A HIERARCHICAL SPECTRAL CLUSTERING AND NON-LINEAR DIMENSIONALITY REDUCTION SCHEME FOR DETECTION OF PROSTATE CANCER FROM MAGNETIC RESONANCE SPECTROSCOPY

by

PALLAVI TIWARI

A thesis 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

Joint Degree of Master of Science

Graduate Program in Biomedical Engineering

Written under the direction of

Dr. Anant Madabhushi

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

Oct, 2008

© 2008

Pallavi Tiwari

ALL RIGHTS RESERVED

ABSTRACT OF THE THESIS

A Hierarchical Spectral Clustering and Non-linear Dimensionality Reduction Scheme for Detection of Prostate Cancer from Magnetic Resonance Spectroscopy

By Pallavi Tiwari

Thesis Director:
Dr. Anant Madabhushi

Magnetic Resonance Spectroscopy (MRS) is a unique non-invasive method which has recently been shown to have great potential in screening of prostate cancer (CaP). MRS provides functional information regarding the concentrations of different biochemicals present in the prostate at single or multiple locations within a rectangular grid of spectra superposed on the structural T2-weighted Magnetic Resonance Imaging (MRI). Changes in relative concentration of specific metabolites including choline, creatine and citrate compared to "normal" levels is highly indicative of the presence of CaP. Most previous attempts at developing computerized schemes for automated prostate cancer detection using MRS have been centered on developing peak area quantification algorithms. These methods seek to obtain area under peaks corresponding to choline, creatine and citrate which is then used to compute relative concentrations of these metabolites. However, manual identification of metabolite peaks on the MR spectra, let alone via automated algorithms, is a challenging

problem on account of low SNR, baseline irregularity, peak-overlap, and peak distortion. In this thesis work a novel computer aided detection (CAD) scheme for prostate MRS is presented that integrates non-linear dimensionality reduction (NLDR) with an unsupervised hierarchical clustering algorithm to automatically identify cancerous spectra. The methodology comprises of two specific aims. Aim 1 is to first automatically localize the prostate region followed in Aim 2 by automated cancer detection on the prostate obtained in Aim 1. In Aim 1, a hierarchical spectral clustering algorithm is used to distinguish between informative and non-informative spectra in order to localize the region of interest (ROI) corresponding to the prostate. Once the prostate ROI is localized, in Aim 2, a non-linear dimensionality reduction (NLDR) scheme in conjunction with a replicated k-means clustering algorithm is used to automatically discriminate between 3 classes of spectra (normal, CaP, and intermediate tissue classes). Results of qualitative and quantitative evaluation of the methodology over 18 1.5 Tesla (T) *in-vivo* prostate T2-w and MRS studies obtained from the multi-site, multi-institutional ACRIN trial, for which corresponding histological ground truth of spatial extent of CaP is available, reveal that the CAD scheme has a high detection sensitivity (89.60) and specificity (78.98). Results further suggest that the CAD scheme has a higher detection accuracy compared to such commonly used MRS analysis schemes as z-score and PCA.

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. I would like to express my gratitude to Dr. Mark Rosen who has constantly guided this work with his insightful opinions and research discussions that have made this research achieve such a high quality.

All my lab mates at the Laboratory of Computational Imaging and Bioinformatics made it a convivial place to work. In particular, I would like to thank Satish Viswanath for his support and help in the past two years. Thanks to Vikas Menon, a friend and so much more, for having been with me through every stage of this thesis, motivating me and encouraging me at all times. I m tempted to individually thank all my friends who have been with me through thick and thin but the space here would be insufficient to thank them enough. I can't however forget to thank Gaurav, my best friend, for motivating me to enter this world of research where I truly belong.

I cannot finish without saying how grateful I am to my family. Particular thanks, of course, to Brother Anmol, Sister Ruchi and brother in law, Anand. Lastly, and most importantly, I wish to thank my loving parents, Suresh Tiwari and Swati Tiwari who have always supported and encouraged me to do my best in all matters of life. Without them this would have never been possible.

This work was made possible via grants from Coulter Foundation, New Jersey Commission on Cancer Research, National Cancer Institute (R21CA127186-01, R03CA128081-01), the Society for Imaging Informatics in Medicine (SIIM) and the Life Science Commercialization Award from Rutgers University. I personally would also like to acknowledge the ACRIN trial for MRI/MRS data used in this work.

Dedication

"The owner always takes away the best flower in the garden. Others have no right to complain. We have to accept "His" choice."

This work is dedicated to the light who brightens my day- My Mother.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

