defined threshold determined by radiologists [115] is used to classify the spectra as cancer/benign based on $A_{ch+cr}(c)/A_{cit}(c)$ for $c \in C$.

5.1.4 z-score and PCA

Similar to our C-LLE scheme, each $c \in C$ is described by 5 principal components, $S^{PCA}(c)$ which contain 98% of the data variance. Consensus clustering is then applied on $S^{PCA}(c)$, to cluster each $c \in C$ into one of 3 classes. A z-score as described in Chapter 3 at every $c \in C$ is also computed.

5.2 Results and Discussion

5.2.1 Qualitative Results

Figure 5.2 shows CaP detection results on a prostate MRS study via C-LLE, LLE, PCA, z-score, and ICA based peak detection. Figures 5.2(a)-(c) show the comparison of C-LLE (c) with LLE for (a) $\kappa = 5$, and (b) $\kappa = 7$ on a single 2D T2 weighted MRI slice. Figures 5.2(d), (e), (f) show the clustering results obtained via ICA peak detection, z-score and PCA respectively. At each spatial location on the MRI slice a spectral signature was analyzed and that corresponding location was assigned one of 3 colors (for C-LLE, PCA) and one of the two colors (for z-score, ICA based peak detection) based on clustering/classifier results. The white box superposed on 5.2(a)-(f) show the potential cancer space for corresponding slices. In each of Figures 5.2(a)-(f) the red cluster represents the one identified as cancer by each of the different methods. Note in Figure 5.2(c) that the C-LLE result shows excellent sensitivity and specificity and

also appears to reduce the variance and instability in the individual weak embeddings shown in 5.2(a), (b). Figures 5.2(d)-(f) corresponding to results obtained via ICA-based peak detection, z-score method and PCA respectively, show low sensitivity and specificity compared to our C-LLE scheme (Figure 5.2(c)). To assess the validity of C-LLE and PCA, we employed ICA to isolate independent components (IC) from clusters identified as CaP by the two schemes. Figure 5.2(h) shows an IC obtained from the cluster identified as CaP by C-LLE (shown as red in Figure 5.2(c)); Figure 5.2(i) shows the corresponding result obtained via PCA. Note the strong correlation between the ICs obtained via C-LLE (Figure 5.2(g)) and a known CaP MRS signature (Figure 5.2(h)) according to the 5-point scale defined in [116]. Note also the dissimilarity between the spectra obtained by PCA and shown in Figure 5.2(i) compared to those in Figure 5.2(g),(h).

5.2.2 Quantitative Results

Table 5.1 shows average CaP detection sensitivity and specificity over 18 studies obtained from C-LLE ($m = 4$), ICA based peak detection, PCA and z-score. Table 5.1 (b) shows the sensitivity and specificity results averaged across 18 datasets for C-LLE ($m = 3, 4, 5$) compared to LLE by varying the number of dimensions. Note that the C-LLE scheme has a higher sensitivity and specificity across all dimensions which suggests the efficacy of the scheme. The effectiveness of our scheme for detection of prostate cancer is evident from the quantitative results (Table 5.1) with both sensitivity and specificity of close to 87% and 85% respectively compared to current state of the art methods peak detection, PCA and z-score. Table 5.1(b) reveals that C-LLE consistently outperforms traditional LLE across multiple dimensions ($m = 3, 4, 5$).

Method	Sensitivity	Specificity
C-LLE	86.90	85.14
Peak detection	45.29	76.62
PCA	66.95	77.89
z-score	74.74	49.75

(a)

m	Sensitivity		Specificity	
	C-LLE	LLE	C-LLE	LLE
3	83.20	82.07	84.81	81.89
4	86.90	83.38	85.14	81.77
5	84.88	82.10	85.60	81.70

(b)

Table 5.1: (a) Average CaP detection Sensitivity and Specificity of C-LLE ($m = 4$), compared to ICA based peak detection, z-score, and PCA, averaged over a total of 18 MRS studies using the top 7 eigen values. Table 5.1(b). Average CaP detection Sensitivity and Specificity results of C-LLE compared to LLE for dimensions 3, 4 and 5.

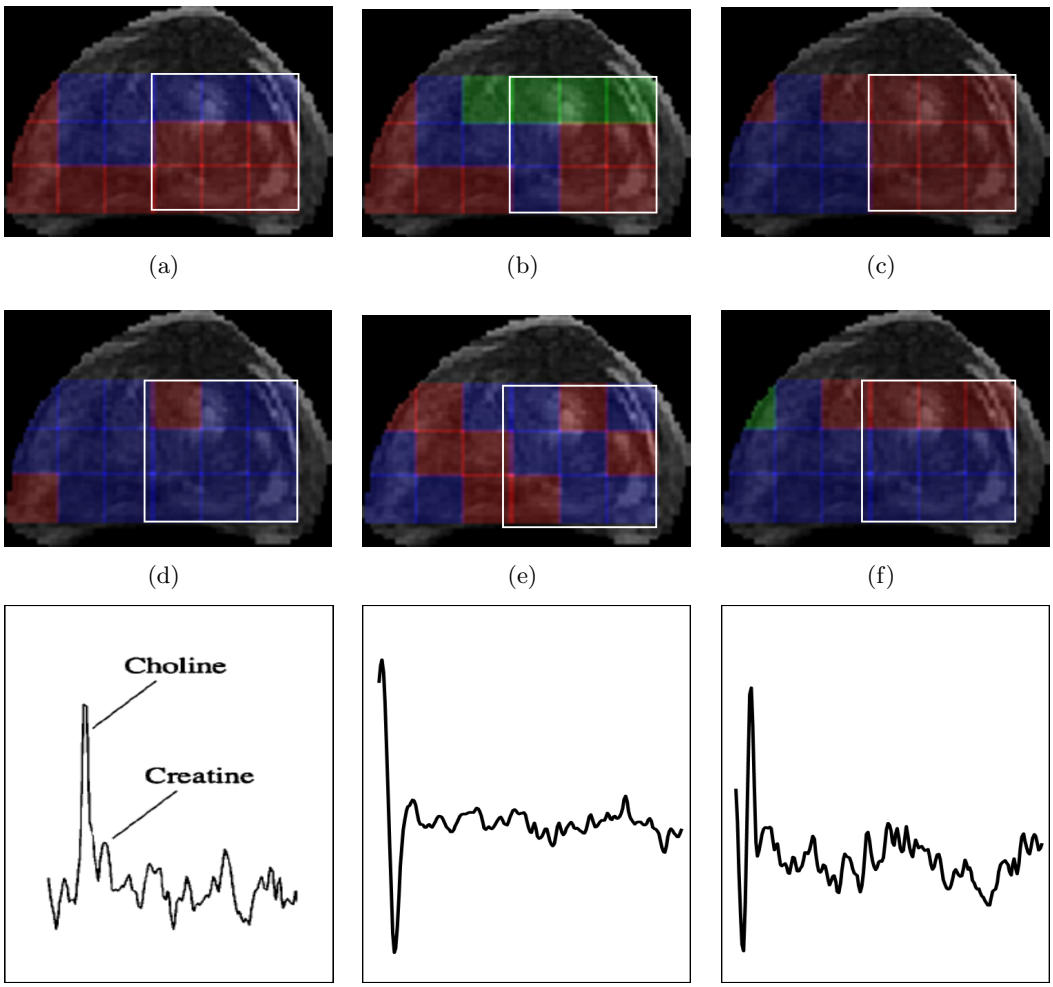


Figure 5.2: Clustering results for (a) LLE ($\kappa = 5$), (b) LLE ($\kappa = 7$), (c) C-LLE, (d) ICA based peak detection, (e) z-score, and (f) PCA. The white box superposed on the T2 MR image corresponds to the ground truth region. In each image the red cluster corresponds to the locations identified as cancer by each of the methods. Fig 5.2(g) Typical CaP MR spectra, (h) IC obtained from clusters identified as CaP in C-LLE, and (i) corresponding IC obtained via PCA.

Chapter 6

Integrating Magnetic Resonance (MR) Imaging (MRI) and MR Spectroscopy: Application to Prostate Cancer Diagnosis

6.1 Materials and Methods

6.1.1 Data Description

A total of 36 1.5 Tesla (T) T2-w MRI, MRS studies were obtained prior to radical prostatectomy. All of these studies were biopsy proven prostate cancer patient studies that were clinically referred for a prostate cancer MR staging exam for improved therapeutic selection. MR imaging was performed by using a 1.5-T whole-body MR imaging unit (Signa; GE Medical Systems, Milwaukee, Wisconsin). The patients were imaged while in the supine position by using a body coil for signal excitation and a pelvic phased-array coil (GE Medical Systems) combined with a balloon-covered expandable endorectal coil (Medrad, Pittsburgh, PA) for signal reception. All MR images were routinely post-processed to compensate for the reception profile of the endorectal and pelvic phased-array coils. A spectroscopic MR imaging volume was then selected by an expert to maximize coverage of the prostate while minimizing the inclusion of peri-prostatic fat and rectal air. Three-dimensional proton (1H) MR spectroscopic imaging data were acquired by using a water and lipid-suppressed double-spin-echo point-resolved spectroscopic sequence optimized for the quantitative detection of both choline and citrate. Water and lipid suppression was achieved by using the band selective inversion with gradient dephasing technique [117]. To eliminate signals from adjacent tissues, especially periprostatic lipids and the rectal wall [118], outer voxel saturation pulses also were used. Data sets were acquired as $16 \times 8 \times 8$ phase-encoded spectral arrays

Scale	Spectra	Patients
Labeled Spectra	4242	36
Scale 1	820	32
Scale 2	1300	28
Scale 3	1096	36
Scale 4	574	34
Scale 5	452	23

Table 6.1: Number of spectra and patients for each scale as annotated by the expert.

(1024 voxels) by using a nominal spectral resolution of $0.240.34cm^3$, 1000/130, and a 17-minute acquisition time.

6.1.2 Pre-processing

Three-dimensional, MR spectroscopic imaging data were processed and aligned with the corresponding T2-w imaging data using a combination of in-house software and Interactive Display Language (Research Systems, Boulder, Colorado) software tools [118]. The raw spectral data were apodized with a 1-Hz Gaussian function and Fourier transformed in the time domain and in three spatial domains. Choline, creatine, and citrate peak parameters (i.e., peak area, peak height, peak location, and line width) were estimated by using an iterative procedure that was used to first identify statistically significant peaks (those with a signal-to-noise ratio higher than 5) in the magnitude spectrum. The frequency shift that best aligns the spectral peaks with the expected locations of choline, creatine, citrate, and residual water is then estimated. Subsequently, the spectra are phased by using the phase of the residual water and the metabolite resonances. Baseline values were corrected by using a local nonlinear fit to the non-peak regions of the spectra. Subsequent feature extraction and classification steps were performed using algorithms developed within the MATLAB (The MathWorks, Inc.) programming environment.

6.1.3 Ground Truth annotations

For all the studies considered in this work, ex vivo whole mount histological sections obtained from radical prostatectomy specimens were available. The "ground truth" CaP extent on the MR imaging was manually delineated by an expert (JK) by visually registering corresponding histological and radiological sections; correspondence between sections having been determined manually by visually determining anatomical fiducials on the histology and the imaging. Having delineated the CaP extent on the MR imaging, an expert spectroscopist then labeled the spectral voxels within the CaP annotated regions on the MRI/MRS according to the 5-point scale. Figure 6.1 shows the standardized 5-point scale developed by Jung et al. [119] which was used to visually classify each spectrum as being either (a) definitely benign (scale 1), (b) probably benign (scale 2), (c) equivocal (scale 3), (d) probably cancer (scale 4), and (e) definitely cancer (scale 5). In this study, all spectra labeled (4, 5) were assumed to be CaP and all spectra labeled as (1, 2) were assumed as benign. The voxels labeled as 3 and atrophic (A) were assumed to be indeterminate and consequently excluded from our analysis. The 36 studies comprised 2120 class 1, 2 and 1026 class 4, 5 spectra (Table 6.1). The class labels for the individual spectral voxels, assigned via a combination of manual registration of histology and MRI and subsequent visual inspection, were used as the surrogate ground truth for CaP extent on the MRI/MRS. This ground truth surrogate is then used for training and evaluation of the MaWERiC classifiers.

6.2 Methodology

The MaWERiC scheme comprises of 4 modules: C.1 wavelet feature extraction, C.2 data representation using PCA, C.3 data combination, and C.4 data classification (Figure 4.1). In the subsequent sub-sections, we will describe each of these modules in detail.

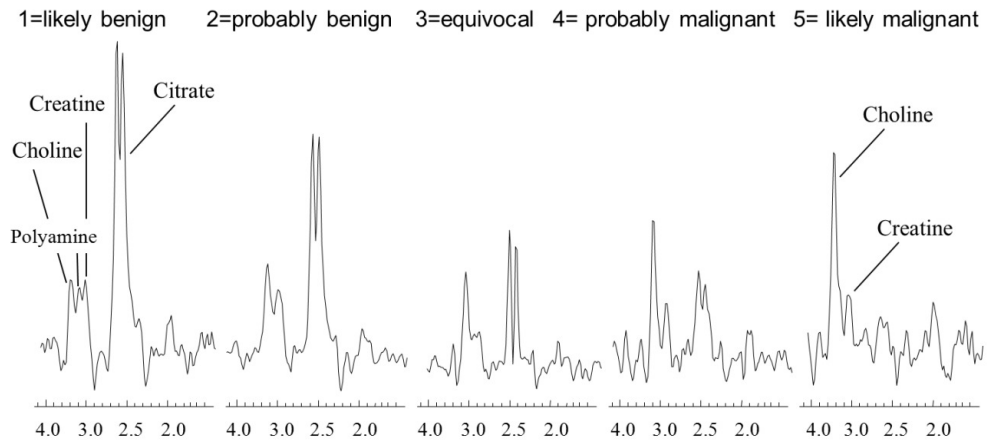


Figure 6.1: Illustration of the standardized five point scale spectra where Figures 6.1(a) - (e) correspond respectively to (a) likely benign (scale 1), (b) probably benign (scale 2), (c) equivocal (scale 3), (d) probably malignant (scale 4), and (e) likely malignant (scale 5) prostate spectra (Figure reproduced from Jung et al. with permission of the author).

6.2.1 Wavelet feature extraction

Haar wavelet features for MRS

The wavelet feature extraction for MRS is performed as detailed in 7.3.

Gabor wavelet features for T2-w MRI

At every metavoxel $c \in C$, a total of 54 Gabor features, $\boldsymbol{F}_k^{T2w}(c), k \in \{1, ..., 54\}$ are obtained at 9 different scales and 6 orientations similar to that shown in [67] and are represented by a Gabor feature vector $\boldsymbol{F}^{T2w}$. Further details on the implementation of Gabor texture features for feature extraction can be found in [67].

6.2.2 Lower dimensional data representation using Principal Component Analysis

At each metavoxel $c \in C$, the high-dimensional MRS wavelet feature vector $\boldsymbol{F}^{MRSw}$ is reduced to transformed Eigen vector $\boldsymbol{F}_{PCA}^{MRSw}(c) = [e_1, e_2, e_3, ..., e_M]$ using PCA, where

$[e_1, e_2, e_3, ..., e_M]$ represent the Eigen vectors obtained from Eigen value decomposition of the data ranked in order of greatest variance. Thus the first m Eigen vectors that represent a pre-specified percentage of the variance in the data are extracted, while the remaining Eigen vectors are discarded. The high dimensional T2-w MRI Gabor feature vector $\boldsymbol{F}^{T2w}(c)$ is similarly reduced to a lower dimensional representation, $\boldsymbol{F}^{T2w}_{PCA}(c)$ using PCA.

6.2.3 Data combination

Owing to the physical and dimensionality differences between the MR spectra and the T2-w MRI features, the MaWERiC meta-classifier is created in the joint T2-w MRI and MRS Eigen space obtained via PCA. Following the mapping of $\boldsymbol{F}^{T2w}$ and $\boldsymbol{F}^{MRSw}$ to reduced dimensional Eigen vector representations, $\boldsymbol{F}^{T2w}_{PCA}$ and $\boldsymbol{F}^{MRSw}_{PCA}$, a new concatenated feature vector $\boldsymbol{F}^{Int}_{PCA} = [\boldsymbol{F}^{T2w}_{PCA}, \boldsymbol{F}^{MRSw}_{PCA}]$ is obtained.

6.2.4 Data Classification using a Random Forest classifier

We define $\mathbf{h}^{\rho}(c)$ as the binary prediction result for the RF classifier, at each threshold $\rho \in [0, 1]$ such that $\mathbf{h}^{\rho}(c) = 1$ when $\mathbf{h}(c) \geq \rho$, 0 otherwise. Further details of RF classifier are detailed in Section 3.8.1.

6.3 Experimental Design and Evaluation

6.3.1 Comparative Strategies

In the following sub-sections, we evaluate and compare the individual modules (feature extraction, classification, data integration) comprising MaWERiC with (i) other feature extraction schemes [67, 120] used in the context of automated CaP detection for individual T2-w MRI, MRS modalities, (ii) a data integration scheme similar to a COD scheme presented in [83] that combines MRS metabolite features with T2-w MRI intensities, (iii) a decision integration strategy and (iv) two other ensemble classifiers, SVMs [101] and probabilistic boosting trees (PBT) [99]. Comparative feature extraction strategies (T2-w MRI, MRS) Below, we discuss some of the feature extraction and

quantification methods previously proposed in the context of MRS [71, 72] and T2-w MRI [11, 67, 24] and that we have implemented in order to quantitatively compare against MaWERiC. The results of these comparative studies will be described later in the Results Section.

Metabolic Peak features for MRS

In the clinic, radiologists typically assess presence of CaP on MRS based on the choline (A_{ch}), creatine (A_{cr}), citrate peaks (A_{cit}) and the $\frac{A_{ch}+cr}{A_{cit}}$ ratio. Variations in these values from predefined normal ranges ($\frac{A_{ch}+cr}{A_{cit}} < 1$) is highly indicative of the presence of the disease [121, 122]. To compare MaWERiC with metabolic features used clinically, we created a metabolic feature vector for MRS, by calculating area under the choline (A_{ch}), creatine (A_{cr}), citrate (A_{cit}) peaks using the composite trapezoidal rule and recording the corresponding ratios ($\frac{A_{ch}}{A_{cr}}$, $\frac{A_{ch+cr}}{A_{cit}}$) [120]. Each $c \in C$ is then defined by a metabolite feature vector $\boldsymbol{F}^{MRS}(c) = [A_{ch}, A_{cr}, A_{cit}, \frac{A_{ch}}{A_{cr}}, \frac{A_{ch+cr}}{A_{cit}}]$.

Texture features for T2-w MRI

Below is a brief description of other individual texture features which have previously been explored in conjunction with classifiers for discriminating between CaP and normal areas on T2-w MRI [11, 67]. A more detailed description of the texture features is provided in Table 3.1.

- Non-steerable Gradient Thirteen non-steerable gradient features for each voxel on the T2-w MRI scene were obtained via convolution of the T2-w MRI scene with the Sobel, Kirsch and standard derivative operators at every spatial location [123].
- First Order Statistical: A total of 8 first-order statistical features including mean, median, standard deviation, and range of gray scale image intensities within a sliding window neighborhood of 3×3 pixels centered around each spatial location in the T2-w MRI scene were extracted [123].
- Second Order Statistical: A total of 13 Haralick features including energy, entropy, inertia, contrast, correlation, sum average, sum variance, sum entropy, difference

average, difference variance, difference entropy, local homogeneity and average deviation were extracted within a sliding window neighborhood of 3x3 pixels centered around each voxel in the T2-w MRI scene [95].

For each class of texture features (first order statistical, second order statistical, non-steerable gradient), corresponding T2-w texture feature vectors $\boldsymbol{F}^{t2\tau_i}, i \in \{1,2,3\}$, are constructed at every $c \in C$. A combined ensemble of texture features is defined as $\boldsymbol{F}^{t2\tau} = [\boldsymbol{F}^{T2w}, \boldsymbol{F}^{t2\tau_1}, \boldsymbol{F}^{t2\tau_2}, \boldsymbol{F}^{t2\tau_3}]$ obtained by concatenating all textural attributes obtained from T2-w MRI. PCA was used to reduce each individual texture feature, $\boldsymbol{F}^{t2\tau_i}, i \in \{1,2,3\}$ to the corresponding low dimensional representation, $\boldsymbol{F}^{t2\tau_i}_{PCA}(c)$, and this was then used for classification.

6.3.2 Comparative Data integration strategies

Classifier combination (COI)

Classifiers $\mathbf{h}^{T2w}(c), \mathbf{h}^{MRS}(c)$ are individually trained on $\boldsymbol{F}^{T2w}_{PCA}(c)$ and $\boldsymbol{F}^{MRS}(c)$, for all $c \in C$. The independence assumption [82] can then be invoked to fuse $\mathbf{h}^{T2w}_{PCA}(c)$ and $\mathbf{h}^{MRS}(c)$ at each $c \in C$, and at every threshold ρ as $\mathbf{h}^{Int}_{d}(c) = \mathbf{h}^{T2w}_{PCA}(c) \times \mathbf{h}^{MRS}(c)$, $\mathbf{h} \in \{\textbf{RF,PBT,SVM}\}$.

Data combination (COD) via MRS metabolic area and ratio features and T2-w image intensity

A combined feature vector $\boldsymbol{F}^{Int}(c) = [\mathbf{F}^{MRS}(c), \boldsymbol{F}^{T2w}(c)]$ is obtained by concatenating the MRS metabolite area and ratio features ($\boldsymbol{F}^{MRS}(c)$) with the mean intensity feature ($\boldsymbol{F}^{T2}(c)$) for each metavoxel $c \in C$. RF classifier along with PBT and SVM classifiers (described in the next section) are then trained using $\boldsymbol{F}^{Int}(c)$ to obtain the meta-classifiers $\mathbf{h}^{Int}(c), \mathbf{h} \in \{\textbf{PBT,SVM,RF}\}$.

6.3.3 Comparative classifier strategies

Probabilistic Boosting Tree (PBT) classifier

The conditional probability that any $c \in C$ belongs to the CaP class, given the combined MRI-MRS feature vector, $\boldsymbol{F}^{Int}_{PCA}(c)$, is calculated at each node based on the learned hierarchical tree as described in Section 3.8.2.

Support vector machine (SVM) classifier

Similarly, the conditional probability that any $c \in C$ belongs to the CaP class, given the combined MRI-MRS feature vector, $\boldsymbol{F}^{Int}_{PCA}(c)$, is calculated using an SVM classifier (details in Section 3.8.3).

6.4 Performance Measures

The classification performance of MaWERiC strategy was compared against related state-of-the-art feature extraction, classifier, and data fusion strategies via (a) area under the Receiver Operating Characteristic (ROC) [124] curve (μ^{AUC}), and (b) classification accuracy (μ^{Acc}) at the operating point on the ROC curve. Both performance measures were reported for voxel-level classification.

6.4.1 Classifier Accuracy

Based on the binary prediction results obtained from the classifier, ROC curves representing the trade-off between CaP detection sensitivity and specificity can be generated. Each point on the curve corresponds to the voxel-level CaP detection sensitivity and specificity of the classifier ($\mathbf{h}^{\rho}(c)$) for some $\rho \in [0, 1]$. The operating point Θ on the ROC curve is defined as value of ρ which yields detection sensitivity and specificity that is closest to 100%. A 3-fold, randomized cross-validation procedure was employed for evaluating performance of MaWERiC against other strategies. Hence for the 36 patient studies considered in this study, 3 sets of spectra each obtained from 12 different studies were constituted. During a single run of cross-validation, 2 out of the 3 sets (corresponding to 24 studies) were chosen for training the classifier while the remaining set of

12 patient studies were used for independent testing. Classifier results were generated on a per voxel basis. This process was repeated until all voxels from all 36 studies were classified once within a single run of cross-validation. This randomized cross-validation process was then repeated a total of 25 times for different training and testing sets. The mean and standard deviation of classifier AUC values (μ^{AUC}) were recorded over these 25 runs. Additionally, the classifier accuracy (μ^{Acc}) at the operating point of the ROC curve was also recorded.

6.5 Experimental Setup

6.5.1 Experiment 1: Comparison of MaWERiC against uni-modal classifiers (T2-w MRI, MRS)

MaWERiC was compared against individual feature extraction strategies for T2-w MRI and MRS. Individual features obtained from T2-w MRI and MRS were also quantitatively evaluated against each other to determine the best performing T2-w MRI and MRS features in terms of μ^{AUC} and μ^{Acc}.

6.5.2 Experiment 2: Comparison of MaWERiC against other COD and COI strategies

MaWERiC was compared against current state of the art COD, and COI strategies, involving direct combination of metabolic features with T2-w image intensities and combination of individual classifier predictions respectively for MRI-MRS integration, where binary predictions from the two uni-modal classifiers were combined using a dot product operation to obtain the final classification.

6.5.3 Experiment 3: Comparison of classifiers (RFs against PBTs and SVMs)

Performance of SVMs and PBT classifiers was compared against the RF classifier (employed for MaWERiC), and across other comparative studies (uni-modal T2-w MRI,

MRS strategies in Experiment 1 and COD, COI strategies in experiment 2) using μ^{AUC} and μ^{Acc} measures.

6.6 Results

6.6.1 Experiment 1: Comparing MaWERiC against uni-modal classifiers (T2-w MRS, MRS)

Qualitative results of classifications obtained from Gabor T2-w MRI ($\mathbf{h}^{T2w}$), Metabolic MRS features ($\mathbf{h}^{MRS}$), COD ($\mathbf{h}^{Int}$), COI ($\mathbf{h}_d^{Int}$), and MaWERiC ($\mathbf{h}_{PCA}^{Int}$) using a RF classifier are shown in Figure 6.2. Probability heat maps for each strategy were obtained, where the spatial locations shown in red (Figures 6.2(b)-(f)) were identified as having a higher probability of CaP as determined by classifiers $\mathbf{h}^{T2w}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}_d^{Int}$, $\mathbf{h}_{PCA}^{Int}$ and on a single T2-w slice. Locations shown in blue were identified as having a higher probability of being benign by the classifiers. The white outline in Figure 6.2(a) shows the ground truth (outlined with a white rectangle) for CaP as annotated by an expert. Note the high CaP detection sensitivity and specificity of MaWERiC (Figure 6.2(f)) compared to individual uni-modal T2-w MRI (Figure 6.2(b)) and MRS (Figure 6.2(c)).

Figure 6.3(a) shows the AUC results, while Figure 6.3(b) shows the accuracy results for different feature extraction strategies ($\mathbf{h}^{T2w}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}_d^{Int}$, and $\mathbf{h}_{PCA}^{Int}$) obtained via a RF classifier over 25 runs of cross validation using box-and-whiskers plots respectively. Note that $m = 15$ was used to reduce the dimensionality of T2w MRI, MRS features since it captured 93% of the variance across MRS and T2w MRI features. Hence, the dimensionality of MaWERiC used for evaluation was . Table 6.3 shows the quantitative results in terms of AUC and accuracy across various feature extraction and classifier strategies ($\mathbf{h}^{T2w}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}_d^{Int}$, $\mathbf{h}_{PCA}^{Int}$) under evaluation. The μ^{AUC} and μ^{Acc} results shown in Table 6.3 across 25 iterations of 3-fold cross validation suggest higher CaP detection accuracy using MaWERiC ($\mu^{AUC} = 0.89 \pm 0.02$, $\mu^{Acc} = 0.83 \pm 0.03$) against both T2-w MRI ($\mu^{AUC} = 0.55 \pm 0.02$, $\mu^{Acc} = 0.54 \pm 0.01$) and MRS ($\mu^{AUC} = 0.77 \pm 0.03$, $\mu^{AUC} = 0.72 \pm 0.02$) for a RF classifier. Note that

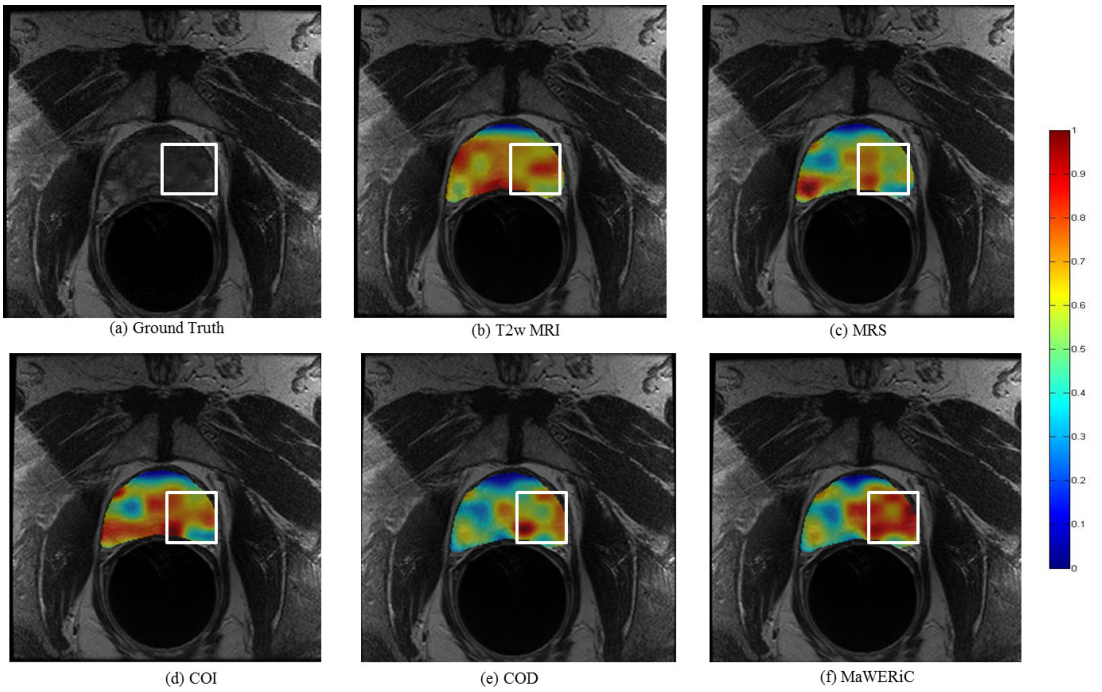
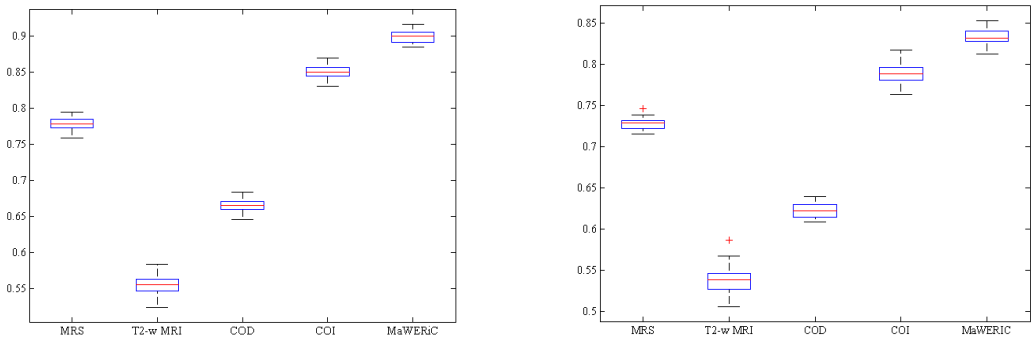


Figure 6.2: (a) Original T2-w image with MRS grid superposed and labeled according to the five point scale (2 = probably benign, 3= indeterminate, 4= probably cancer, 5= definitely cancer, A = atrophy), (b)-(d) probability heat map results superposed on a single T2 slice by interpolating the CaP probabilities at MRS resolution to a pixel level T2w MRI resolution using Gaussian smoothing, (b) T2-w MRI wavelet classifier ($\mathbf{h}^{T2w}$), (c) MRS classifier ($\mathbf{h}^{MRS}$), (d) COD scheme involving integration of MRI mean intensity + MRS metabolic features ($\mathbf{h}^{Int}$), (e) decision level integration ($\mathbf{h}_d^{Int}$), and (f) MaWERiC ($\mathbf{h}_{PCA}^{Int}$) respectively. Locations shown in red correspond to those identified by the classifiers as CaP while those shown in blue correspond to metavoxels classified as benign. Note that the white outline in Figures 6.2(b)-(f) denotes the spatial extent of CaP shown on the T2-w slice. Also note the high detection sensitivity and specificity of CaP probability (Figure 6.2(f)) compared to other classifiers ($\mathbf{h}^{T2w}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}_d^{Int}$) under evaluation.

a higher accuracy for MaWERiC was observed across the other two classifiers (SVM and PBT) as well. Table 6.4 shows the p-values of paired student t-tests conducted over μ^{AUC} values for comparing statistical significant difference of MaWERiC against all the other comparative feature extraction strategies ($\mathbf{h}^{T2w}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}_d^{Int}$), with the null hypothesis being equal classification performance from MaWERiC when compared to the other feature extraction strategies. Significantly superior performance for MaWERiC ($p < 0.05$) was observed for all pairwise comparisons ($\mathbf{h}_{PCA}^{Int}$ - $\mathbf{h}^{T2w}$, $\mathbf{h}_{PCA}^{Int}$ - $\mathbf{h}^{MRS}$, $\mathbf{h}_{PCA}^{Int}$ - $\mathbf{h}^{Int}$, $\mathbf{h}_{PCA}^{Int}$ - $\mathbf{h}_d^{Int}$). Table 6.2 shows the individual μ^{AUC} and μ^{Acc} values (obtained across 25 runs of 3-fold cross validation) using each set of texture features (1st order statistical ($\mathbf{h}^{t2\tau_1}$), 2nd order statistical ($\mathbf{h}^{t2\tau_2}$), Gradient ($\mathbf{h}^{t2\tau_3}$) and Gabor



(a) (b)

Figure 6.3: Box-and-whisker plot results of (a) AUC, and (b) accuracy obtained over 25 runs of 3 fold cross validation across 36 studies for the different feature extraction strategies using a RF classifier. Note that the red line in the middle of each box reflects the median value, while the box is bounded by 25 and 75 percentile of AUC (a) and accuracy (b) values. The whisker plot extends to the minimum and maximum values (obtained across all 25 runs) outside the box and the outliers are denoted as the red plus symbol for different feature extraction methods.

($\mathbf{h}^{t2w}$)), extracted from T2-w MRI across the three sets of classifiers (SVM, RF and PBT). Except in the case of the PBT classifier, Gabor ($\mathbf{h}^{t2w}$) was found to outperform the other first, second-order statistical and gradient texture features ($\mathbf{h}^{t2\tau_1}$, $\mathbf{h}^{t2\tau_2}$, and $\mathbf{h}^{t2\tau_3}$) for both the RF and SVM classifiers.

6.6.2 Experiment 2: Comparing MaWERiC against peak integration/average MR intensities based COD

The qualitative results in Figure 6.2 and box-plots in Figure 6.3 suggest that MaWERiC ($\mathbf{h}^{Int}_{PCA}$) (Figure 6.2(f)) yields a higher detection accuracy compared to state-of-the-art COD ($\mathbf{h}^{Int}$) (Figure 6.2(d)) and COI ($\mathbf{h}^{Int}_{d}$) (Figure 6.2(e)) strategies.

Table 6.3 demonstrates the quantitative results which suggest a significantly higher CaP detection accuracy of MaWERiC, $\mathbf{h}^{Int}_{PCA}$ ($\mu^{AUC} = 0.89 \pm 0.02$, $\mu^{Acc} = 0.83 \pm 0.03$) compared to both COD, ($\mu^{AUC} = 0.66 \pm 0.02$, $\mu^{AUC} = 0.62 \pm 0.02$), and COI, ($\mu^{AUC} = 0.85 \pm 0.03$, $\mu^{AUC} = 0.78 \pm 0.03$) integration strategies using a RF classifier. MaWERiC results were found to be significantly better than the other comparative feature extraction strategies ($\mathbf{h}^{T2w}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}^{Int}_{d}$) across the two classifiers (SVM and PBT) as well.

Classifier	Feature	Mean AUC	Mean accuracy
PBT	Gabor	0.549 ± 0.019	0.545 ± 0.017
	Gradient	0.560 ± 0.018	0.556 ± 0.012
	1st order statistical	0.566 ± 0.013	0.554 ± 0.016
	2nd order statistical	0.534 ± 0.016	0.538 ± 0.012
RF	Gabor	0.554 ± 0.112	0.546 ± 0.014
	Gradient	0.547 ± 0.115	0.545 ± 0.014
	1st order statistical	0.544 ± 0.011	0.537 ± 0.013
	2nd order statistical	0.544 ± 0.011	0.537 ± 0.013
SVM	Gabor	0.513 ± 0.030	0.659 ± 0.013
	Gradient	0.495 ± 0.038	0.499 ± 0.029
	1st order statistical	0.493 ± 0.026	0.498 ± 0.027
	2nd order statistical	0.504 ± 0.041	0.511 ± 0.033

Table 6.2: Mean AUC and accuracy values with standard deviation for different texture and wavelet features obtained using PBT, RF and SVM classifier across 25 iterations of 3 fold cross validation.

1	MRS ($\boldsymbol{F}^{MRS}$)	PBT	0.78 ± 0.03	0.72 ± 0.02
2	T2w MRI ($\boldsymbol{F}^{T2w}_{PCA}$)	PBT	0.54 ± 0.01	0.54 ± 0.01
3	COD ($\boldsymbol{F}^{Int}$)	PBT	0.72 ± 0.03	0.67 ± 0.03
4	COI ($\boldsymbol{F}^{Int}_{d}$)	PBT	0.82 ± 0.01	0.81 ± 0.03
5	MaWERiC($\boldsymbol{F}^{Int}_{PCA}$)	PBT	0.88 ± 0.03	0.81 ± 0.03
6	MRS ($\boldsymbol{F}^{MRS}$)	RF	0.77 ± 0.03	0.72 ± 0.02
7	T2w MRI ($\boldsymbol{F}^{T2w}_{PCA}$)	RF	0.55 ± 0.02	0.54 ± 0.01
8	COD ($\boldsymbol{F}^{Int}$)	RF	0.66 ± 0.02	0.62 ± 0.02
9	COI ($\boldsymbol{F}^{Int}_{d}$)	RF	0.85 ± 0.03	0.78 ± 0.03
10	MaWERiC($\boldsymbol{F}^{Int}_{PCA}$)	RF	0.89 ± 0.02	0.83 ± 0.03
11	MRS ($\boldsymbol{F}^{MRS}$)	SVM	0.73 ± 0.01	0.76 ± 0.14
12	T2w MRI ($\boldsymbol{F}^{T2w}_{PCA}$)	SVM	0.51 ± 0.03	0.65 ± 0.01
13	COD ($\boldsymbol{F}^{Int}$)	SVM	0.68 ± 0.14	0.71 ± 0.08
14	COI ($\boldsymbol{F}^{Int}_{d}$)	SVM	0.81 ± 0.22	0.77 ± 0.18
15	MaWERiC($\boldsymbol{F}^{Int}_{PCA}$)	SVM	0.90 ± 0.24	0.84 ± 0.11

Table 6.3: Mean AUC and accuracy results of different feature extraction and classification techniques used for comparing different methods in this study against MaWERiC across 25 iterations of 3 fold cross validation across three classifier strategies (PBT, RF, SVM).

6.6.3 Experiment 3: Comparing PBT against SVM and RF classifiers

PBTs, SVMs, and RFs demonstrated similar AUC and accuracy results across all feature combination strategies ($\mathbf{h}^{T2w}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}_d^{Int}$, $\mathbf{h}_{PCA}^{Int}$). Both RF and SVM demonstrated higher μ^{AUC} and μ^{Acc} for MaWERiC ($\mathbf{h}_{PCA}^{Int}$) compared to PBT but the results from all three classifiers for were not found to be statistically significantly different from each other. Although slightly higher μ^{AUC} and μ^{Acc} were obtained using SVM classifier, RF was employed for MaWERiC due to its stable performance across different classifier iterations. Results from SVM classifier were found to have a high standard deviation across both accuracy (0.24 for SVM against 0.02 for RF) and AUC values (0.11 for SVM against 0.03 for RF).

6.7 Discussion

To the best of our knowledge, MaWERiC is the first CDS that provides a systematic framework for quantitative combination of structural information from T2-w MRI (imaging) with metabolic information from MRS (non-imaging) for improved CaP detection. The few COI and COD based data integration techniques previously explored in the literature [80, 83] are limited in applicability due to the ad-hoc approaches employed for overcoming dimensionality differences across modalities. For instance, in [83], Simonetti et al. quantitatively combined MRI and MRS by directly concatenating features obtained from the two heterogeneous data sources. However the differing dimensionalities of MRI, MRS features were not accounted for in this study, suggesting that the classifier may have been biased towards the MRS features (8 MRS versus 4 MRI features). Another approach for combining binary decisions, COI, makes an unrealistic assumption of independence across the two data modalities, although complimentary information is acquired simultaneously from the two or more sources about the same disease. MaWERiC was evaluated on 36 1.5 T in-vivo MRS, T2-w MRI patient studies on a per-voxel basis and results thus obtained were compared against 4 other feature extraction strategies, using (i) MRS metabolic features, (ii) T2-w Gabor wavelet features, (iii) a COD scheme involving integration of MRS metabolic features with mean

Method	MRS	T2w MRI	COI	COD
MaWERiC	1.60×10^{-12}	4.06×10^{-24}	0.0017	5.62×10^{-12}

Table 6.4: p-values obtained by pairwise t-test for evaluating presence of statistically significant differences in AUC for MaWERiC against the other 4 methods (Gabor MRI, Metabolic MRS, COI and COD schemes) under evaluation using a RF classifier.

image intensity from T2-w MRI, and (iv) a COI scheme which combined the independent classification results obtained from T2-w MRI and MRS. We also evaluated three classifiers, SVM, PBT and RF, across all the aforementioned 4 strategies (uni-modal MRI, MRS, COD and COI scheme) to identify the best classifier. MaWERiC was found to significantly outperform all the other 4 feature extraction (individual MRS, T2-w MRI) COD and COI strategies, for all 3 classifiers.

To overcome concerns of bias and over-fitting of the data, we iteratively divided 36 patient studies into training and testing sets via a three-fold cross validation scheme. μ^{AUC} and μ^{Acc} values over 25 iteration runs were then reported for all 15 combinations of feature extraction, classifier, and data fusion strategies (see Table 6.2). In the following subsections, we discuss the detection results of MaWERiC with respect to the other feature extraction, data fusion, and classification strategies considered.

6.7.1 Experiment 1: Comparing MaWERiC against uni-modal classifiers (T2-w MRI, MRS)

MaWERiC was found to significantly outperform a uni-modal classifier trained on Gabor features for T2-w MRI. MaWERiC also outperformed a MRS uni-modal classifier trained on clinically used metabolic MRS features. Our results were consistent with several multi-modal integration studies [84, 125, 80, 126, 127, 128] which have suggested that combining orthogonal, complementary pieces of information from different modalities can improve classification accuracy compared to uni-modal data channels [129, 130, 131, 132, 13, 18].

Our results demonstrate that MRS metabolite peak area and ratio features yield better classifiers (at a meta-voxel level) compared to a Gabor texture based T2-w MRI classifier. Our findings are consistent with [66] where μ^{AUC} of 0.68 was obtained using

T2-w MRI compared to μ^{AUC} of 0.80 obtained using MRS, the metabolic peaks having been identified by visual inspection of 2 expert readers. In a related study [133], MRS ratios of metabolite concentrations (μ^{AUC}=0.89) were shown to outperform visually identified, hypo-intense T2-w MRI features (μ^{AUC}=0.85) for CaP detection on a total of 65 patient studies. Note that in these studies the AUC evaluation was done on a per patient basis, as opposed to a voxel-based evaluation, as in our work. Our findings (Figure 6.3 (a)-(b)) suggest that T2-w MRI texture features alone may not be enough to identify CaP signatures on the prostate. Our findings are also consistent with recent 1.5 T and 3T multi-parametric clinical studies [134, 135, 136] which reported sensitivity (at the patient level) in the range of 0.45-0.55 and specificity in the range of 0.80-0.90 from T2-w MRI.

6.7.2 Experiment 2: Comparing MaWERiC against other Data fusion Strategies

MaWERiC versus decision combination (COI)

MaWERiC outperformed a decision level combination scheme [80, 102] in terms of μ^{AUC} and μ^{Acc}. The decision level classifier was obtained by combining the binary class decisions (AND operation) from the individual uni-modal classifiers. Decision level integration while helping to overcome the curse of dimensionality (since all the input information is reduced to a scalar valued decision), tends to implicitly treat the data channels as independent. More specifically, in case of T2-w MRI, MRS, data is acquired simultaneously providing complementary (structural and metabolic) information from each spatial location about the same disease. Decision-level fusion strategies may thus be unable to exploit the synergy between these complimentary data streams. By contrast, data level fusion strategies not only exploit the complementary information spread across the different modalities, but are also able to leverage the cross-talk between the data channels [82].

MaWERiC versus data integration using metabolic MRS and MRI intensity features (COD)

The only other work that we are aware of where MRI and MRS features were quantitatively combined at data-level has been for brain tumor detection [83]. However, in this approach [83] MRS features (obtained via PCA, ICA and quantification) were directly combined with 4 intensity features from multi-protocol MRI, possibly causing the classifier to be biased towards MRS features. Although MaWERiC was compared only against the best performing COD strategy (quantification + MR intensities), one of 4 presented in [83], our superior results suggested that directly aggregating multi-modal, heterogeneous data from very different sources without accounting for differences in feature dimensionality and relative scaling, can adversely impact classifier performance [82]. This is especially true if the constituent classifier features are high dimensional or are unevenly scaled. The superior classifier accuracy of MaWERiC compared to a COD meta-classifier trained using just T2-w MR image intensities and metabolic peak area features (Figures 6.2(a)-(b)), may be attributable to the uniform scaling and data representation provided by the MaWERiC framework.

Since the high dimensional data could be embedded into a reduced space of arbitrary dimensions, we evaluated MaWERiC across different numbers of Eigen vectors, $m \in \{5, 10, 15, 20\}$; The MaWERiC classifier was found to consistently outperform the COD classifier [83] across different values of m. $m = 15$ was chosen as the number of low dimensional embedding vectors on which to project the high dimensional T2w MRI and MRS features, since it accounts for up to 93% of the variance in the data. Note that no significant differences in and for the MaWERiC classifier were observable for , these values accounting for more than 93% of variance in the data. Figure 6.4 shows the variation in μ^{AUC} (y-axis) and μ^{Acc} (x-axis) of MaWERiC using a random forest classifier across different values of MRS dimensions, from $m = 5$ to 40 ($m = 40$ captures 99.8% variance for MRS), with dimension for T2-w MRI fixed at $m = 15$ (captures 99.8% T2w MRI variance). As can be seen from Figure 6.4, the highest AUC and accuracy was obtained when dimensions ($m = 15$) were same for both T2-w

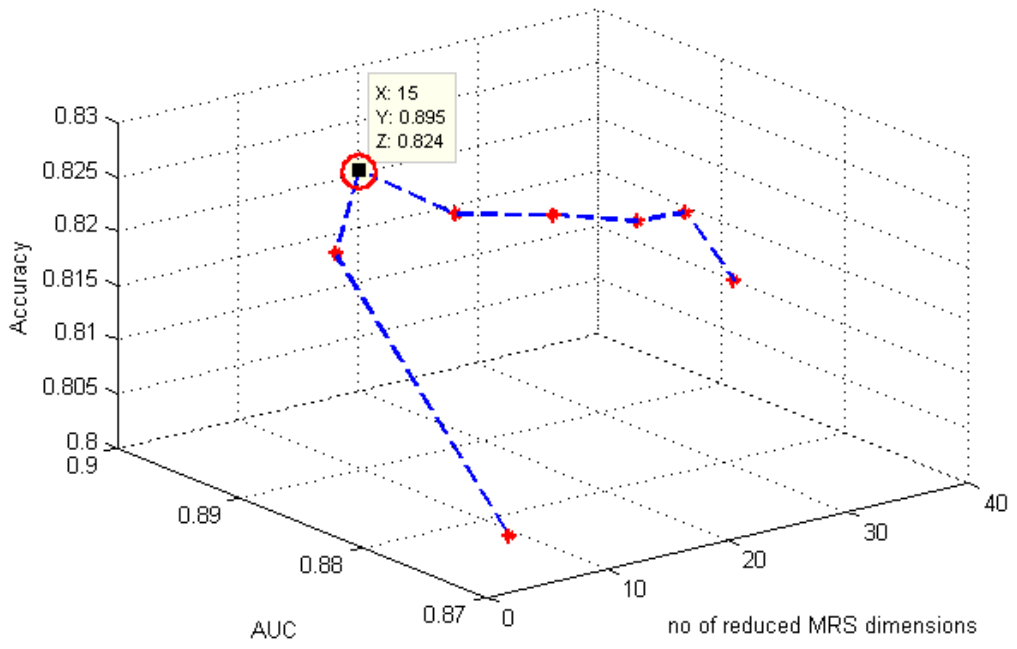


Figure 6.4: 3D Plot showing variation of AUC (y-axis) and accuracy (z-axis) values of MaWERiC across different PCs (x-axis) for MRS (m is fixed as 15 for T2w MRI as it captures 98.8% MRI variance). Note that the highest AUC and accuracy values were obtained when the same number of dimensions were used for both T2- MRI and MRS at m =15.

MRI and MRS. It is important to note here that our choice of number of Eigen vectors was based of maximizing classifier accuracy while using a minimal number of attributes, based on the guiding principle of Occams razor [137].

6.7.3 Experiment 3: Comparing RF against SVM and PBT classifiers

The three classifiers considered in this study, PBTs, RFs and SVMs are all relatively new, state of the art classifier ensembles that have been shown to be useful in different medical imaging applications [138, 11, 67, 120, 110, 71, 72] . The advantage of these classifier ensembles is that they are able to incorporate information from multiple channels and data sources easily. The RF classifier was employed as the ensemble of choice within MaWERiC due to its improved and stable performance over SVM and PBT classifiers (Table 6.3). The RF classifier is known to be able to reduce data variance and hence is able to provide substantial performance improvement over other ensemble classifier strategies [109]. RF classifiers have also shown to be relatively more stable

across different levels of noise compared to other classifier ensembles [109].

It was observed that μ^{AUC} obtained via $\mathbf{h}^{Int}_{PCA}$ was statistically significantly different from $\mathbf{h}^{Int}_{T2w}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$ and, $\mathbf{h}^{Int}_{d}$ across all three classifiers (Table 6.3), although no statistically significant difference was observed across the 3 classifiers (results not shown). These results suggest that the detection performance was more a function of the choice of the feature set and/or fusion strategy (data or decision level), rather than the choice of classifiers.

Although the results obtained via our MaWERiC data integration scheme significantly outperformed a number of state of the art feature extraction and fusion strategies for MRS and T2-w MRI, we also acknowledge a few limitations of our study: (1) the spectra belonging to scale 3 (identified by the expert as being indeterminate) and voxels identified as atrophic (A) were not considered for classification. We believe that spectra classified as intermediate might provide some clinical insights about the disease specific features, a topic which will be explored in future work. (2) Alternative wavelet-based (apart from Haar and Gabor) and other feature extraction strategies (e.g. independent component analysis (ICA) [71, 72]) were not considered. However, our choice of Haar wavelets for MRS and Gabor wavelets for T2-w MRI was motivated by previous demonstrations of their successful employment in building accurate classifiers for CaP detection [67, 107]. (3) While PCA was employed to obtain a uniform, homogeneous space for representation of the different modalities, newer NLDR methods [91, 139, 73] have been shown to yield better low dimensional data representations compared to PCA [73]. However, these NLDR methods are highly sensitive to the parameter selection and selecting the optimal parameters for two modalities would have been a challenge. (4) Ground truth for evaluation was delineated on a per-MRS voxel by an expert, after considering the disease extent mapped on the radiological imaging from corresponding histopathology. Another way of more robustly and accurately estimating spatial extent of disease on the MRI is by spatially co-registering ex vivo whole mount radical prostatectomy sections with corresponding in vivo pre-operative MRI. Our group has previously developed elastic registration algorithms for handling deformations between ex vivo histology and pre-operative MRI [67]. However in this study, this strategy

could not be leveraged due to the non-availability of digital pathology resources for digitization of whole mount histology glass slides.

Chapter 7

Quantitative Integration of Multi-parametric MRI: Application to Prostate Cancer Grading

7.1 Overview

Figure 7.1 provides an overview of the hierarchical scheme that leverages SeSMiK-GE for CaP detection and grading. In Step 1, a RF classifier is trained on the low dimensional representation obtained via SeSMiK-GE to obtain a probability of each spatial location on T2w MRI/MRS scene as being either cancer or benign. A probabilistic pairwise Markov model (PPMM) algorithm [140] is then leveraged to impose spatial constraints to the RF classifier result, yielding a lesion segmentation. CaP lesions identified in Step 1, in Step 2 are then further distinguished as high or low grade CaP via the RF classifier trained on the SeSMiK-GE derived low dimensional representation of the data. Our assumption is that by first localizing the CaP ROI, we can achieve better discriminability between high and low grade CaP regions, as opposed to a non hierarchical three-class classifier (i.e. attempting to directly distinguish between normal, high and low grade CaP).

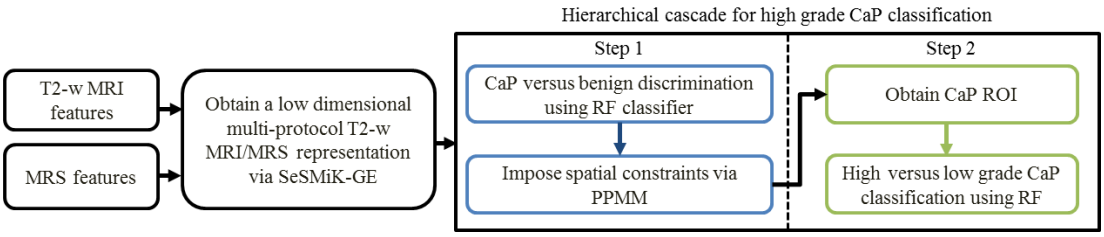


Figure 7.1: Flowchart showing the hierarchical classification strategy employed in this work for CaP detection and grading. In Step 1, CaP ROI is identified using RF and PPMM classifier trained on the SeSMiK-GE derived low dimensional data representation. In Step 2, CaP regions identified in Step 1, are further discriminated as high and low grade CaP.

7.2 Description of Semi-Supervised Multi Kernel Graph Embedding (SeSMiK-GE)

In the following subsections, we describe the detailed mathematical formulation of our SeSMiK-GE strategy. We first provide a brief overview of (1) kernel graph embedding, (2) semi-supervised, and (3) multi-kernel learning strategies. We subsequently describe how these different strategies are combined within the SeSMiK-GE framework for data integration and constructing meta-classifiers.

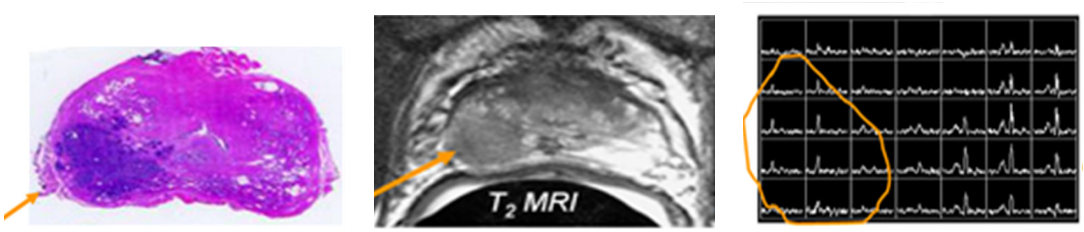
7.3 Experimental Design

7.3.1 Data Description

A total of 29 1.5 Tesla (T) endorectal T2w MRI, MRS studies were obtained prior to radical prostatectomy. The 29 1.5 T studies comprised a total of 960 CaP and 1365 benign metavoxels. Of the 29 studies, 12 studies were found to have high grade CaP with 188 low and 310 high grade CaP metavoxels. All of these studies were biopsy proven prostate cancer patient studies that were clinically referred for a CaP MR staging exam for improved therapeutic selection. All MP MR imaging and spectroscopy scans were performed using the acquisition details provided in Section 6.1.1.

7.3.2 Manual Ground Truth annotation of voxels

For all the studies considered in this work, *ex-vivo* whole mount histological sections obtained from radical prostatectomy specimens were available (Figure 7.2). The "ground truth" CaP and grade extent on the MR imaging was manually delineated and graded by an expert (JK) (more than 25 year experience in the field of MP MRI for CaP detection and grading) by visually registering corresponding histological (Figure 7.2(a)) and radiological sections (Figure 7.2(b)); correspondence between sections having been determined manually by visually determining anatomical fiducials (urethra, verumontanum, large benign prostatic hyperplasia (BPH) nodules) on the histology and the imaging. Having delineated the CaP extent on the MR imaging, the individual regions



(a) (b) (c)

Figure 7.2: (a) Ground truth for CaP extent as defined through the histopathological analysis of hemotoxylin and eosin stained tissue section. The histological CaP extent in Figure 7.2(a) is then visually registered onto the corresponding T2w MRI (Figure 7.2(b)) and MRS sections (Figure 7.2(c)) by an expert using histology as a visual reference.

were then graded based on the 5-point scale (Figures 6.1, 7.3). The class labels for the individual spectral voxels, assigned via a combination of manual registration of histology and MRI and subsequent visual inspection, were used as the surrogate ground truth for CaP detection and grading on the MRI/MRS and employed for subsequent training and evaluation of the SeSMiK-GE classifier.

Since digitized images of the radical prostatectomy histologic sections were not available for this study (Figure 7.2(a) only represents one single illustrative example), deformable image registration methods for spatially aligning *ex-vivo* histology and pre-operative MRI [142, 143] could not be employed to more precisely map the extent of CaP on to MRI. The 5-point scale (described in Chapter 6 Figure 6.1) used by the expert for annotation purposes in this work has previously been used as surrogate ground truth in several clinical papers for CaP evaluation in the absence of digitized whole mount histology sections [144, 145, 146].

Annotations for High vs. low Gleason score

Of the labels that were annotated as CaP, spectra were further identified as being of high or low grade (using a scale of 1 to 5). Figure 7.3 shows the five-point scale used to annotate different grades of CaP by our expert spectroscopist (JK), (a) Gleason score 6 (3+3) (scale 1), (b) Gleason score 7 (3+4) (scale 2), (c) Gleason score 7 (4+3) (scale 3), (d) Gleason score 8 (4+4) (scale 4), and (e) Gleason score 9 (>4+4) (scale 5). Again, the voxels labeled (1, 2) were assumed to be low grade and voxels labeled (3, 4, 5) were

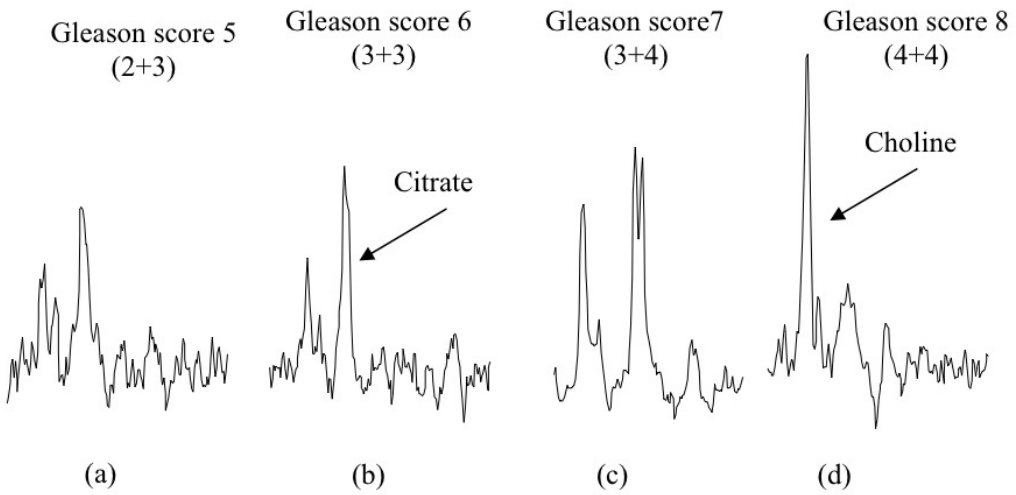


Figure 7.3: Illustration of the five point scale used to annotate ground truth for each spectrum as (a) Gleason score 5 (scale 1), (b) Gleason score 6 (scale 2), (c) Gleason score 7 (scale 3), and (d) Gleason score 8 (scale 4), where increasing Gleason score reflects greater severity of CaP.

assumed to be high grade.

7.3.3 Feature extraction from MRI and MRS

Feature extraction from MRS

For each $c \in C$, $\boldsymbol{F}(c) = [f_a(c)|a \in \{1, ...U\}]$, represents the MR spectral vector, reflecting the frequency component of each of U metabolites. The corresponding spectral data matrix is given as $\mathcal{F} = [\mathbf{F}_1(c); \mathbf{F}_2(c), ...; \mathbf{F}_n(c)] \in \mathbb{R}^{n \times U}$ where $n = |C|$, $|C|$ is the cardinality of C.

Feature extraction from T2w MRI

T2w MRI images were first corrected for bias field (due to the inhomogeneous endorectal coil reception) using a post-processing correction strategy as given in [147]. 34 texture features were then extracted for each metavoxel $c \in C$ based on responses to various gradient filters and gray level co-occurrence operators. These features were chosen based on their previous demonstrated discriminability between CaP and benign regions on T2w MRI [23]. Figure 7.4 shows an example of a texture feature (Figure 7.4 (b)) extracted from the original T2w MRI image (Figure 7.4 (a)) that captures some of the subtle differences between low and high Gleason grade CaP regions, both of which

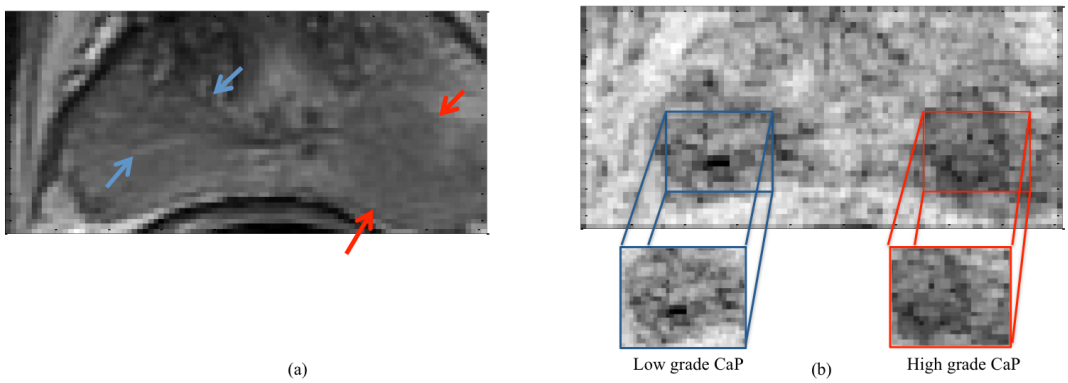


Figure 7.4: (a) Ground truth for high (red arrows) and low grade (blue arrows) CaP extent on a single T2w MRI section. (b) A grey-level texture feature for the corresponding section used to illustrate subtle, yet existing texture differences for low and high grade CaP regions on the same section.

appear as hypointense on the original T2w MRI image. A brief summary of T2w MRI features extracted in this work is provided in Table 3.1.

Feature extraction for T2w MRI was performed in two steps. In Step 1, we calculated the feature scenes $\hat{\mathcal{G}}_u = (\hat{C}, \hat{f}_u)$ for each $\hat{\mathcal{C}}$ by applying the feature operators $\Phi_u, u \in \{1, \ldots, 34\}$ within a local neighborhood associated with every $\hat{c} \in \hat{C}$. 13 gradient (Kirsch, Sobel, Directional filters), 8 first order statistical (Grey Level features with window sizes, 3×3, and 5×5) and 13 Haralick features were extracted at each $\hat{c} \in \hat{C}$. In Step 2, pixel level features (obtained in Step 1) are re-sampled to a lower MRS voxel level resolution. For each MRS voxel $c \in C$, a T2w MRI texture feature vector is obtained by taking the average of the feature values within the corresponding metavoxel as $g_u(c) = \frac{1}{|R_{cd}|} \sum_{\hat{c} \in R_{cd}} \left[\hat{f}_u(\hat{c})\right]$. The corresponding T2w MRI feature vector is then expressed as $\boldsymbol{G}(c) = [g_u(c) | u \in \{1, \ldots, 34\}], \forall c \in C$, and the MRI data matrix is given as $\mathcal{G} = [\boldsymbol{G}_1; \boldsymbol{G}_2; ...; \boldsymbol{G}_n] \in \mathbb{R}^{n \times 34}$. Note that T2w MRI and MRS are in implicit spatial alignment with each other but re-sampling of the T2w MRI features is necessitated by the resolution differences across the imaging and non-imaging MRI protocols.

7.3.4 Hierarchical Classification of high grade CaP via MP SeSMiK-GE signature

Random forest (RF) classifier

Detailed description of RF classifier is provided in Section 3.8.1 (Chapter 3). The final class likelihood that c belongs to CaP, via the RF classifier, is obtained $\mathbf{h}(c)$.

Imposition of spatial constraints via Probabilistic pairwise Markov Random Field (PPMM)

In a Bayesian framework, the restriction of contextual information to local neighborhoods is called the Markov property, and a system of sites that obeys this property is termed MRF [46]. Our group has previously presented a novel extension of MRF, called probabilistic pairwise Markov random model (PPMM) that formulates Markov priors in terms of probability densities, instead of the typical potential functions, facilitating the creation of more sophisticated priors [140]. We modeled the local neighborhood constraints of a CaP voxel existing close to another CaP voxel, using PPMMs based on the assumption that a CaP voxel would have a higher probability of co-existing with another CaP voxel compared to a benign voxel. PPMMs are applied to the output of the RF classifier ($\mathbf{h}(c)$) to obtain a spatially constrained classifier output, ($\tilde{\mathbf{h}}(c)$), which accurately delineates CaP presence using fused MP low dimensional representation $\bar{\boldsymbol{\alpha}}$.

Algorithm for Hierarchical Classifier using SeSMiK-GE

Algorithm *SeSMiK-GE* is first called to obtain the fused MP T2-MRI/MRS representation $\boldsymbol{F}^{T2MRS} = \bar{\boldsymbol{\alpha}}$ for each $c \in C$. The algorithm for hierarchical classification of high grade CaP $HierarchHighGradeCaP$ is presented below.

Algorithm *HierarchHighGradeCaP*
Input: $\mathcal{F}$, $\mathcal{G}$, $\mathcal{N}$, d, $\hat{\boldsymbol{\beta}} = [\hat{\beta}_1, \hat{\beta}_2]$
Output: $\tilde{\mathbf{h}}(c)$, $\hat{\mathbf{h}}(c)$
begin

0. Obtain $\boldsymbol{F}^{T2MRS} = SeSMiK\text{-}GE([\mathcal{F}, \mathcal{G}, \mathcal{N}, d, \hat{\boldsymbol{\beta}}])$
1. Compute $\mathbf{h}(c)$ for each $c \in C$ using RF
2. Apply PPMM to obtain $\tilde{\mathbf{h}}(c)$ for each $c \in C$
3. $\rho = 0$
4. while $\rho \leq 1$
5. obtain $\gamma_{(\rho)} = [SN_{(\rho)} + SP_{(\rho)}]$
6. $\rho = \rho + 0.1$
7. *endwhile*
8. Obtain $\nu = \arg\max_{\rho}[\gamma_{(\rho)}]$
9. $A = \{c|\tilde{\mathbf{h}}(c) \geq \nu\}$, $A \subset C$, $\nu \in [0, 1]$
10. $\forall c \in A$, compute $\hat{\mathbf{h}}(c)$ using RF
11. **return** $\tilde{\mathbf{h}}(c)$, $\hat{\mathbf{h}}(c)$

end

In Step 0, fused MP T2w MRI/ MRS signature $\boldsymbol{F}^{T2MRS}$ is obtained by calling the *SeSMiK-GE* algorithm, which is subsequently employed to train a RF classifier to obtain CaP probabilities, $\mathbf{h}(c)$, for each $c \in C$ in Step 1. PPMM is applied in Step 2 on $\mathbf{h}(c)$ to obtain $\tilde{\mathbf{h}}(c)$ for every $c \in C$. In Steps 4 to 7, classifier detection sensitivity ($SN_{(\rho)}$), specificity ($SP_{(\rho)}$), and the sum $\gamma_{(\rho)} = [SN_{(\rho)} + SP_{(\rho)}]$ are calculated at every threshold ρ, $\rho \in [0, 1]$. The thresholds are performed in intervals of 0.1 to obtain 11 $SN_{(\rho)}$, $SP_{(\rho)}$, and γ values. In Step 8, ν is obtained as the threshold that maximizes γ and is subsequently used in Step 9, to identify a subset of voxels, $A \subset C$, that have a CaP probability greater than ν. All the voxels in A are then further distinguished, in Step 10, as belonging to either high or low grade CaP using a RF classifier.

7.3.5 Implementation Details and Classifier Training

Parameter Selection for SeSMiK-GE

For the SSL module, 40% of the total training samples were randomly selected to obtain $\tilde{W}_{MRS}$ and $\tilde{W}_{MRI}$. For the MKL module, a Gaussian kernel was used to obtain

K_{MRI} and K_{MRS}, and the corresponding weights $\hat{\beta}_1$ and $\hat{\beta}_2 = (1 - \hat{\beta}_1)$, $\beta_1, \beta_2 \in [0, 1]$ that optimized CaP accuracy on 40% of the same training set (used for $\hat{W}$). The set of $\hat{W}$ were learned using a hierarchical brute force strategy (see Section 4.2.3), and subsequently used to obtain $\boldsymbol{F}^{T2MRS}$.

Classifier Training

Two independent cross validation strategies, leave-one-out (LOO) and three-fold cross validation, were used for evaluation of CaP classifiers. In LOO, each classifier was trained on 28 CaP studies, while one study was used for testing [76]. This process was repeated until all 29 studies were classified once within a single run of LOO cross-validation on a per voxel basis. Similarly, for three-fold-cross-validation, 29 patient studies were divided into three sets such that two of the three sets (corresponding to 20 studies) were chosen for training the classifier, while the remaining set of 9 studies was used for independent testing. This process was repeated until all 29 studies were classified once within a single run of cross-validation. The three-fold randomized cross-validation process was then repeated 25 times for different training and testing sets.

A total of 12 out of the 29 CaP patient studies were found to have high grade CaP. Both LOO, and three-fold cross validation strategies were similarly independently used for evaluation of high Gleason grade CaP classifiers over 12 patient studies. For LOO, 11 studies were used to training and 1 study was used for testing at a time, until all 12 patient studies were classified. For three-fold-cross-validation, 8 studies were used for training and 4 for testing, until all 12 patient studies are classified within a single cross-validation run. Three-fold-cross-validation was similarly repeated 25 times over different training and testing sets for classifying patients with high Gleason grade CaP on a per-voxel basis.

Performance Evaluation Measures

ROC curves representing the trade-off between CaP detection sensitivity and specificity were independently generated for each of the classifiers considered in this work for both leave-one-out and three-fold cross validation strategies as described in Section **??**.

7.3.6 Comparative Data Integration Strategies

In the following sub-sections, we evaluate and compare SeSMiK with other feature extraction schemes used in the context of automated CaP detection for (i) individual MRS [73], T2w MRI [23] modalities, (ii) a data integration (COD) scheme for combining MRS with T2w MRI features, and (iii) a decision integration strategy where the individual uni-modal classifier outputs are fused to obtain a combined classification output [58]. A summary of all the comparative strategies used for CaP detection and grading in this work is given in Table 7.1.

Uni-modal T2w MRI, MRS classifiers

The high dimensional MRS and T2w MRI feature vectors, $\mathbf{F}$ and $\mathbf{G}$ respectively, are reduced to corresponding low dimensional representations, $\mathbf{F}^{MRS}$, and $\mathbf{F}^{T2}$ using GE [73] for each $c \in C$. $\mathbf{F}^{MRS}$ and $\mathbf{F}^{T2}$ were used to train uni-modal T2w MRI, MRS classifiers, $\mathbf{h}^{T2}$ and $\mathbf{h}^{MRS}$ respectively. Similarly, $\tilde{\mathbf{h}}^{MRS}$ and $\tilde{\mathbf{h}}^{T2}$ were obtained using a PPMM classifier on $\mathbf{h}^{T2}$ and $\mathbf{h}^{MRS}$. Corresponding uni-modal grading classifiers for T2w MRI and MRS were obtained as $\hat{\mathbf{h}}^{T2}$ and $\hat{\mathbf{h}}^{MRS}$.

Classifier combination (COI)

The independence assumption can be invoked to fuse $\mathbf{h}^{MRS}$ and $\mathbf{h}^{T2}$ at each $c \in C$ as $\mathbf{h}^{IntD}(c) = \mathbf{h}^{T2}(c) \times \mathbf{h}^{MRS}(c)$. PPMM classifier was employed on the decision classifier output $\mathbf{h}^{IntD}(c)$ to obtain $\tilde{\mathbf{h}}^{IntD}(c)$, and similarly, $\hat{\mathbf{h}}^{IntD}(c)$ was obtained for high grade CaP classification via COI for each $c \in C$.

Data combination (COD)

A combined feature vector $\boldsymbol{F}^{Int}(c) = [\boldsymbol{F}^{MRS}(c), \boldsymbol{F}^{T2}(c)]$ is obtained by concatenating MRS and T2w MRI reduced Eigenfeatures for each metavoxel $c \in C$. A RF classifier is then trained using $\boldsymbol{F}^{Int}(c)$ to obtain the CaP meta-classifier $\mathbf{h}^{Int}(c)$, followed by the corresponding MRF classifier output as $\tilde{\mathbf{h}}^{Int}(c)$. The corresponding high grade COD classifier is obtained as $\hat{\mathbf{h}}^{Int}(c)$ for each $c \in C$.

Index	Feature extraction strategy	CaP Classifier	PPMM	Grade classifier
1.	MRS classifier	$\mathbf{h}^{MRS}$	$\tilde{\mathbf{h}}^{MRS}$	$\hat{\mathbf{h}}^{MRS}$
2.	T2w MRI classifier	$\mathbf{h}^{T2}$	$\tilde{\mathbf{h}}^{T2}$	$\hat{\mathbf{h}}^{T2}$
3.	Classifier combination	$\mathbf{h}^{IntD} = \mathbf{h}^{MRS} \times \mathbf{h}^{T2}$	$\tilde{\mathbf{h}}^{IntD}$	$\hat{\mathbf{h}}^{IntD}$
4.	Data combination classifier	$\mathbf{h}^{Int} = [\mathbf{h}^{MRS}, \mathbf{h}^{T2}]$	$\tilde{\mathbf{h}}^{Int}$	$\hat{\mathbf{h}}^{Int}$
5.	SeSMiK-GE classifier	$\mathbf{h}^{T2MRS}$	$\tilde{\mathbf{h}}^{T2MRS}$	$\hat{\mathbf{h}}^{T2MRS}$

Table 7.1: Summary of different feature extraction and classifier techniques compared in this work against the SeSMiK-GE classifier.

7.3.7 Experimental Evaluation

Experiment 1. CaP detection via SeSMiK-GE

We compared the performance of PPMM based SeSMiK-GE classifier ($\tilde{\mathbf{h}}^{T2MRS}(c)$) against each of the individual PPMM spatially constrained classifier outputs for T2w MRI ($\tilde{\mathbf{h}}^{T2}(c)$), MRS ($\tilde{\mathbf{h}}^{MRS}(c)$) as well as COD ($\tilde{\mathbf{h}}^{Int}(c)$), and COI ($\tilde{\mathbf{h}}^{IntD}(c)$) classifiers in accurately identifying CaP regions. This was done via (a) area under the ROC [150] curve (φ^{AUC}), and (b) classification accuracy (φ^{Acc}) at the operating point on the ROC curve.

Experiment 2. High grade CaP detection via SeSMiK-GE

ROC analysis across both LOO and three-fold cross validation was independently performed to compare the performance of SeSMiK-GE in accurately identifying high grade CaP ($\hat{\mathbf{h}}^{T2MRS}(c)$) against other classifier strategies ($\hat{\mathbf{h}}^{T2}(c)$, $\hat{\mathbf{h}}^{MRS}(c)$, $\hat{\mathbf{h}}^{Int}(c)$, $\hat{\mathbf{h}}^{IntD}(c)$) via φ^{AUC} and φ^{Acc}.

7.4 Results and Discussion

7.4.1 Experiment 1: CaP detection via SeSMiK-GE

Figure 7.5 (a), (c) show AUC results (φ^{AUC}) while Figure 7.5 (b), (d) show accuracy results (φ^{Acc}) for LOO and three-fold cross validation across various feature extraction and classifier strategies ($\tilde{\mathbf{h}}^{T2}$, $\tilde{\mathbf{h}}^{MRS}$, $\tilde{\mathbf{h}}^{Int}$, $\tilde{\mathbf{h}}^{IntD}$, $\tilde{\mathbf{h}}^{T2MRS}$) using box-and-whiskers plots. The mean (φ_{μ}^{AUC}) and standard deviation (ζ^{AUC}) of AUC values, and classifier accuracy (φ_{μ}^{Acc}) at the operating point of the ROC curve were also recorded. A higher

φ_{μ}^{AUC} and φ_{μ}^{Acc} was obtained using PPMM based SeSMiK-GE ($\varphi_{\mu}^{AUC} = 0.89 \pm 0.07$, $\varphi_{\mu}^{Acc} = 0.84 \pm 0.14$) compared to the individual T2w MRI ($\varphi_{\mu}^{AUC} = 0.54 \pm 0.18$, $\varphi_{\mu}^{Acc} = 0.58 \pm 0.20$), MRS ($\varphi_{\mu}^{AUC} = 0.61 \pm 0.20$, $\varphi_{\mu}^{Acc} = 0.58 \pm 0.22$), COD ($\varphi_{\mu}^{AUC} = 0.64 \pm 0.22$, $\varphi_{\mu}^{Acc} = 0.62 \pm 0.18$), and COI classifiers ($\varphi_{\mu}^{AUC} = 0.62 \pm 0.07$, $\varphi_{\mu}^{Acc} = 0.45 \pm 0.07$) across different dimensions $d \in \{5, 10, 15, 20\}$ (results reported for $d = 15$) using a LOO cross-validation strategy.

Similar results were obtained using a three-fold cross validation strategy, with a higher φ_{μ}^{AUC} and φ_{μ}^{Acc} reported using PPMM based SeSMiK-GE ($\varphi_{\mu}^{AUC} = 0.85 \pm 0.02$, $\varphi_{\mu}^{Acc} = 0.84 \pm 0.03$) compared to the individual T2w MRI ($\varphi_{\mu}^{AUC} = 0.57 \pm 0.02$, $\varphi_{\mu}^{Acc} = 0.58 \pm 0.01$), MRS ($\varphi_{\mu}^{AUC} = 0.76 \pm 0.01$, $\varphi_{\mu}^{Acc} = 0.67 \pm 0.01$), COD ($\varphi_{\mu}^{AUC} = 0.77 \pm 0.01$, $\varphi_{\mu}^{Acc} = 0.66 \pm 0.02$), and COI classifiers ($\varphi_{\mu}^{AUC} = 0.64 \pm 0.01$, $\varphi_{\mu}^{Acc} = 0.51 \pm 0.01$) across different dimensions. Note that although LOO-cross-validation strategy yielded higher φ_{μ}^{AUC} and φ_{μ}^{Acc} values, variance across φ_{μ}^{AUC} and φ_{μ}^{Acc} was significantly reduced with a three-fold cross-validation strategy.

CV strategy	$\varphi^{AUC}_{\tilde{\mathbf{h}}^{T2}} - \varphi^{AUC}_{\tilde{\mathbf{h}}^{T2MRS}}$	$\varphi^{AUC}_{\tilde{\mathbf{h}}^{MRS}} - \varphi^{AUC}_{\tilde{\mathbf{h}}^{T2MRS}}$	$\varphi^{AUC}_{\tilde{\mathbf{h}}^{COD}} - \varphi^{AUC}_{\tilde{\mathbf{h}}^{T2MRS}}$	$\varphi^{AUC}_{\tilde{\mathbf{h}}^{COI}} - \varphi^{AUC}_{\tilde{\mathbf{h}}^{T2MRS}}$
LOO	1.87×10^{-5}	0.001	0.005	2.8×10^{-6}
3-fold	1.4×10^{-7}	1.4×10^{-7}	1.4×10^{-7}	1.4×10^{-7}

(a)

CV strategy	$\varphi^{Acc}_{\tilde{\mathbf{h}}^{T2}} - \varphi^{Acc}_{\tilde{\mathbf{h}}^{T2MRS}}$	$\varphi^{Acc}_{\tilde{\mathbf{h}}^{MRS}} - \varphi^{Acc}_{\tilde{\mathbf{h}}^{T2MRS}}$	$\varphi^{Acc}_{\tilde{\mathbf{h}}^{COD}} - \varphi^{Acc}_{\tilde{\mathbf{h}}^{T2MRS}}$	$\varphi^{Acc}_{\tilde{\mathbf{h}}^{COI}} - \varphi^{Acc}_{\tilde{\mathbf{h}}^{T2MRS}}$
LOO	1.87×10^{-4}	6.09×10^{-4}	8.8×10^{-4}	4.17×10^{-5}
3-fold	2.02×10^{-7}	2.09×10^{-4}	2.09×10^{-4}	1.47×10^{-7}

(b)

Table 7.2: Table showing the p-values of statistical significance obtained using a pairwise Wilcoxon signed ranked test across the two cross-validation (CV) strategies (LOO and 3-fold) while comparing classifiers $\tilde{\mathbf{h}}^{T2}$, $\tilde{\mathbf{h}}^{MRS}$, $\tilde{\mathbf{h}}^{Int}$, $\tilde{\mathbf{h}}^{IntD}$, with $\tilde{\mathbf{h}}^{T2MRS}$ for (a) φ^{AUC} and (b) φ^{Acc} at $d = 15$.

Figure 7.5 suggests that MP COD based data integration strategies ($\tilde{\mathbf{h}}^{Int}$, and $\tilde{\mathbf{h}}^{T2MRS}$) yield higher CaP detection φ^{AUC} and φ^{Acc} as compared to uni-modal $\tilde{\mathbf{h}}^{T2}$ and $\tilde{\mathbf{h}}^{MRS}$ classifiers [45, 24, 46] and a COI based data integration classifier. Table 7.2 shows the p values of statistical significance obtained using a non-parametric Wilcoxon signed test [151] for comparing φ_{μ}^{AUC} and φ_{μ}^{Acc} for $\tilde{\mathbf{h}}^{T2MRS}$ against $\tilde{\mathbf{h}}^{T2}$, $\tilde{\mathbf{h}}^{MRS}$, $\tilde{\mathbf{h}}^{Int}$, and $\tilde{\mathbf{h}}^{IntD}$ across LOO and three-fold-cross validation strategies at $d = 15$. The p-value was appropriately adjusted at $p = 0.001$ on account of multiple testing via a Bonferroni

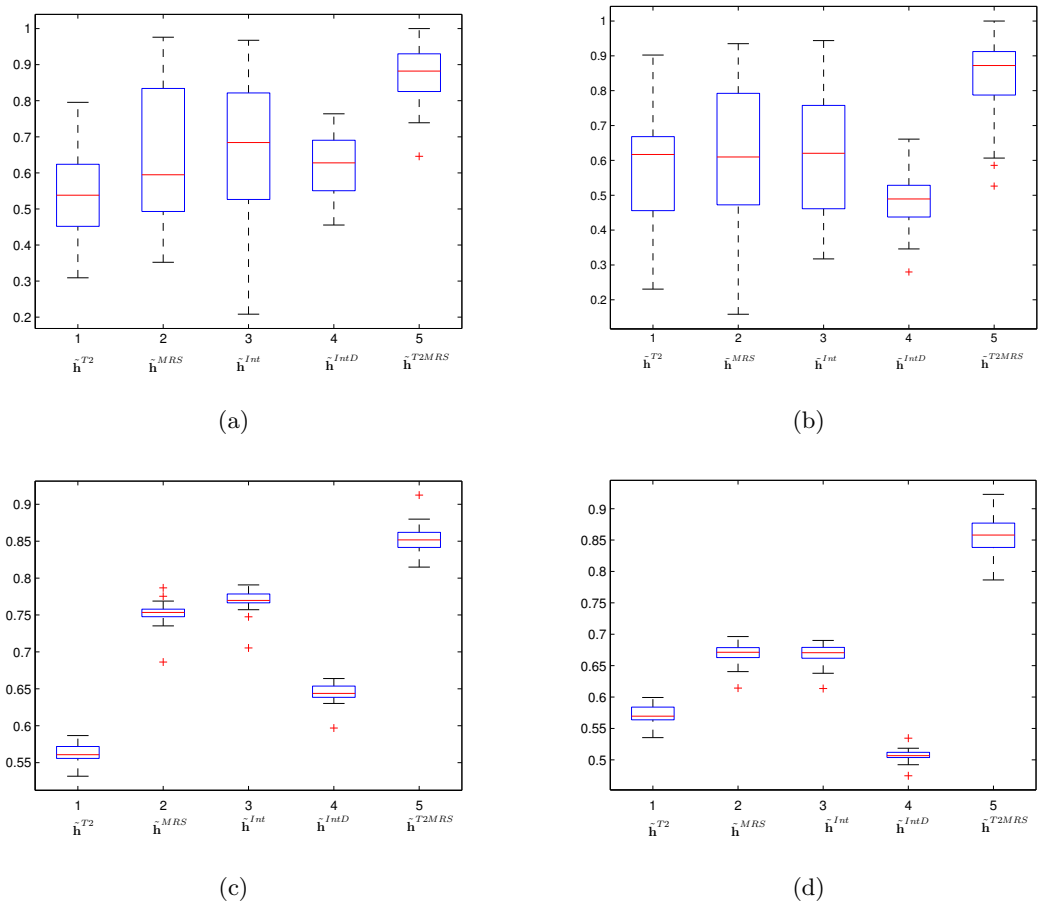


Figure 7.5: Box-and-whisker plot results of AUC (Fig. 7.5(a)) and accuracy (Fig. 7.5(b)) obtained over 29 studies via a leave-one-out cross validation strategy for $\tilde{\mathbf{h}}^{T2}$, $\tilde{\mathbf{h}}^{MRS}$, $\tilde{\mathbf{h}}^{Int}$, $\tilde{\mathbf{h}}^{IntD}$, and $\tilde{\mathbf{h}}^{T2MRS}$. Fig. 7.5(c) and (d) shows the box-whisker-plots for three-fold cross validation strategy over 25 cross-validation runs for AUC and accuracy respectively. Note that the red line in the middle of each box reflects the median value while the box is bounded by 25 and 75 percentile of AUC (Fig. 7.5(a), (c)) and accuracy (Fig. 7.5(b), (d)) values. The whisker plot extends to the minimum and maximum values outside the box and the outliers are denoted as the red plus symbol for different feature extraction strategies.

test. It is worth noting that the results across different $d \in \{10, 15, 20\}$ were found to be consistent, where $\tilde{\mathbf{h}}^{T2MRS}$ consistently significantly outperformed ($p < 0.001$) the other classifiers across all reduced dimensions (results not shown). We believe that the high sensitivity and specificity of SeSMiK-GE compared to the other data integration strategies (COI and COD) is due to (a) combining and weighting individual kernel contributions within the MKL formation, and (b) employing partial label information for SSL which improves class differentiability.

Finally, it should be noted that employing PPMM classifier $\tilde{\mathbf{h}}^{T2MRS}$ ($\varphi_{\mu}^{Acc} = 0.72 \pm 0.16$) (significantly improved CaP detection accuracy ($\sim$20% improvement for $d = 15$) over not imposing any spatial constraints $\mathbf{h}^{T2MRS}$ ($\varphi_{\mu}^{Acc} = 0.89 \pm 0.07$). The improvement was consistent across $\tilde{\mathbf{h}}^{Int}$, $\tilde{\mathbf{h}}^{IntD}$, $\tilde{\mathbf{h}}^{MRS}$, and $\tilde{\mathbf{h}}^{T2}$ over both LOO and three-fold-cross-validation, and resonates with similar findings in [46].

7.4.2 Experiment 2: CaP grading via SeSMiK-GE

Figure 7.6 shows qualitative results for the hierarchical classification strategy for identifying high grade CaP. Figures 7.6 (a), (d), (g) show the CaP and high grade CaP ground truth as annotated by an expert. Yellow outline in Figures 7.6 (a), (d), (g) defines the CaP extent while the black outline denotes the high grade CaP extent for three different T2w MRI sections from three different patients. The corresponding probability heat maps for CaP classification are shown in Figures 7.6 (b), (e), (h), where the spatial locations shown in red were identified by the respective classifiers as having a higher probability of CaP presence. Locations shown in blue were those identified as having a higher probability of being benign by the SeSMiK-GE classifier. Within the high probability CaP regions (red) in Figures 7.6 (b), (e), (h), a probability of high grade CaP is further assigned, the corresponding probability heat maps for which are shown in Figures 7.6 (c), (f), (i). Here, the spatial locations shown in red (Figures 7.6 (c), (f), (i)) were identified as having a higher probability of high grade CaP, while locations shown in blue were identified as having a higher probability of being low grade using $\hat{\mathbf{h}}^{T2MRS}$. Note the high detection accuracy in accurately identifying CaP and high grade CaP using our hierarchical classification strategy across all three T2w MRI

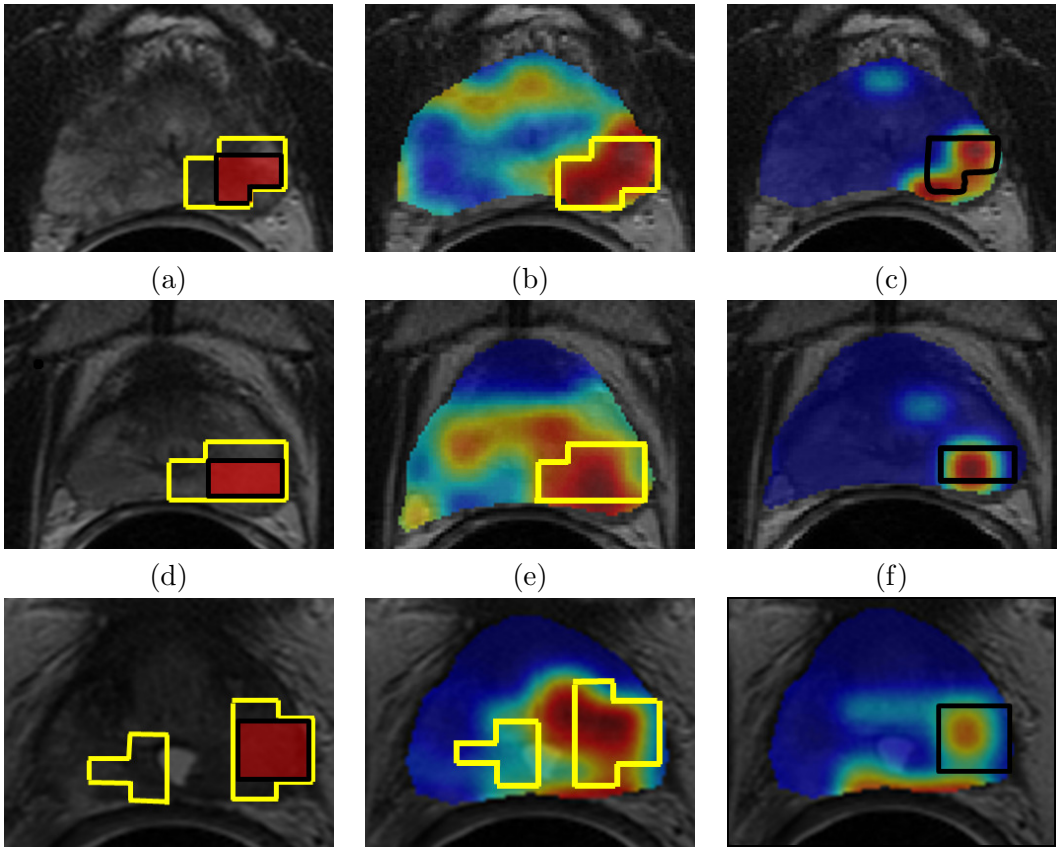


Figure 7.6: (a), (d) and (g) show three different T2w MRI sections with CaP ground truth (as annotated by an expert) outlined in yellow, while the high grade CaP ground truth outlined in red. Figures 7.6(b), (e) and (h) show the probability heat map results corresponding to CaP classification on T2w MRI sections in (a), (d) and (g) respectively for three different T2w MRI studies. Figures 7.6(c), (f) and (i) show the probability heat maps corresponding to high grade CaP classification performed within the spatial locations identified as high probabilistic CaP regions (shown in red) in Figures 7.6(b), (e) and (h) respectively. Note that locations shown in red in Figures 7.6(b), (e) and (h) correspond to those identified by $\mathbf{h}^{T2MRS}$ as CaP while in Figures 7.6(c), (f) and (i) as those identified as high grade CaP by $\hat{\mathbf{h}}^{T2MRS}$. Similarly the spatial locations shown in blue in Figures 7.6(b), (e) and (h) correspond to spatial locations classified as benign and as low grade CaP in Figures 7.6(c), (f) and (i).

sections.

Figure 7.7 (a) shows φ^{AUC} while Figure 7.7 (b) shows φ^{Acc} values across various feature extraction and classifier strategies ($\hat{\mathbf{h}}^{T2}$, $\hat{\mathbf{h}}^{MRS}$, $\hat{\mathbf{h}}^{Int}$, $\hat{\mathbf{h}}^{IntD}$, $\hat{\mathbf{h}}^{T2MRS}$) for high grade CaP using box-and-whiskers plots obtained via a LOO-cross-validation strategy. Corresponding results for φ^{AUC} and φ^{Acc} obtained via a three-fold cross-validation strategy are shown in Figure 7.7 (c) and (d) respectively. A higher AUC ($\varphi_{\mu}^{AUC} = 0.84 \pm 0.07$) was obtained using SeSMiK-GE for classifying high grade CaP via the hierarchical classification strategy, compared to the individual T2w MRI ($\varphi_{\mu}^{AUC} = 0.54 \pm 0.13$), MRS ($\varphi_{\mu}^{AUC} = 0.59 \pm 0.19$), COD ($\varphi_{\mu}^{AUC} = 0.62 \pm 0.18$), and COI classifiers ($\varphi_{\mu}^{AUC} = 0.61 \pm 0.07$), using a LOO cross-validation strategy. Table 7.3 shows the p-values obtained by performing a paired non-parametric Wilcoxon test for comparing φ^{AUC} obtained from $\hat{\mathbf{h}}^{T2MRS}$ with other classifier strategies ($\hat{\mathbf{h}}^{T2}$, $\hat{\mathbf{h}}^{MRS}$, $\hat{\mathbf{h}}^{Int}$, $\hat{\mathbf{h}}^{IntD}$) using LOO and three-fold-cross-validation strategies at $d = 15$.

CV strategy	$\varphi^{AUC}_{\hat{\mathbf{h}}^{T2}} - \varphi^{AUC}_{\hat{\mathbf{h}}^{T2MRS}}$	$\varphi^{AUC}_{\hat{\mathbf{h}}^{MRS}} - \varphi^{AUC}_{\hat{\mathbf{h}}^{T2MRS}}$	$\varphi^{AUC}_{\hat{\mathbf{h}}^{COD}} - \varphi^{AUC}_{\hat{\mathbf{h}}^{T2MRS}}$	$\varphi^{AUC}_{\hat{\mathbf{h}}^{COI}} - \varphi^{AUC}_{\hat{\mathbf{h}}^{T2MRS}}$
LOO	2.3×10^4	0.001	0.002	3.8×10^4
3-fold	1×10^4	0.001	0.001	2.1×10^4

Table 7.3: Table showing the p-values of statistical significance using a pairwise Wilcoxon signed test across the two cross-validation (CV) strategies (LOO and 3-fold) while comparing classifiers $\hat{\mathbf{h}}^{T2}$, $\hat{\mathbf{h}}^{MRS}$, $\hat{\mathbf{h}}^{Int}$, $\hat{\mathbf{h}}^{IntD}$, with $\hat{\mathbf{h}}^{T2MRS}$ for φ^{AUC} at $d = 15$.

Figure 7.8 shows the average ROC curves obtained using $\mathbf{h}^{T2}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}^{T2MRS}$ and $\tilde{\mathbf{h}}^{T2MRS}$ for CaP versus benign classification (Figure 7.8(a)) and using $\hat{\mathbf{h}}^{T2}$, $\hat{\mathbf{h}}^{MRS}$, $\hat{\mathbf{h}}^{Int}$, and $\hat{\mathbf{h}}^{T2MRS}$ for low versus high grade classification (Figure 7.8 (b)) for $d = 15$. Again note the improvement in AUC using the MRF based SeSMiK-GE CaP classifier ($\tilde{\mathbf{h}}^{T2MRS}$) compared to SeSMiK-GE CaP classifier with no spatial constraints ($\mathbf{h}^{T2MRS}$) (Figure 7.8 (a)).

We attribute the high φ^{AUC} and φ^{Acc} obtained via $\hat{\mathbf{h}}^{T2MRS}$ to the hierarchical classification of high grade CaP which systematically hones-in on the CaP region of interest by eliminating other confounders that might otherwise affect classification [73]. The hierarchical scheme is especially relevant in cases where the morphologic differences between the two classes are subtle. These differences may not be appreciable in the presence of other object classes which first need to be eliminated (benign confounders)

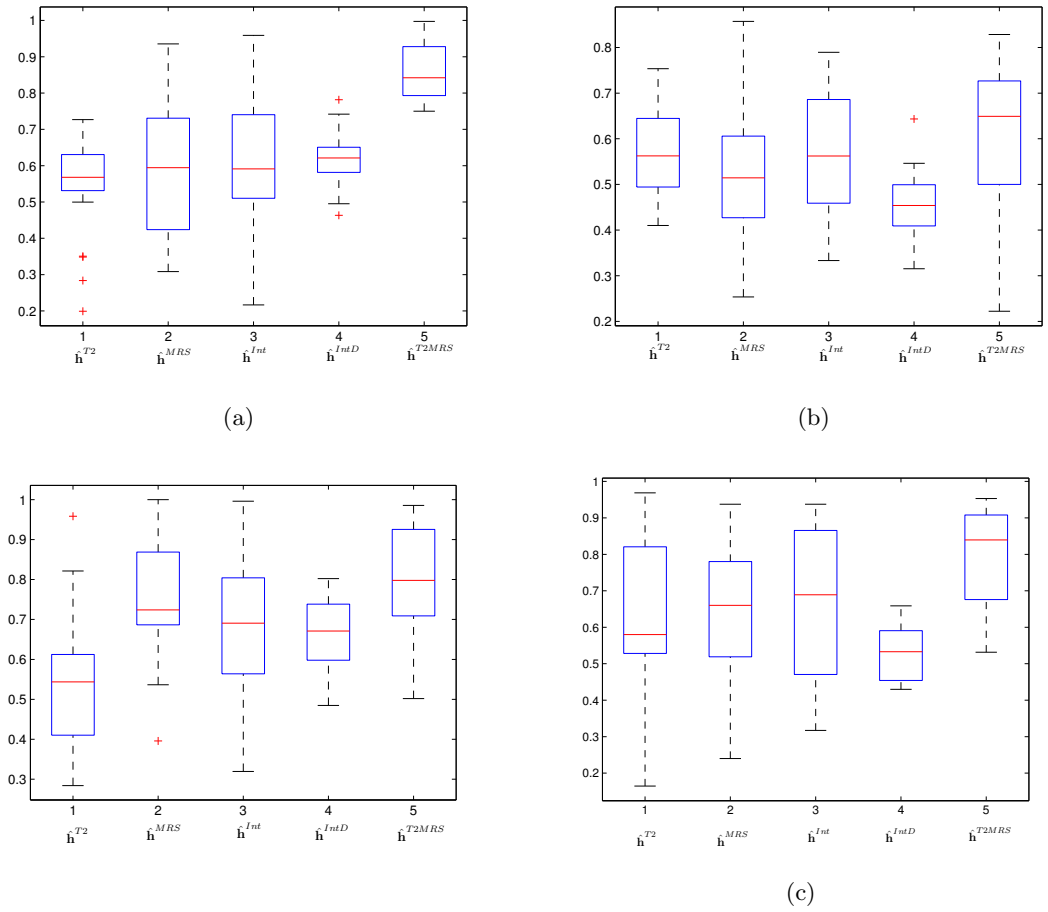
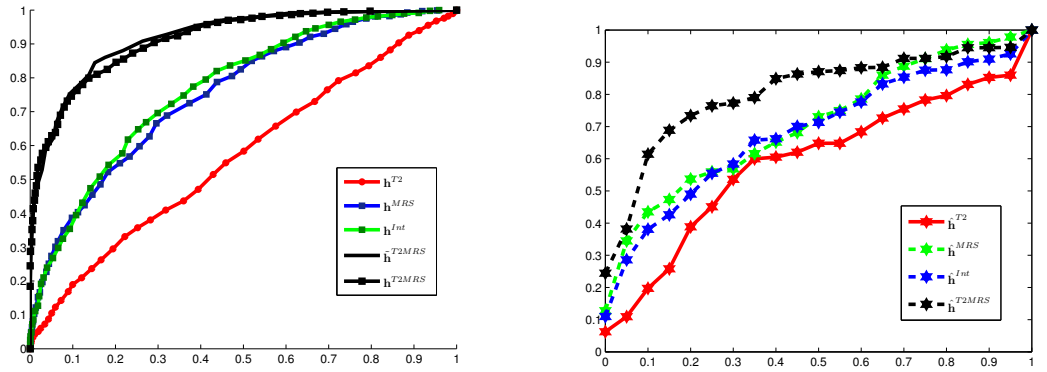
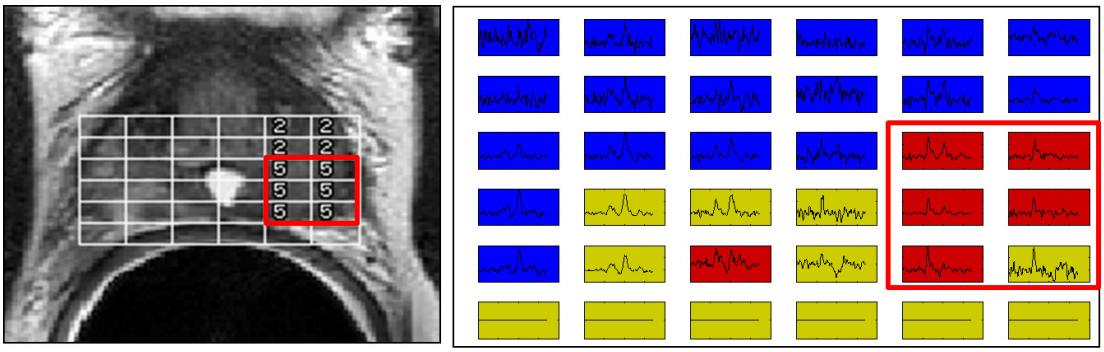


Figure 7.7: Box-and-whisker plot results of AUC (Fig. 7.7(a)) and accuracy (Fig. 7.7(b)) obtained over 12 studies via a leave-one-out cross validation strategy for $\hat{\mathbf{h}}^{T2}$, $\hat{\mathbf{h}}^{MRS}$, $\hat{\mathbf{h}}^{Int}$, $\hat{\mathbf{h}}^{IntD}$, and $\hat{\mathbf{h}}^{T2MRS}$. Corresponding results obtained via a three-fold-cross-validation are shown in Fig. 7.7(c) and (d). Note that the red line in the middle of each box reflects the median value while the box is bounded by 25 and 75 percentile of AUC (Fig. 7.7(a), (c)) and accuracy (Fig. 7.7(b), (d)) values. The whisker plot extends to the minimum and maximum values outside the box and the outliers are denoted as the red plus symbol for different feature extraction strategies.



(a) (b)

Figure 7.8: ROC curves for (a) CaP versus benign classification using $\mathbf{h}^{T2}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}^{T2MRS}$, $\tilde{\mathbf{h}}^{T2MRS}$ and (b) high versus low grade CaP using $\hat{\mathbf{h}}^{T2}$, $\hat{\mathbf{h}}^{MRS}$, $\hat{\mathbf{h}}^{Int}$, and $\hat{\mathbf{h}}^{T2MRS}$ for $d = 15$.



(a) (b)

Figure 7.9: (a) High grade CaP ground truth (outlined in red) as annotated by an expert, where label 2 = Gleason score 7 (3+4), and label 5 = Gleason score 9 (>4+4) spectra on a single T2w MRI section. Figure 7.9(b) shows the corresponding classification result obtained by thresholding the probability values at the operating point ν on MRS grid, where red corresponds to high probability of high grade CaP and blue corresponds to high probability of low grade CaP. Note the high detection sensitivity and specificity obtained via SeSMiK-GE in accurately localizing high grade CaP region. Also note the elevated choline peak in all the metavoxels identified as high grade (in red). Elevation in choline has clinically been shown to be correlated with high grade CaP

thereby allowing for the subtle differences to be accentuated. Our results are also consistent with studies that have demonstrated a higher accuracy of high grade CaP detection using MP MRI/MRS compared to individual protocols [34, 35, 36].

Figure 7.9 shows the classification result obtained for one of the T2w MRI/MRS sections, subsequently plotted back onto the corresponding MRS grid, where the red square represents the ground truth extent for high grade CaP. Figure 7.9 (b) shows the corresponding classification result obtained via SeSMiK-GE on a MRS grid, which demonstrates the high sensitivity and specificity obtained via SeSMiK-GE for high grade CaP detection. Note that the spectra identified as low grade and high grade CaP (in red) by the classifier appear to be qualitatively different in terms of the relative concentrations of choline, creatine and citrate.

Chapter 8

Quantitative Integration of MP-MRI: Application to Evaluating Treatment Related Changes in Prostate

8.1 Overview

In this paper we present a novel MP quantitative data integration scheme for evaluating pre-, post-RT related changes, that, (1) provides a common framework across pre-, post-patient studies via registration, segmentation and classification modules, to overcome the aforementioned challenges associated with pre-, post- data alignment, (2) accurately quantifies post-RT imaging marker changes ("hotspots") on a per-voxel level, on T2w MRI, MRS, and DWI, and (3) intelligently combines "changes in imaging markers" across individual MP-MRI modalities for accurately assessing pre-, post-RT changes ("hotspots"), identified as, (a) successful treatment (CaP on pre-RT, no CaP on post-RT), (b) new CaP foci (no CaP on pre-RT, CaP on post-RT) and (c) local recurrence (CaP on pre-RT, CaP on post-RT but in a different location relative to the site of treatment). These treatment changes are captured by differences in imaging markers across different imaging modalities (T2w, MRS, DWI), and are combined by optimally weighting contributions of each MP-MRI modality based on their ability to accurately capture post-RT changes.

8.2 System Overview and Data Description

8.2.1 System Overview

Figure 8.1 presents an overview of our scheme illustrating the registration, segmentation, quantification and integration modules. In Module 1, T2w MRI, MRS and DWI pre-, post-treatment are brought into alignment using a spatially constrained 3-D affine

registration scheme [152], that accounts for changes in overall shape and size of the prostate pre-, and post-RT. Module 2 involves accurate delineation of the prostate region of interest (ROI) on pre-, post-RT using a robust statistical shape model [153] that is able to compensate for the loss of image resolution post-RT. In Module 3, a difference map is generated for each of the individual T2w MRI, MRS and DWI protocols by taking an absolute difference of the corresponding imaging markers pre-, post-RT MRI on a per-voxel basis. The difference maps thus obtained on each of these imaging markers are then intelligently weighted based on their respective ability in capturing treatment related changes to yield a weighted, combined MP-MRI imaging marker difference map (Module 4).

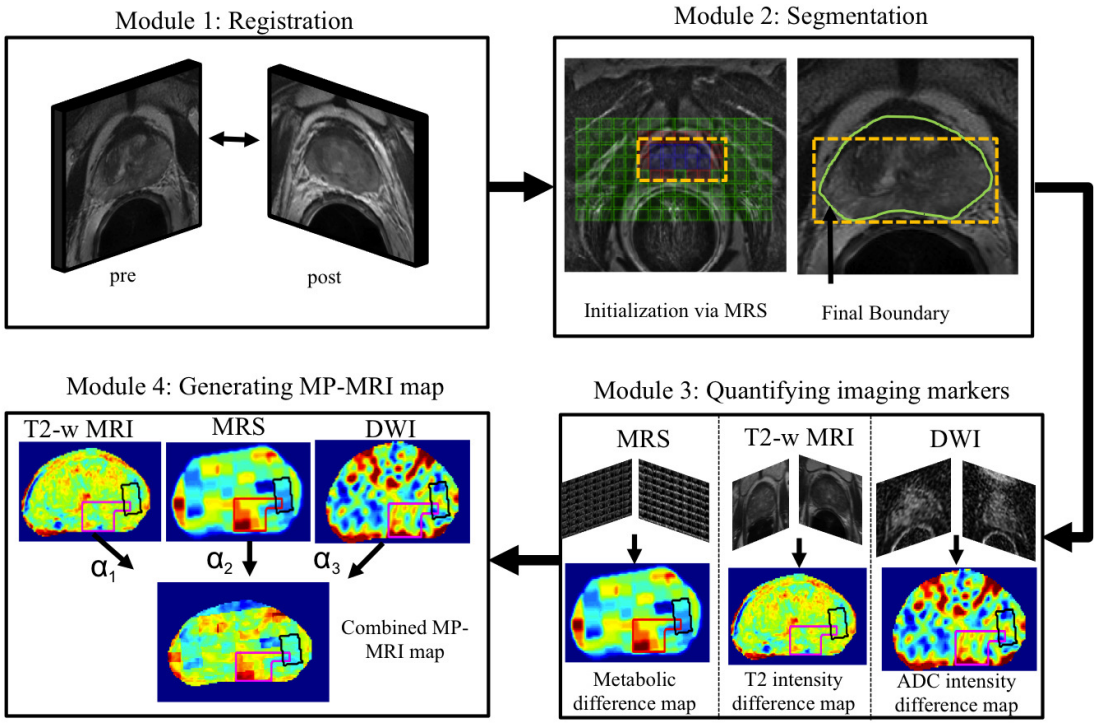


Figure 8.1: Flowchart of the new strategy showing different modules, registration (Module 1), segmentation (Module 2), quantification of imaging markers (Module 3). The individual imaging marker difference maps are integrated via a weighted combination scheme to yield an integrated difference map, one that is more reflective of the disease specific changes, pre-, post, RT (Module 4).

8.2.2 Notation

We denote a pre-treatment T2w prostate MRI scene as $\mathcal{C}_{t2}^{pre} = (C, f_{t2}^{pre})$, where $f_{t2}^{pre}(c)$ is the associated intensity at every voxel location c on a 3D grid C. MRS data corresponding to $\mathcal{C}_{t2}^{pre}$ is obtained at a much coarser resolution, such that multiple voxels are located within a single MRS voxel. To obtain a voxel based MRS representation, we linearly interpolate acquired MRS information to a T2w MRI voxel resolution, thus yielding the corresponding spectral scene $\mathcal{C}_{mrs}^{pre} = (C, G^{pre})$, where $G^{pre}(c)$ is the associated MRS metabolic vector at every voxel $c \in C$ (representing the concentrations of different biochemicals, such as creatine, citrate, and choline). The DWI MRI scene is similarly defined as $\mathcal{C}_{adc}^{pre} = (C, f_{adc}^{pre})$, where $f_{adc}^{pre}(c)$ is the associated ADC intensity at every voxel location c on a 3D grid C (interpolated to a T2w MRI voxel resolution). Post-RT T2w MRI prostate image scene, $\widehat{\mathcal{C}}_{t2}^{post}$, is registered to $\mathcal{C}_{t2}^{pre}$ to yield the corresponding registered T2w MRI scene (post-RT) as $\mathcal{C}_{t2}^{post} = (C, f_{t2}^{post})$. MRS information (interpolated to T2w voxel resolution) is transformed to yield $\mathcal{C}_{mrs}^{post} = (C, G^{post})$. DWI image is similarly transformed to yield the corresponding registered DWI image as $\mathcal{C}_{adc}^{post} = (C, f_{adc}^{post})$.

8.2.3 Data Description

A total of 14 *in vivo* endorectal MP-MRI patient datasets were acquired at the University of California, San Francisco between 1998-2007. All patients underwent external beam radiotherapy after initial MRI (1.5 Tesla, GE Signa, endorectal coil), with supplementary, neo-adjuvant hormonal therapy. Post-RT, patients were reimaged via MRI (3 Tesla, GE Signa, endorectal coil). An expert spectroscopist labeled the spectral voxels as CaP and benign on the MRI/MRS pre- and post-RT studies, which was used as surrogate ground truth labels for CaP extent. Five out of the 14 studies included T2w, MRS, and ADC maps (from DWI), while the remaining 9 comprised MRS and T2w MRI alone. A total of 51 slices from the 14 patient studies (with T2w MRI and MRS information, acquired between 1998-2002) with CaP (pre-RT) constituted Dataset 1, and a total of 20 slices with CaP (pre-RT) from the 5 patients studies (with ADC,

Treatment response	Residual CaP	New CaP occurrence	Number of studies
Successful treatment	N	N	5
Partially successful treatment	Y	N	3
Local recurrence	Y	Y	6
	N	Y	

Table 8.1: Different RT outcomes with corresponding effects on CaP presence and extent.

T2w and MRS information, acquired between 2002-2007) constituted Dataset 2. Based on the pre-, and post-RT CaP label information, regions of definite treatment change ("hotspots") were defined on each of the 51 images for Dataset 1, and 20 images for Dataset 2. These included regions of (1) successful treatment, (2) partially successful treatment, and (3) local recurrence (detailed in Table 8.1).

8.3 Registration and Segmentation of Multi-parametric MR imagery

8.3.1 Registration of pre- and post-treatment MR imagery

Registration of $\mathcal{C}_{t2}^{pre}$ to $\widehat{\mathcal{C}}_{t2}^{post}$ is uniquely complicated by (1) changes in the overall shape and size of the prostate gland (which is known to shrink, post-RT [39]), (2) differing acquisition parameters, and (3) changes in imaging markers due to RT effects. We attempt to address these challenges by employing a mutual information (MI) based 3D registration scheme [152], comprising of following steps,

1. Post-RT MP-MRI data is first down-sampled to pre-RT image resolution. Bounding boxes containing the prostate on $\mathcal{C}_{t2}^{pre}$ and $\widehat{\mathcal{C}}_{t2}^{post}$ are then manually selected.

2. A spatially constrained MI similarity measure is used to drive the affine transformation of $\widehat{\mathcal{C}}_{t2}^{post}$ onto $\mathcal{C}_{t2}^{pre}$. Only those voxels of $\mathcal{C}_{t2}^{pre}$ and $\widehat{\mathcal{C}}_{t2}^{post}$ that fall within the bounding box (selected in Step 1) are considered in the calculation of MI (chosen for its robustness to non-linear intensity relationships [152]).

3. A 3D affine transformation with 12 degrees of freedom, encoding rotation, translation, shear, and scale, is implemented (as presented in [152]) to accurately align the prostate between $\mathcal{C}_{t2}^{pre}$ and $\widehat{\mathcal{C}}_{t2}^{post}$.

Alignment of T2w and ADC maps is done based on available voxel sizes and locations (automatically extracted from DICOM headers). In the case of post-RT ADC maps, the 3D transformation from Step 3 is applied to map all the data into the pre-treatment coordinate frame C (associated with $\mathcal{C}_{t2}^{pre}$). However, alignment of T2w MRI and MRS is done in 2D (since MRS, T2 are implicitly aligned pre-registration in 2D), by applying the transformation obtained in Step 3 along the X and Y-direction to obtain $\mathcal{C}_{mrs}^{post}$.

8.3.2 Automated segmentation of prostate capsule on T2w MRI data

This module utilizes a novel, fully automated Active Shape Model (ASM) scheme for delineation of the prostate capsule on *in vivo* T2w MR imagery [153]. This technique, developed by our group and presented in [153], leverages multi-protocol data as follows,

1. First, a texture-based SVM classifier is constructed to be able to classify voxels within the prostate ROI.

2. A single midgland slice is selected from each test study. Corresponding MRS data is identified as either prostatic or extra-prostatic via a replicated k-means spectral clustering scheme [153]. This yields a bounding box of spectra from within the prostate.

3. The SVM classifier from Step 1 is used to identify prostatic voxels within the bounding box identified in Step 2, resulting in a boundary initialization.

4. The ASM transforms a known mean shape of the prostate (detailed in [153]) to the boundary initialization from Step 3, resulting in the gland capsule segmentation for this slice.

5. This segmentation is extended to the base and apex to yield a delineation of the prostate ROI (as described in [153]) on $\mathcal{C}_{t2}^{pre}$, $\mathcal{C}_{t2}^{post}$ as well as on the ADC map.

8.4 Weighted combination of imaging marker difference maps for identifying treatment related changes post-therapy

8.4.1 Feature extraction for individual T2, MRS and DWI protocols

A difference map for each of the imaging markers, extracted from each of the individual MR protocols (T2, MRS, DWI) is computed for every scene $\mathcal{C}$, for every $c \in C$ as,

$$\Delta_i = |\mathcal{C}_i^{pre} - \mathcal{C}_i^{post}|, \tag{8.1}$$

where $i \in \{1, 2, ..., n\}$ is the imaging marker evaluated, $\mathcal{C}$ is the 3D image scene associated with each imaging marker i, and n is the total number of imaging markers.

A. Structural (T2w): All T2w MRI images, pre-, post- RT, were first corrected for bias-field and underwent intensity standardization [23]. A difference map (Δ_1) is then calculated by taking an absolute difference (L_1-norm) of T2w signal intensities across pre-, post- RT image scenes at every $c \in C$, using Equation 8.1.

B. Metabolic (MRS): Kurhanewicz et al [42] have suggested that ratios of area under the choline (A_{ch}), creatine (A_{cr}) and citrate peaks (A_{cit}) are highly indicative of the presence of CaP. However, only choline (A_{ch}) and creatine (A_{cr}) are considered post-RT, due to the known absence of citrate [43]. We calculated A_{ch} and A_{cr} using a composite trapezoidal rule within the pre-defined metabolic ranges on each spectrum, both pre-, and post-RT. A ratio of A_{ch} /A_{cr} was recorded and used to obtain a difference map (Δ_2) by taking a L_1 norm of A_{ch}/A_{cr} pre-, and post-RT as illustrated in Equation 8.1.

C. Functional (ADC): DWI images were corrected for bias field and intensity standardized pre-, post-RT. Equation 8.1 is then employed to compute an ADC difference map (Δ_3) by taking an absolute difference of ADC signal intensities values across pre-, and registered post- RT on a per voxel basis.

8.4.2 Generating a combined multi-parametric weighted map

Individual difference maps ($\Delta_1, \Delta_2, \Delta_3$) obtained from T2w, MRS and DWI, allow for quantification of the changes in imaging markers across each of the individual protocols.

A MP-MRI weighted map can thus be obtained by leveraging each of the different marker difference maps as,

$$\mathcal{C}_{map} = \sum_{i=1}^{n} \alpha_i \times \Delta_i, \tag{8.2}$$

where $\alpha_i, i \in \{1, 2, .., n\}$ reflects the contribution of each of the individual n imaging markers obtained via different MP-MRI protocols (T2w MRI, ADC, MRS) employed simultaneously to accurately quantify treatment-specific changes on MP-MRI.

Optimization of weights:

Weights are obtained via a rigorous optimization of imaging marker difference maps from a set of training images ($\tilde{D}$), and the learned weights are then assigned to the test image to obtain $\hat{D}$.

1. For each of the training images, $\tilde{D}_i$, a binary mask, $\tilde{D}_i^{\theta}$, is created by thresholding the intensity values between 0 and 1, such that $\tilde{D}_i^{\theta} = \tilde{D}_i \geq \theta$ for each $i, i \in \{1, ...n\}$.

2. Sensitivity and specificity values at each threshold $\theta, \theta \in [0, 1]$ are recorded and a ROC is obtained for each $\tilde{D}_i$, for each $i, i \in \{1, ..., n\}$. Binary map at the operating point of the ROC curve $\tilde{D}_i^{\nu}$ for each i is recorded, where ν is the operating point of the ROC curve.

3. $\tilde{D}_i^{\nu}$ is then used to create a MP-MRI map as, $\tilde{D}_{map} = \sum_{i=1}^{n} \alpha_i \times \tilde{D}_i^{\nu}$, where weights are such that $\sum_{i=1}^{n} \alpha_i = 1, \alpha_i \in [0, 1]$.

	Pre-RT appearance	Post-RT appearance
T2w	low T2w signal intensity in peripheral zone	Hypo-intense regions, smooth texture No change in residual CaP regions
MRS	elevated levels of choline (A_{ch})/creatine (A_{cr}) reduced levels of citrate (A_{cit})	Nearly absent A_{cit}, polyamines Residual CaP has elevated A_{ch}, A_{cr}
DWI	significantly low ADC compared to benign	Residual CaP has increased ADC Residual CaP lower ADC

Table 8.2: Summary of qualitative changes in MP-MR imaging parameters pre- and post-RT, and the corresponding quantitative features used in this work to characterize each of the marker differences for different protocols.

4. Positive predictive value (Φ) is recorded for different values of α_i, $\alpha \in [0,1]$ based on the overlap of training MP-MRI image, $\tilde{D}_{map}$, with respect to expert delineated ground truth, D_{GT}. The values of α_i that maximize Φ in accurately identifying treatment changes are obtained as:

$$\{\tilde{\alpha}_1, ..., \tilde{\alpha}_n\} = \arg \max_{\alpha_1,...,\alpha_n} (\Phi) \tag{8.3}$$

5. The maximum likelihood estimate (MLE) of $\tilde{\alpha}_i$ (mode of the distribution assuming each $\tilde{\alpha}_i$ is normally distributed) for each $i, i \in \{1, ..., n\}$ across all training images, $\hat{\alpha}_i$, is then used to obtain the combined test image, $\hat{D} = \sum_{i=1}^{n} \hat{\alpha}_i \times \hat{D}_i$.

8.5 Results and Discussion

A leave-one-out cross validation strategy was used, where at each iteration, slices from a single patient study were held out from testing, while the remaining slices were used for training. The cross-validation process was repeated until all slices from all patient studies are evaluated. During each run of cross validation, test image, $\hat{D}$, is evaluated via ROC analysis for threshold $\theta, \theta \in [0,1]$ based on overlap of thresholded binary image with respect to the expert delineated ground truth labels, D_{GT}, on a per-voxel basis.

8.5.1 Experiment 1: Quantifying changes in individual imaging markers post-RT

Figure 8.2 shows scaled absolute difference images of T2w intensities (Figure 8.2(c), (i)), ADC values (Figure 8.2(d), (j)), ch/cr metabolites (Figure 8.2(e), (k)), and weighted MP-MRI maps (Figures 8.2(f), (l)) for single 2D slices from two different patient studies. The results were evaluated based on the overlap of ground truth labels ("hotspots") on a per-voxel basis. CaP on pre-RT is outlined in magenta, while the CaP on post-RT is outlined in black. Note that ADC (Figure 8.2(d), (j)) appears to identify more true positive regions (RT-changes) as compared to T2w MRI (Figure 8.2(c), (i)) and MRS (Figure 8.2(e), (k)). MRS (Figure 8.2(e), (k)) appears to pick fewer false positives associated with RT-related changes, as compared to ADC (Figure 8.2(d), (j)), and T2w MRI (Figure 8.2(c), (i)) across the two slices.

Method	AUC	Accuracy
T2w intensity map	49.6 ± 9.0	51.7 ± 7.0
ch/*cr* map	55.9 ± 20.0	55.2 ± 11.3
MP-MRI map	**61.6 ± 7.5**	**63.0 ± 4.3**

(a)

Method	AUC	Accuracy
T2w intensity map	54.4 ± 8.3	54.9 ± 7.8
ch/*cr* map	68.6 ± 10.3	67.5 ± 9.0
ADC map	70.8 ± 7.0	70.2 ± 10.2
MP-MRI map	**73.2 ± 6.9**	**72.5 ± 8.0**

(b)

Table 8.3: Average AUC and accuracy obtained via leave-one-out cross validation for quantifying pre-, post-RT changes, obtained from T2w image intensity map, *ch*/*cr* map, ADC difference map and the weighted MP map for, (a) Dataset 1 (T2w MRI and MRS) over a total of 14 T2w MRI-MRS studies and, (b) Dataset 2 (T2w, MRS and DWI) over a total of 5 patient studies. Note that Dataset 2 which was acquired later compared to Dataset 1 used a more optimized set of MRI protocols, resulting in superior image quality.

Table 8.3(a) shows the mean AUC and accuracy for Dataset 1 across 14 patient studies (with only T2w MRI and MRS data), while Table 8.3(b) shows the mean AUC and accuracy obtained for Dataset 2 (with DWI, T2w MRI and MRS) for 5 patient studies, obtained via a leave-one-out cross validation. It is worth noting that ADC difference maps outperformed T2w MRI and MRS difference maps, with an AUC of 70.2%, compared to 67.5% for MRS and 54.9% for T2w MRI, in accurately quantifying RT-related changes. The qualitative and quantitative results presented in this work corroborate with the findings in [44] which suggests that difference in ADC values might be an important imaging marker in evaluating pre-, post-RT changes.

8.5.2 Experiment 2: Quantifying changes via weighted combination of MP-MRI

The results presented in Figure 8.2 and Table 8.3 suggest that MP-MRI map outperformed each of the individual imaging marker differences $(\Delta_1, \Delta_2, \Delta_3)$ across both the datasets, in accurately quantifying treatment changes. The improved performance of combined MP-MRI map clearly indicates the efficacy of optimal weighted combination of imaging markers in identifying treatment specific changes pre-, post RT. It is also interesting to note that a much higher AUC was obtained for Dataset 2 (AUC = 73.2%),

as compared to Dataset 1 (AUC = 61.6%), when ADC was incorporated along with T2w MRI and MRS to create the weighted MP-MRI map. This is not surprising as a number of groups [154, 155, 156] have shown that that the inclusion of DWI in addition to the other MRI protocols, significantly improves CaP detection. Since Dataset 2 only comprised 5 patient studies, AUC values obtained via MP-MRI were not found to be statistically significantly different compared to AUC values obtained from DWI different maps using a paired student t-test.

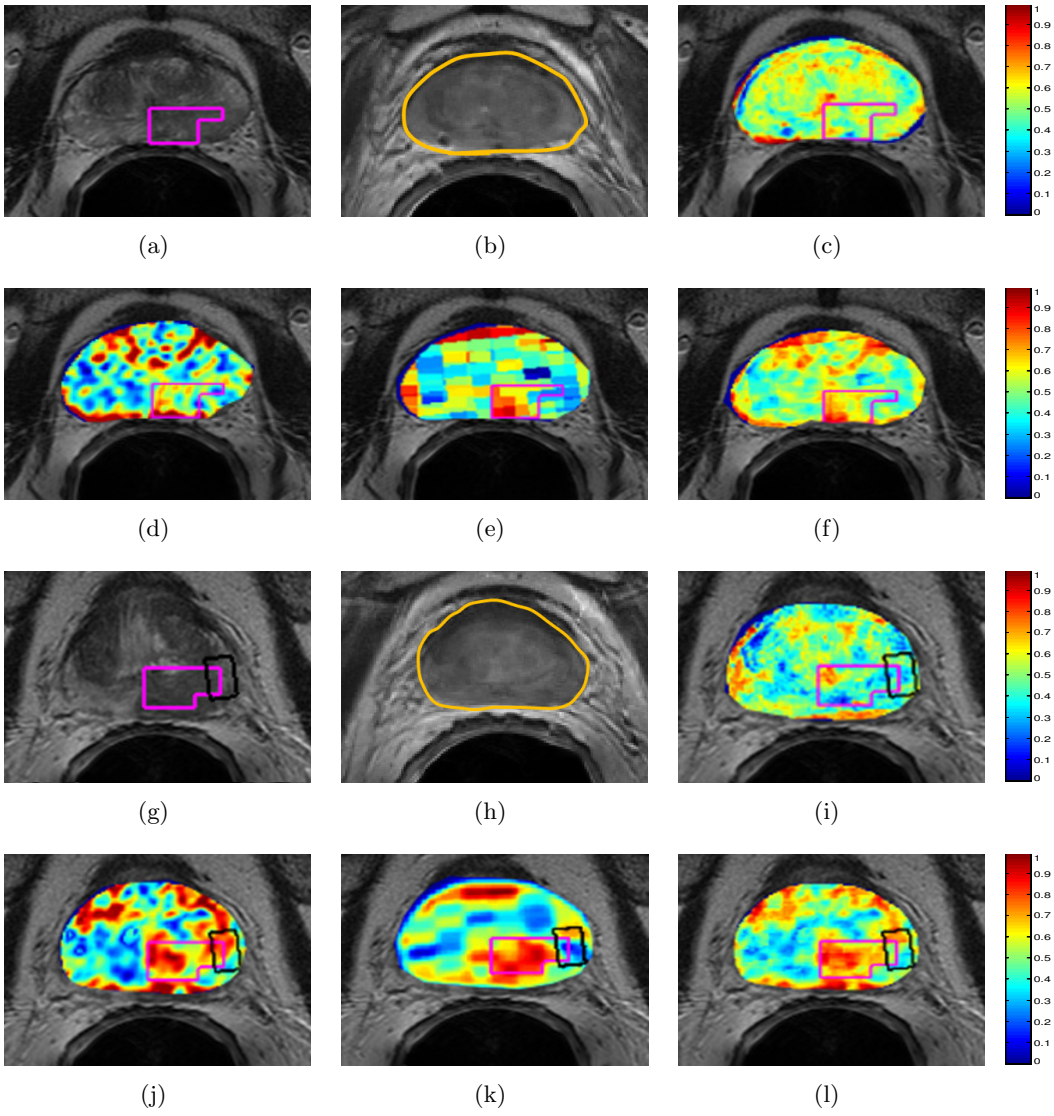


Figure 8.2: (a), (g) show pre-RT T2w MRI image with the pre, post-RT CaP labels delineated in magenta (pre-RT CaP) and black (post-RT CaP) respectively. (b), (h) demonstrate the corresponding registered post-RT T2w MRI image with the segmented prostate region shown with the orange boundary. (c), (i) correspond to the scaled absolute T2w MRI image intensity difference heat maps. (d) and (j) show the corresponding difference heatmaps obtained by taking a difference of ADC values pre-, post- RT. Similarly, (e) and (k) show the heatmaps for metabolic marker A_{ch}/A_{cr}, and (f) and (l) show the corresponding weighted MP-MRI maps for two single slices from two different patient studies.

Chapter 9

Concluding Remarks

In this thesis, we have presented a suite of novel feature extraction and data integration strategies for quantitative integration of imaging and non-imaging data. Specific applications of this work were shown in the context of integrating multi-parametric MRI for (a) prostate cancer diagnosis,(b) Gleason grading, and (c) evaluating radiation-therapy treatment related changes in the prostate. The major goals accomplished in this work include,

- Developing and evaluating novel data representation strategies to provide a common framework for data representation across imaging and non-imaging protocols.
- Theoretic and algorithmic development of two novel data integration strategies, (1) Multi-modal Wavelet Embedding Representation for data Combination (MaWERiC), and (2) Semi-Supervised Multi-Kernel Graph Embedding (SeSMiK-GE), which overcome the differences in resolution and dimensionalities across heterogeneous multi-parametric data. MaWERiC assumes equal contributions across the protocols, while SeSMiK-GE is a weighted data combination strategy, which optimizes weights across the protocols depending on their diagnostic capabilities.
- Application of MaWERiC and SeSMiK for integrating structural and metabolic information from MP MRI/MRS for (a) identifying high probability CaP regions, and (b) further classifying the high probability CaP regions as being high or low grade CaP. To the best of our knowledge, our work is the first application of a computerized decision support classifier for identifying high grade CaP using MP MRI.
- Developing and evaluating a MP-MRI map which optimally weights contributions

from differences of individual imaging markers in accurately evaluating pre-, post-RT prostate cancer via MP-MRI.

The C-LLE algorithm was introduced in Chapter 3, which aims to overcome the limitations of traditional LLE by obtaining a stable low dimensional representation of high dimensional data by integrating object adjacencies from across multiple low dimensional unstable data projections. The spirit behind C-LLE, is rooted in the classifier ensembles literature where multiple weak classifiers with high variance and bias are combined to create a strong classifier with low bias and variance.

In Chapter 5, we demonstrated the excellent performance of C-LLE algorithm for CaP detection on prostate MRS compared to ICA based peak detection scheme, PCA, z-score, and traditional LLE. One issue that was not addressed in this work is the intrinsic dimensionality at which to combine the multiple weak embeddings in C-LLE. While, C-LLE is computationally more expensive compared to traditional LLE, it is not computationally prohibitive except for very large or dense datasets. In future work we will address some of these issues and also explore the applicability of C-LLE to other classification problems.

In Chapter 4, two novel data integration strategies, MaWERiC and SeSMiK were introduced.

MaWERiC is specifically geared towards quantitative integration of imaging and non-imaging data. MaWERiC comprises of two transformation modules, (i) wavelet transformation and (ii) principal component analysis which together provide a platform for uniform and homogeneous data integration across modalities. The homogeneous, low-dimensional representation of disparate data sources obtained via MaWERiC is then combined in the Eigen space.

In Chapter 6, a three-fold cross-validation performed over 25 iterations and the corresponding pairwise t-test on a total of 36 1.5 Tesla in-vivo T2-w MRI, MRS studies demonstrated that the MaWERiC classifier significantly outperforms (a) either modality individually, (b) decision combination obtained by combining individual classifier decisions from both modalities, and (c) a classifier combining metabolite peak area and

ratio features from MRS and T2-w MR image intensities.

SeSMiK-GE is a slight improvement over MaWERiC in that it overcomes the differences in resolution and dimensionalities across heterogeneous multi-parametric data, while linearly combining different protocols, by leveraging (1) multi-kernel learning, (2) semi-supervised learning, and (3) dimensionality reduction methods under a single unified framework. In Chapter 7, quantitative evaluation of the SeSMiK-GE classifier revealed a significantly higher detection accuracy in identifying both CaP and high-grade CaP regions as compared to (a) individual uni-modal T2w MRI, MRS modalities, (b) decision combination obtained by combining individual classifier decisions from both modalities, and (c) a classifier combining MRS features and T2w MR texture features.

In Chapter 8, we extended our integrated MP-MRI approach to optimally weighting contributions from differences of individual imaging markers in accurately evaluating pre-, post-RT CaP via MP-MRI. Different MRI protocols from pre- and post-RT MRI scans were first affinely registered. Functional, structural, and metabolic difference maps were then obtained individually using DWI, T2w, and MRS respectively, by taking a scaled absolute difference of the imaging markers pre-, post- RT. A combined weighted MP-MRI map is then created by leveraging differences across multiple imaging markers. Quantitative evaluation against expert delineated ground truth of treatment changes, yielded a high AUC and accuracy for the weighted MP-MRI map as compared to individual imaging markers. We believe that such an accurate per-voxel based quantitative evaluation of treatment changes pre-, post-RT will have a high clinical impact in monitoring treatment effectiveness, and could be used to modify treatment regimen early, in cases of studies with new foci or recurrence of CaP.

To conclude, MP-MRI holds great potential as a screening tool for prostate cancer diagnosis, prognosis, and has more recently been found to be effective in identifying treatment related changes in the prostate. However, there is still a great need of computerized decision support tools to optimize and quantify imaging signatures obtained across the different MP-MRI imaging protocols to assist radiologists in making a more informed and improved disease diagnosis. The methods developed in this work will thus

have significant translational and clinical benefits towards improved, early CaP diagnosis, prognosis and early identification of treatment related changes in the prostate.

Chapter 10

Future Work

In future work, we intend to extend the strategies developed in this work to more generalizable imaging and non-imaging, multi-scale, multi-modality protocols for characterizing different diseases. Below we more specifically describe the future work associated with each of the goals presented in this work. Towards goals 1 and 2, we developed two novel strategies MaWERiC and SeSMiK for combining T2w MRI and MRS for prostate cancer diagnosis and prognosis. Both SeSMiK and MaWERiC are developed to provide a general framework for potentially integrating any combination of heterogeneous data modalities, independent of scales and dimensions. In future work, we will look at applying and extending MaWERiC and SeSMiK in the context of other biomedical applications such as integration of -omics with imaging data for improved disease characterization.

The SeSMiK methodology described in this dissertation , to the best of our knowledge, is the first CDS strategy for distinguishing high versus low grade CaP patients, and could have far reaching implications for CaP patients trying to decide on the appropriate treatment option. The ability to identify low grade disease *in-vivo* might allow CaP patients to opt for active surveillance rather than immediately opting for aggressive therapy. Recent studies have demonstrated a correlation between ADC values and Gleason grade [157]. While DWI was not considered while implementing SeSMiK in this work, the SeSMiK-GE framework could be easily extended to accommodate additional imaging protocols. We are currently working on extending SeSMiK to intelligently combining T2w MRI, MRS and ADC maps for predicting Gleason grade *in-vivo*.

We additionally demonstrated the utility of MP-MRI signatures (T2w, MRS and DWI) for evaluating treatment-related changes in CaP patients who opt for IMRT as

a treatment modality. In future work, we aim to incorporate additional MR protocols (such as DCE MRI) to further improve efficacy of quantitative treatment evaluation using MP-MRI. We also aim to evaluate our quantitative scheme for other treatment modalities such as proton beam therapy and brachytherapy to quantitatively compare the different treatment modalities in accurately quantifying treatment related changes. The work presented here should set the stage for ultimately developing image based predictors for early treatment response and potentially long-term patient outcome. Additionally, the framework could be applied in the context of clinical trials for evaluating the comparative effectiveness of different prostate cancer treatment modalities.

Chapter 11

Appendix

Abbreviation	Description
MRI	Magnetic Resonance Imaging
MP-MRI	Multi-parametric MRI
CaP	Prostate Cancer
MaWERiC	Multimodal Wavelet Embedding Representation for data Combination
SeSMiK	Semi-Supervised Multi-Kernel
RT	Radiation Therapy
MRS	Magnetic Resonance Spectroscopy
CDS	Computerized Decision Support
DWI	Diffusion Weighted Imaging
ADC	Apparent diffusion coefficient
IMRT	Intensity Modulated Radiation Therapy
COD	Combination of Data
COI	Combination of Interpretation
DR	Dimensionality Reduction
MKL	Multi-kernel Learning
NLDR	Non-linear dimensionality reduction
PCA	Principal Component Analysis
ICA	Independent Component Analysis
LLE	Locally Linear Embedding
C-LLE	Consensus- Locally Linear Embedding
MLE	Maximum Likelihood Estimation
GE	Graph embedding
SVM	Support Vector Machine
PBT	Probabilistic Boosting Tree
RF	Random Forest classifier
PPMM	Pairwise Probabilistic Markov Model
SSDR	Semi-supervised dimensionality reduction
SSGE	Semi-supervised graph embedding
LOO	Leave one out
CV	Cross validation
IMRT	Intensity modulated radiation therapy

Table 11.1: List of abbreviations used in this dissertation.

Symbol	Description
$\mathcal{F}$	data matrix
η_h	high-pass filter
η_l	low-pass filter
$\mathcal{H}_h$	high-filter coefficients
$\mathcal{L}_l$	low-filter coefficients
M	dimensionality of spectrum
$z(c)$	z-score at every $c \in C$
$S^{LLE}(c)$	low-dimensional embedding vector obtained via LLE
$S_\kappa(c)$	low-dimensional embedding at κ
Φ_B	Texture feature operator
H	Co-occurrence matrix
V_t	weak clustering at iteration t
$\hat{V}$	stable clustering
κ	Neighborhood parameter
$\phi(c, \mathcal{S}_k)$	weak predictor
$\mathcal{S}_k$	Bootstrapped training set
$\phi^{Bag}(c)$	Strong bagged classifier
$\hat{S}(c)$	stable consensus embedding
$\boldsymbol{F}^{T2w}$	Gabor feature vector for T2w MRI
$\boldsymbol{F}^{T2\tau_1}$	non-steerable Gradient T2w MRI feature vector
$\boldsymbol{F}^{T2\tau_2}$	first order statistical T2w MRI feature vector
$\boldsymbol{F}^{T2\tau_3}$	Haralick T2w MRI feature vector
$\boldsymbol{F}^{T2t}$	T2w feature vector $\boldsymbol{F}^{T2t}= [\boldsymbol{F}^{T2w}, \boldsymbol{F}^{T2\tau_1}, \boldsymbol{F}^{T2\tau_2}, \boldsymbol{F}^{T2\tau_3}]$
$\boldsymbol{F}^{MRS}$	Metabolic vector $\boldsymbol{F}^{MRS} =[A_{ch}, A_{cr}, A_{cit}, A_{ch}/A_{cr}, A_{ch+cr}/A_{cr}$
$\boldsymbol{F}^{MRSw}$	MRS wavelet feature vector
$\boldsymbol{F}^{T2w}_{PCA}$	T2w MRI Eigen feature vector obtained via PCA on $\boldsymbol{F}^{T2w}$
$\boldsymbol{F}^{MRSw}_{PCA}$	MRS Eigen feature vector obtained via PCA on $\boldsymbol{F}^{MRSw}$
$\boldsymbol{F}^{Int}$	Combined feature vector $\boldsymbol{F}^{Int} = [\boldsymbol{F}^{T2w}, \boldsymbol{F}^{MRS}]$
$\boldsymbol{F}^{Int}_{PCA}$	Wavelet PCA based MaWERiC feature vector
$\hat{\mathcal{C}}$	3D MRI scene
$\mathcal{C}$	3D MR spectral scene
$\hat{C}$	3D grid of MRI voxels
C	3D grid of metavoxels
$\hat{c}$	Voxel location in $\hat{C}, \hat{c} \in \hat{C}$
c	A metavoxel in $C, c \in C$

Symbol	Description
$g_u(c)$	mean feature value at metavoxel c
G	Undirected weighted graph
K	Kernel matrix
ϕ	Pairwise kernel mapping
$\mathcal{F}_m$	Original data matrix for protocol m
$F(c)$	Feature vector at metavoxel c
D	High dimensional feature space
d	Low dimensional feature space
$\mathcal{D}$	Diagonal weight matrix
L	Laplacian matrix, $L = \mathcal{D} - W$
n	Number of data points
y	Output embedding in original space
W	Similarity weight matrix
$\boldsymbol{\alpha}$	Output embedding in kernel space
λ	Eigenvalues
ω_l	class label for point l, $\omega \in \{0, 1\}$
$\mathcal{N}$	Neighborhood parameter for SSDR
μ^{AUC}	mean classification AUC
μ^{Acc}	mean classification accuracy
$\tilde{W}_m$	Modified weight matrix for modality m
M	Number of modalities/kernels
β_m	weight for kernel m, $\beta \in [0, 1]$
A_{ch}	Area under choline peak
A_{cr}	Area under creatine peak
A_{cit}	Area under citrate peak
$\hat{\mathcal{D}}$	Combined diagonal matrix
$\hat{K}$	Combined kernel matrix
$\hat{W}$	Combined weight matrix
K_{MRI}	MRI kernel matrix
K_{MRS}	MRS kernel matrix

Symbol	Description
$\boldsymbol{f}_a(c)$	MRS signal intensity at c
a	Frequency index
$\hat{\mathcal{G}}$	T2w MRI feature scene
Φ_u	Feature operator, $u \in \{1, ..., 34\}$
δ	Optimization parameter
σ	Scaling parameter
$\hat{\boldsymbol{\beta}}$	Subset of optimal weights $[\hat{\beta_m}]$
$\bar{\boldsymbol{\alpha}}$	Fused low dimensional representation
$\boldsymbol{F}^{MRS}$	MRS feature vector for each $c \in C$
$\boldsymbol{F}^{T2MRS}$	Low dimensional SeSMiK-GE feature vector
$\boldsymbol{G}$	Average T2w feature vector
$\boldsymbol{G}^{T2}$	Low dimensional T2w MRI feature vector
$\mathbf{h}_\rho(c)$	Binary CaP classifier output at ρ
$\tilde{\mathbf{h}}_\rho(c)$	Binary MRF classifier output at ρ
$\hat{\mathbf{h}}_\rho(c)$	High grade CaP classifier output at ρ
ν	Threshold for hierarchical classification
Θ	Operating point of ROC curve
$\mathcal{C}_q^{pre}$	pre-treatment MRI scene, $q \in \{T2, mrs, adc\}$
$\widehat{\mathcal{C}}_q^{post}$	post-treatment MRI scene, $q \in \{T2, mrs, adc\}$
$\mathcal{C}_q^{post}$	registered post-treatment MRI scene, $q \in \{T2, mrs, adc\}$

Table 11.2: List of commonly used notation and symbols in this dissertation.

References

[1] P. Tiwari, M. Rosen, and A. Madabhushi. Consensus-locally linear embedding (c-lle): Application to prostate cancer detection on magnetic resonance spectroscopy. In *Proc. MICCAI*, volume 5241, pages 330–338, 2008.

[2] P. Tiwari, A. Madabhushi, et al. Spectral embedding based probabilistic boosting tree (ScEPTre): classifying high dimensional heterogeneous biomedical data. In *MICCAI*, volume 12, pages 844–51, 2009.

[3] P. Tiwari, A. Madabhushi, et al. Semi supervised multi kernel (SeSMiK) graph embedding: identifying aggressive prostate cancer via magnetic resonance imaging and spectroscopy. In *MICCAI*, volume 13(3), pages 666–73, 2010.

[4] P. Tiwari, S. Viswanath, J. Kurhanewicz, and A. Madabhushi. Weighted combination of multi-parametric mri for evaluating radiation therapy changes in the prostate. In *Workshop on Prostate Cancer Imaging, Proc. MICCAI*, 2011.

[5] P. Tiwari, S. Viswanath, G Lee, and A. Madabhushi. Multi-modal data fusion schemes for integrated classification of imaging and non-imaging biomedical data. In *Proc. ISBI*, 2011.

[6] P. Tiwari, S. Viswanath, J. Kurhanewicz, A. Sridhar, and A. Madabhushi. Multimodal wavelet embedding representation for data combination (MaWERiC): integrating magnetic resonance imaging and spectroscopy for prostate cancer detection. *NMR Biomed*, page Accepted, 2011.

[7] P. Tiwari, J. Kurhanewicz, and A. Madabhushi. Multi-kernel graph embedding for detecting and grading prostate cancer in vivo using multi-protocol magnetic resonance imaging. *Medical Image Analysis*, 2012 (under review).

[8] P. Tiwari, S. Viswanath, J. Kurhanewicz, and A. Madabhushi. Combining multi-parametric mri markers for evaluating different radiation therapy related changes in the prostate. *International Journal of Radiation Oncology, Biology and Physics*, 2012 (under review).

[9] A. Madabhushi, S. Doyle, G. Lee, A. Basavanhally, J. Monaco, S. Masters, J. Tomaszewski, and M. Feldman. Integrated diagnostics: a conceptual framework with examples. *Clin Chem Lab Med*, 48(7):989–98, 2010.

[10] V. Braun, S. Dempf, R. Tomczak, A. Wunderlich, R. Weller, and H. P. Richter. Multimodal cranial neuronavigation: direct integration of functional magnetic resonance imaging and positron emission tomography data: technical note. *Neurosurgery*, 48(5):1178–81; discussion 1181–2, 2001.

[11] Ian Chan, William Wells Iii, Robert V. Mulkern, Steven Haker, Jianqing Zhang, Kelly H. Zou, Stephan E. Maier, and Clare M. C. Tempany. Detection of prostate cancer by integration of line-scan diffusion, t2-mapping and t2-weighted magnetic resonance imaging; a multichannel statistical classifier. *Medical Physics*, 30(9):2390–2398, 2003.

[12] G. Hong and Y. Zhang. Comparison and improvement of wavelet-based image fusion. *International Journal of Remote Sensing*, 29(3):673–691, 2008.

[13] G. Villeirs, W. Oosterlinck, E. Vanherreweghe, and G. De Meerleer. A qualitative approach to combined magnetic resonance imaging and spectroscopy in the diagnosis of prostate cancer. *Eur J Radiol*, 73(2):352–6.

[14] P. Carroll, F. Coakley, and J. Kurhanewicz. Magnetic resonance imaging and spectroscopy of prostate cancer. *Reviews in Urology*, 8, 2006. S4-S10.

[15] A. Heerschap, G. J. Jager, M. Van der Graaf, J. O. Barentsz, J. Rosetten, G. De La Oosterhof, E. Ruijter, and J. Ruijs. In vivo proton mr spectroscopy reveals altered metabolite content in malignant prostate tissue. *Anticancer Research*, 17:1455–1460, 1997.

[16] Eberhardt S. Zakian, K. L., H. Hricak, A. Shukla-Dave, S. Kleinman, M. Muruganandham, K. Sircar, M. W. Kattan, Scardino P. T. Reuter, V. E., and J. A. Koutcher. Transition zone prostate cancer: Metabolic characteristics at h mr spectroscopic imaging initial results. *Radiology*, 229:241–247, 2003.

[17] A. Wetter, T. Engl, D. Nadjmabadi, K. Fliessbach, T. Lehnert, J. Gurung, W. Beecken, and J. Vogl. Combined mri and mr spectroscopy of the prostate before radical prostatectomy. *AJR, Am. J. Roentgenol*, 187:724–730, 2006.

[18] L. Wang, H. Hricak, M. W. Kattan, H. N. Chen, P. T. Scardino, and K. Kuroiwa. Prediction of organ-confined prostate cancer: incremental value of mr imaging and mr spectroscopic imaging to staging nomograms. *Radiology*, 238(2):597–603, 2006.

[19] J. Luts, T. Laudadio, A. J. Idema, A. W. Simonetti, A. Heerschap, D. Vandermeulen, J. A. Suykens, and S. Van Huffel. Nosologic imaging of the brain: Segmentation and classification using mri and mrsi. *NMR in Biomedicine*, 22:374–390, 2009.

[20] W. Arjan, J. Willem, S. Fabien, J. Jack, H. Arend, and M. Lutgarde. Combination of feature-reduced mr spectroscopic and mr imaging data for improved brain tumor classification. *NMR in Biomedicine*, 18(1):34–43, 2005.

[21] F. May, T. Treumann, P. Dettmar, R. Hartung, and J. Breul. Limited value of endorectal magnetic resonance imaging and transrectal ultrasonography in the staging of clinically localized prostate cancer. *BJU International*, 87(1):66–69, 2001.

[22] M. McLean, J. Srigley, D. Banerjee, P. Warde, and Y. Hao. Interobserver variation in prostate cancer gleason scoring: Are there implications for the design of clinical trials and treatment strategies. *Clinical Oncology*, 9(4):222–225, 1997.

[23] A. Madabhushi, M. D. Feldman, D. N. Metaxas, J. Tomaszeweski, and D. Chute. Automated detection of prostatic adenocarcinoma from high-resolution ex vivo mri. *IEEE Transactions of Medical Imaging*, 24:1611–1625, 2005.

[24] P. Vos, T Hambrock, J. Barenstz, and H. Huisman. Computer-assisted analysis of peripheral zone prostate lesions using t2-weighted and dynamic contrast enhanced t1-weighted mri. *Physics in Medicine and Biology*, 55(6):1719.

[25] D. Gleason. Classification of prostatic carcinomas. *Cancer Chemotherapy Reports*, 50:125–128, 1966.

[26] J. Epstein. An update of the gleason grading system. *Journal of Eurology*, 13:57–59, 2006.

[27] D. Bostwick. Grading prostate cancer. *American Journal of Clinical Pathology*, 102:38–56, 1994.

[28] H. Welch, E. Fisher, D. Gottlieb, and M. Barry. Detection of prostate cancer via biopsy in the medicare-seer population during the psa era. *Journal of the National Cancer Institute*, 99:1395–1400, 2007.

[29] A. Potosky, W. Davis, R. Hoffman, J. Stanford, R. Stephenson, D. Penson, and L. Harlan. Five-year outcomes after prostatectomy or radiotherapy for prostate cancer: The prostate cancer outcomes study. *J. Natl. Cancer Inst.*, 96:1358–1367, 2004.

[30] E. A. Klein. What is insignificant prostate carcinoma. *Cancer*, 101:1923–1925, 2004.

[31] Klotz. L. Active surveillance for prostate cancer: For whom? *Journal of Clinical Oncology*, 23:8165–8169, 2005.

[32] J. Kurhanewicz, M. Swanson, S. Nelson, and D. Vigneron. Combined magnetic resonance imaging and spectroscopic imaging approach to molecular imaging of prostate cancer. *Journal of Magnetic Resonance Imaging*, pages 451–463, 2002. 16.

[33] J. Kurhanewicz, J. Vigneron, H. Hricak, P. Narayan, P. Carroll, and S. Nelson. Three dimensional h-1 mr spectroscopic imaging of the in situ human prostate with high (0.24-0.7-cm3) spatial resolution. *Radiology*, 198:795–805, 1996.

[34] D. L. Langer, T. H. van der Kwast, A. J. Evans, A. Plotkin, J. Trachtenberg, and Haider M. A. Wilson, B. C. Prostate tissue composition and mr measurements: Investigating the relationships between adc, t2, ktrans, ve, and corresponding histologic features. *Radiology*, 255:485–494, 2010.

[35] Shukla-Dave A., Hricak H., Kattan M., Pucar D., Kuroiwa K., Chen H., Spector J., Koutcher J., Zakian K., and Scardino P. The utility of magnetic resonance imaging and spectroscopy for predicting insignificant prostate cancer: An initial analysis. *BJU International*, 99:786–793, 2007.

[36] A. Shukla-Dave, H. Hricak, N. M. Ishill, C. Moskowitz, M. Drobnjak, V. Reuter, K. Zakian, P. Scardino, and C. Cordon-Cardo. Correlation of mr imaging and mr spectroscopic imaging findings with ki-67, phospho-akt, and androgen receptor expression in prostate cancer. *Radiology*, 250:803–812, 2009.

[37] K. L. Zakian, K. Sircar, H. Hricak, H. N. Chen, A. Shukla-Dave, S. Eberhardt, M. Muruganandham, L. Ebora, M. Kattan, V. Reuter, P. Scardino, and J. Koutcher. Correlation of proton mr spectroscopic imaging with gleason score based on septsection pathologic analysis after radical prostatectomy. *Radiology*, 234:804–814, 2005.

[38] L. Wang, Y. Mazaheri, J. Zhang, N. M. Ishill, and Hricak H. Kuroiwa, K. Assessment of biologic aggressiveness of prostate cancer: Correlation of mr signal intensity with gleason grade after radical prostatectomy. *Radiology*, 246:168–176, 2008.

[39] D. Pucar et al. The role of imaging in the detection of prostate cancer local recurrence after radiation therapy and surgery. *Curr Opin Urol*, 18(1):87–97, 2008.

[40] B. Pickett et al. Use of MRI and spectroscopy in evaluation of external beam radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys*, 60(4):1047–55, 2004.

[41] Park BK Park W. Song I, Kim CK. Assessment of response to radiotherapy for prostate cancer: value of diffusion-weighted mri at 3 t. *AJR Am J Roentgenol.*, 194(6):477–82., 2010.

[42] J. Kurhanewicz et al. Locally recurrent prostate cancer after external beam radiation therapy: Diagnostic performance of 1.5-t endorectal mr imaging and mr spectroscopic imaging for detection. *Radiology*, 256(2):485–492, 2010.

[43] F. V. Coakley et al. Endorectal MR imaging and MR spectroscopic imaging for locally recurrent prostate cancer after external beam radiation therapy: preliminary experience. *Radiology*, 233(2):441–8, 2004.

[44] I. Song et al. Assessment of Response to Radiotherapy for Prostate Cancer: Value of Diffusion-Weighted MRI at 3 T. *Am. J. Roentgenol.*, 194(6):W477–482, 2010.

[45] I. Chan, W. Wells III, R. Mulkern, S. Haker, J. Zhang, K. Zou, S. Maier, and C. Tempany. Detection of prostate cancer by integration of line-scan diffusion, t2-mapping and t2-weighted magnetic resonance imaging; a multichannel statistical classifier. *Medical Physics*, 30:2390–2398, 2003.

[46] X. Liu, D. L. Langer, M. A. Haider, Y. Yang, M. N. Wernick, and I. S. Yetik. Prostate cancer segmentation with simultaneous estimation of markov random field parameters and class. *IEEE Trans Med Imaging*, 28(6):906–915, 2009.

[47] D. Langer, T. Kwast, A. Evans, J. Trachtenberg, B. Wilson, and M. 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*, 30:327–334, 2009.

[48] A.C. Westphalen et al. T2-Weighted endorectal magnetic resonance imaging of prostate cancer after external beam radiation therapy. *Int Braz J Urol*, 35(2):171–80, 2009.

[49] A. Zapotoczna et al. Current role and future perspectives of magnetic resonance spectroscopy in radiation oncology for prostate cancer. *Neoplasia*, 9(6):455–63, 2007.

[50] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.

[51] J. Tenenbaum, V. De Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2323–2326, 2000.

[52] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:888–905, 2000.

[53] S. Viswanath, B. Bloch, E. Genega, N. Rofsky, R. Lenkinski, J. Chappelow, R. Toth, and A. Madabhushi. A Comprehensive Segmentation, Registration, and Cancer Detection Scheme on 3 Tesla In Vivo Prostate DCE-MRI. In *Proc. MICCAI*, pages 662–669, 2008.

[54] G. Lee, C. Rodriguez, and A. Madabhushi. Investigating the Efficacy of Nonlinear Dimensionality Reduction schemes in Classifying Gene- and Protein-Expression Studies. *IEEE J CBB*, 5(3):1–17, 2008.

[55] M. Sugiyama, T. Idé, S. Nakajima, and J. Sese. Semi-supervised local fisher discriminant analysis for dimensionality reduction. *Advances in Knowledge Discovery and Data Mining*, pages 333–344, 2008.

[56] D. Sun and D. Zhang. A new discriminant principal component analysis method with partial supervision. *Neural Processing Letters*, 30:103–113, 2009.

[57] H. Zhao. Combining labeled and unlabeled data with graph embedding. *Neurocomputing*, 69:2385–2389, 2006.

[58] T. Rohlfing, A. Pfefferbaum, E. Sullivan, and CR. Maurer. Information fusion in biomedical image analysis: Combination of data vs combination of interpretations. *Information Processing in Medical Imaging*, 3565:150–161, 2005.

[59] G. Lee, S. Doyle, J. Monaco, A. Madabhushi, M. D. Feldman, S. R. Master, and J. Tomaszewski. A knowledge representation framework for integration, classification of multi-scale imaging and non-imaging data: Preliminary results in predicting prostate cancer recurrence by fusing mass spectrometry and histology. *in Proc. ISBI*, pages 77–80, 2009.

[60] R. Verma, E. Zacharaki, Y. Ou, H. Cai, S. Chawla, E. Lee, S. Melhem, R. Wolf, and C. Davatzikos. Multiparametric tissue characterization of brain neoplasms and their recurrence using pattern classification of mr images. *Academic Radiology*, 15(8):966–977, 2008.

[61] G. Lanckriet, M. Deng, N. Cristianini, M. Jordan, and W. Noble. Kernel-based data fusion and its application to protein function prediction in yeast. *Pacific Symposium on Biocomputing*, 9:300–311, 2004.

[62] Y. Lin, Tyng-Luh L., and Chiou-Shann F. Multiple kernel learning for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99, 2010.

[63] R. Duda, P. Hart, and D. Stork. Pattern classification (2nd edition). *New York: Wiley-Interscience*, 2000. 680.

[64] B. Scholkopf and A. Smola. Learning with kernels: Support vector machines, regularization, optimization, and beyond. *MIT Press*, 2001.

[65] M. S. Cohen, R. M. DuBois, and M. M. Zeineh. Rapid and effective correction of rf inhomogeneity for high field magnetic resonance imaging. *Hum Brain Mapp*, 10(4):204–11, 2000.

[66] J. J. Futterer, S. W. Heijmink, T. W. Scheenen, J. Veltman, H. J. Huisman, P. Vos, C. A. Hulsbergen-Van de Kaa, J. A. Witjes, P. F. Krabbe, A. Heerschap, and J. O. Barentsz. Prostate cancer localization with dynamic contrast-enhanced mr imaging and proton mr spectroscopic imaging. *Radiology*, 241(2):449–58, 2006.

[67] Jonathan Chappelow, B. Nicolas Bloch, Neil Rofsky, Elizabeth Genega, Robert Lenkinski, William DeWolf, and Anant Madabhushi. Elastic registration of multimodal prostate mri and histology via multiattribute combined mutual information. *Medical Physics*, 38(4):2005–2018.

[68] H. Ratiney, M. Sdika, Y. Coenradie, S. Cavassila, D. van Ormondt, and D. Graveron-Demilly. Time-domain semi-parametric estimation based on a metabolite basis set. *NMR Biomed*, 18(1):1–13, 2005.

[69] J. van der Veen, R. de Beer, P. Luyten, and D. van Ormondt. Accurate quantification of in vivo 31p nmr signals using the variable projection method and prior knowledge. *Magn Reson Med*, 6(1):92–8, 1988.

[70] L. Vanhamme, A. van den Boogaart, and S. Van Huffel. Improved method for accurate and efficient quantification of mrs data with use of prior knowledge. *J Magn Reson*, 129(1):35–43, 1997.

[71] B. Michael Kelm, H. Menze Bjoern, M. Zechmann Christian, T. Baudendistel Klaus, and A. Hamprecht Fred. Automated estimation of tumor probability in prostate magnetic resonance spectroscopic imaging: Pattern recognition vs quantification, 2007.

[72] J. Luts, T. Laudadio, A. J. Idema, A. W. Simonetti, A. Heerschap, D. Vandermeulen, J. A. Suykens, and S. Van Huffel. Nosologic imaging of the brain: segmentation and classification using mri and mrsi. *NMR Biomed*, 22(4):374–90, 2009.

[73] P. Tiwari, M. Rosen, and A. Madabhushi. A hierarchical spectral clustering and nonlinear dimensionality reduction scheme for detection of prostate cancer from magnetic resonance spectroscopy (mrs). *Medical Physics*, 36(9):3927–3939, 2009.

[74] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 1933.

[75] P. Comon. Independent component analysis, a new concept. *Signal Processing*, 36(3):287–314, 1994.

[76] B. M. Kelm, H. M. Bjoern, M. Z. Christian, T. B. Klaus, and A. H. Fred. Automated estimation of tumor probability in prostate magnetic resonance spectroscopic imaging: Pattern recognition vs quantification. *Magnetic resonance in medicine*, 57:150–159, 2007.

[77] X. Liu, D. Langer, M. Haider, Y. Yang, M. Wernick, and I. Yetik. Prostate cancer segmentation with simultaneous estimation of markov random field parameters and class. *IEEE Trans Med Imaging*, 28(6):906–15, 2009.

[78] D. Ampeliotis, A. Antonakoudi, K. Berberidis, E. Z. Psarakis, and A. Kounoudes. A computer-aided system for the detection of prostate cancer based on magnetic resonance image analysis. In *Communications, Control and Signal Processing, 2008. ISCCSP 2008. 3rd International Symposium on*, pages 1372–1377, 2008.

[79] S. Ozer, D. Langer, X. Liu, M. Haider, T. Kwast, A. Evans, Y. Yang, M. Wernick, and I. Yetik. Supervised and unsupervised methods for prostate cancer segmentation with multispectral mri. *Medical Physics*, 37:1873–1883, 2010.

[80] J. Jesneck, L. Nolte, J. Baker, C. Floyd, and J. Lo. Optimized approach to decision fusion of heterogeneous data for breast cancer diagnosis. *Med Phys*, 33(8):2945–54, 2006.

[81] K. Veeramachaneni, L. A. Osadciw, and P. K. Varshney. An adaptive multimodal biometric management algorithm. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 35(3):344–356, 2005.

[82] Richard Duda, Peter Hart, and David Stork. Pattern classification (2nd edition). Technical Report 0471056693, Wiley-Interscience, 2000.

[83] A. Simonetti, W. Melssen, F. Edelenyi, J. van Asten, Arend Heerschap, and Lutgarde M. C. Buydens. Combination of feature-reduced mr spectroscopic and mr imaging data for improved brain tumor classification. *NMR in Biomedicine*, 18(1):34–43, 2005.

[84] Leo Breiman. Bagging predictors. *Mach. Learn.*, 24(2):123–140, 1996.

[85] Wang J. Zha H. Zhang, Z. Adaptive manifold learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011.

[86] T. Damoulas and M. Girolami. Probabilistic multi-class multi-kernel learning: On protein fold recognition and remote homology detection. *Bioinformatics*, 24(10):1264–1270, 2008.

[87] Shi Y., F. Tillmann, D. Anneleen, T. Leon-Charles, S. Johan, D. M. Bart, and M. Yves. L2-norm multiple kernel learning and its application to biomedical data fusion. *BMC Bioinformatics*, 11(1):309–333, 2010.

[88] L. T. Mainardi, D. Origgi, P. Lucia, G. Scotti, and S. Cerutti. A wavelet packets decomposition algorithm for quantification of in vivo (1)h-mrs parameters. *Med Eng Phys*, 24(3):201–8, 2002.

[89] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *Information Theory, IEEE Transactions on*, 38(2):713–718, 1992.

[90] R. Bellman. *Adaptive control processes: a guided tour*. Princeton University Press, 1961.

[91] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–6, 2000.

[92] J. Venna and S. Kaski. Local multidimensional scaling. *Neural Networks*, 19(6):889–899, 2006.

[93] A.C. Bovik, M. Clark, and W.S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE J PAMI*, 12(1):55–73, 1990.

[94] Christoph Busch. Wavelet based texture segmentation of multi-modal tomographic images. *Comput Graph*, 21(3):347–358, 1997.

[95] R. M. Haralick, Dinstein, and K. Shanmugam. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3:610–621, 1973.

[96] J.C. Russ. *The Image Processing Handbook*. CRC Press, 5th edition, 2007.

[97] A. Prinzie and D. Van den Poel. Random forests for multiclass classification: Random multinomial logit. *Expert systems with Applications*, 34(3):1721–1732, 2008.

[98] J. Quinlan. *C4. 5: programs for machine learning*. Morgan Kaufmann, 1993.

[99] T. Zhuowen. Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, 2:1589–1596 Vol. 2, 2005.

[100] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

[101] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[102] T. Rohlfing, A. Pfefferbaum, E. V. Sullivan, and C. R. Maurer. Information fusion in biomedical image analysis: Combination of data vs. combination of interpretations. In *Information Processing in Medical Imaging*, pages 150–161, 2005.

[103] Stephane G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.

[104] J. Zheng and E. Regentova. Wavelet based feature reduction method for effective classification of hyperspectral data. *Information Technology: Coding and Computing [Computers and Communications], 2003. Proceedings. ITCC 2003. International Conference on*, pages 483–487, 2003.

[105] N. Sebe and M. S. Lew. Wavelet based texture classification. *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, 3:947–950 vol.3, 2000.

[106] D. Gabor. Theory of communication. part 1: The analysis of information. *Electrical Engineers - Part III: Radio and Communication Engineering, Journal of the Institution of*, 93(26):429–441, 1946.

[107] Jose Mira, Jose lvarez, Orlando Arevalo Acosta, and Matilde Santos Penas. Classification of biomedical signals using a haar 4 wavelet transform and a hamming neural network. In *Nature Inspired Problem-Solving Methods in Knowledge Engineering*, volume 4528 of *Lecture Notes in Computer Science*, pages 637–646. Springer Berlin / Heidelberg, 2007.

[108] P. Subramani, R. Sahu, and S. Verma. Feature selection using haar wavelet power spectrum. *BMC Bioinformatics*, 7(1):432, 2006.

[109] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[110] R. Diaz-Uriarte and S. Alvarez de Andres. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7:3, 2006.

[111] K. R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf. An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, 12(2):181–201, 2001.

[112] B. Schlkopf, R. Herbrich, and A. Smola. A generalized representer theorem. *Computational Learning Theory*, 2111:416–426, 2001.

[113] T. Lin and H. Zha. Riemannian manifold learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30, 2008.

[114] D. Cai, X. He, Y. Hu, J. Han, and T. Huang. Learning a spatially smooth subspace for face recognition. *Proc. IEEE Conf. Computer Vision and Pattern Recognition Machine Learning (CVPR'07)*, pages 1–7, 2001.

[115] McKnight T., Noworolski S., Vigneron D., and Nelson S. An automated technique for the quantitative assessment of 3d-mrsi data from patients with glioma. *J Magn Reson Imaging*, 13(2):167–77, 2001.

[116] J. A. Jung, F. V. Coakley, D. B. Vigneron, M. G. Swanson, A. Qayyum, V. Weinberg, K. D. Jones, P. R. Carroll, and J. Kurhanewicz. Prostate depiction at endorectal mr spectroscopic imaging: Investigation of a standardized evaluation system. *Radiology*, 233:701–708, 2004.

[117] J. Star-Lack, S. J. Nelson, J. Kurhanewicz, L. R. Huang, and D. B. Vigneron. Improved water and lipid suppression for 3d press csi using rf band selective inversion with gradient dephasing (basing). *Magn Reson Med*, 38(2):311–21, 1997.

[118] Prodromos G. Borboroglu, Stewart W. Comer, Robert H. Riffenburgh, and Christopher L. Amling. Extensive repeat transrectal ultrasound guided prostate biopsy in patients with previous benign sextant biopsies. *The Journal of Urology*, 163(1):158–162, 2000.

[119] J. Jung, F. Coakley, D. Vigneron, M. Swanson, A. Qayyum, V. Weinberg, K. Jones, P. Carroll, and J. Kurhanewicz. Prostate depiction at endorectal mr spectroscopic imaging: investigation of a standardized evaluation system. *Radiology*, 233(3):701–8, 2004.

[120] A. Devos, L. Lukas, J. A. Suykens, L. Vanhamme, A. R. Tate, F. A. Howe, C. Majos, A. Moreno-Torres, M. van der Graaf, C. Arus, and S. Van Huffel. Classification of brain tumours using short echo time 1h mr spectra. *J Magn Reson*, 170(1):164–75, 2004.

[121] A. Heerschap, G. J. Jager, M. van der Graaf, J. O. Barentsz, J. J. de la Rosette, G. O. Oosterhof, E. T. Ruijter, and S. H. Ruijs. In vivo proton mr spectroscopy reveals altered metabolite content in malignant prostate tissue. *Anticancer Res*, 17(3A):1455–60, 1997.

[122] J. Kurhanewicz, D. Vigneron, H. Hricak, P. Narayan, P. Carroll, and S. Nelson. Three-dimensional h-1 mr spectroscopic imaging of the in situ human prostate with high (0.24-0.7-cm3) spatial resolution. *Radiology*, 198(3):795–805, 1996.

[123] John C. Russ. *The Image Processing Handbook, Fifth Edition (Image Processing Handbook)*. CRC Press, Inc., 2006.

[124] M. Zweig and G. Campbell. Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem*, 39(4):561–77, 1993.

[125] Thomas Dietterich. Ensemble methods in machine learning. *Multiple Classifier Systems*, pages 1–15, 2000.

[126] G. Lanckriet, T. De Bie, N. Cristianini, M. Jordan, and W. Noble. A statistical framework for genomic data fusion. *Bioinformatics*, 20(16):2626–35, 2004.

[127] R. Polikar. Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6(3):21–45, 2006.

[128] P. Zonari, P. Baraldi, and G. Crisi. Multimodal mri in the characterization of glial neoplasms: the combined role of single-voxel mr spectroscopy, diffusion imaging and echo-planar perfusion imaging. *Neuroradiology*, 49(10):795–803, 2007.

[129] J. Kurhanewicz, M. Swanson, S. Nelson, and D. Vigneron. Combined magnetic resonance imaging and spectroscopic imaging approach to molecular imaging of prostate cancer. *J Magn Reson Imaging.*, 16(4):451–463, 2002.

[130] J. Scheidler, H. Hricak, D. B. Vigneron, K. K. Yu, D. L. Sokolov, L. R. Huang, C. J. Zaloudek, S. J. Nelson, P. R. Carroll, and J. Kurhanewicz. Prostate cancer: localization with three-dimensional proton mr spectroscopic imaging–clinicopathologic study. *Radiology*, 213(2):473–80, 1999.

[131] A. Shukla-Dave, H. Hricak, M. W. Kattan, D. Pucar, K. Kuroiwa, H. N. Chen, J. Spector, J. A. Koutcher, K. L. Zakian, and P. T. Scardino. The utility of magnetic resonance imaging and spectroscopy for predicting insignificant prostate cancer: an initial analysis. *BJU Int*, 99(4):786–93, 2007.

[132] M. G. Swanson, D. B. Vigneron, T. K. Tran, and J. Kurhanewicz. Magnetic resonance imaging and spectroscopic imaging of prostate cancer. *Cancer Invest*, 19(5):510–23, 2001.

[133] E. Squillaci, G. Manenti, S. Mancino, M. Carlani, M. Di Roma, V. Colangelo, and G. Simonetti. Mr spectroscopy of prostate cancer. initial clinical experience. *J Exp Clin Cancer Res*, 24(4):523–30, 2005.

[134] C. Kim, B. Park, and B. Kim. Localization of prostate cancer using 3t mri: comparison of t2-weighted and dynamic contrast-enhanced imaging. *J Comput Assist Tomogr*, 30(1):7–11, 2006.

[135] J. Kurhanewicz, D. Vigneron, P. Carroll, and F. Coakley. Multiparametric magnetic resonance imaging in prostate cancer: present and future. *Curr Opin Urol*, 18(1):71–7, 2008.

[136] B. Turkbey, P. Pinto, H. Mani, M. Bernardo, Y. Pang, Y. McKinney, K. Khurana, G. Ravizzini, P. Albert, M. Merino, and P. Choyke. Prostate cancer: Value of multiparametric mr imaging at 3 t for detection histopathologic correlation1. *Radiology*, 255(1):89–99.

[137] Shah Mohak. Feature selection with conjunctions of decision stumps and learning from microarray data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99(PrePrints).

[138] G. Carneiro, B. Georgescu, S. Good, and D. Comaniciu. Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. *IEEE Trans Med Imaging*, 27(9):1342–55, 2008.

[139] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–23, 2000.

[140] J. Monaco, J. E. Tomaszewski, M. D. Feldman, I. Hagemann, M. Moradi, P. Mousavi, A. Boag, C. Davidson, P. Abolmaesumi, and A. Madabhushi. High-throughput detection of prostate cancer in histological sections using probabilistic pairwise markov models. *Medical Image Analysis*, 14(4):617–629, 2010.

[141] T. K. Tran, D. B. Vigneron, N. Sailasuta, J. Tropp, P. Roux, J. Kurhanewicz, S. Nelson, and R. Hurd. Very selective suppression pulses for clinical mrsi studies of brain and prostate cancer. *Magnetic Resonance in Medicine*, 43(1):23–33, 2000.

[142] J. Chappelow, N. Bloch, N. Rofsky, E. Genega, R. Lenkinski, W. DeWolf, and A. Madabhushi. Elastic registration of multimodal prostate mri and histology via multiattribute combined mutual information. *Medical Physics*, 38:2005–2018, 2011.

[143] G. Xiao, N. Bloch, J. Chappelow, E. Genega, N. Rofsky, R. Lenkinski, J. Tomaszewski, M. Feldman, M. Rosen, and A. Madabhushi. Determining histology-mri slice correspondences for defining mri-based disease signatures of prostate cancer. *Computerized Medical Imaging and Graphics*, 35:568–578, 2011.

[144] J. Vilanova, J. Comet, C. Barcel-Vidal, J. Barcel, E. Lpez-Bonet, A. Maroto, M. Arzoz, A. Moreno, and J. Areal. Peripheral zone prostate cancer in patients with elevated psa levels and low free-to-total psa ratio: Detection with mr imaging and mr spectroscopy. *Radiology*, 253:135–143, 2009.

[145] D. Yakar, S. Heijmink, Huisman H. Hulsbergen-van de Kaa, C., J. Barentsz, J. Ftterer, and T. Scheenen. Initial results of 3-dimensional 1 h-magnetic resonance spectroscopic imaging in the localization of prostate cancer at 3 tesla should we use an endorectal coil. *Investigative Radiology*, 46, 2011.

[146] T. Scheenen, S. Heijmink, S. Roell, C. HulsbergenVan de Kaa, B. Knipscheer, J. A. Witjes, J. Barentsz, and A. Heerschap. Three-dimensional proton mr spectroscopy of human prostate at 3 t without endorectal coil: Feasibility. *Radiology*, 245, 2007. 507516.

[147] Reed G.D. Kurhanewicz J. Vigneron D.B. Noworolski, S.M. Post-processing correction of the endorectal coil reception effects in mr spectroscopic imaging of the prostate. *Journal of Magnetic Resonance in Imaging*, 32(3):654–662, 2010.

[148] L. Breiman. Random forest. *Machine Learning*, 45:5–32, 2001.

[149] J. Quinlan. C4.5: programs for machine learning. *Morgan Kaufmann Publishers Inc. San Francisco, CA*, 1993.

[150] M. H. Zweig and G. Campbell. Receiver-operating characteristic (roc) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4):561–577, 1993.

[151] J. Demsar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7, 2006. 130.

[152] J. Chappelow, A. Madabhushi, et al. Elastic Registration of Multimodal Prostate MRI and Histology via Multi-Attribute Combined Mutual Information. *Med Phys*, 2010 (Accepted).

[153] R. Toth, A. Madabhushi, et al. A magnetic resonance spectroscopy driven initialization scheme for active shape model based prostate segmentation. *Med Img Anal*, 2010 (In Press).

[154] M. Chen et al. Prostate cancer detection: Comparison of T2-w Imaging, DWI, proton MRSI, and the three techniques combined. *Acta Radiologica*, 49(5):602–610, 2008.

[155] Chan Kyo Kim et al. Prediction of locally recurrent prostate cancer after radiation therapy: Incremental value of 3T diffusion-weighted MRI. *JMRI*, 29(2):391–397, 2009.

[156] Yousef Mazaheri et al. Prostate cancer: Identification with combined diffusion-weighted MRI and 3d 1H MRSI correlation with pathologic findings. *Radiology*, 246(2):480–488, 2008.

[157] D. Gleason, T Hambrock, D. Somford, H. Huisman, I. van Oort, J. Witjes, C. Hulsbergen-van de Kaa, T. Scheenen, and J. Barentsz. Relationship between apparent diffusion coefficients at 3.0-t mr imaging and gleason grade in peripheral zone prostate cancer. *Radiology*, 2011.

Vita

Pallavi Tiwari

2008-2012 Ph.D. in Biomedical Engineering, Rutgers University

2006-08 M.S. in Biomedical Engineering, Rutgers University

2002-2006 B.S. in Biomedical Engineering, Rutgers University

2009-2011 Department of Defense Prostate Cancer Research Program Pre-doctoral Fellow, Department of Biomedical Engineering, Rutgers University

2006-2009 Research Assistant, Department of Biomedical Engineering, Rutgers University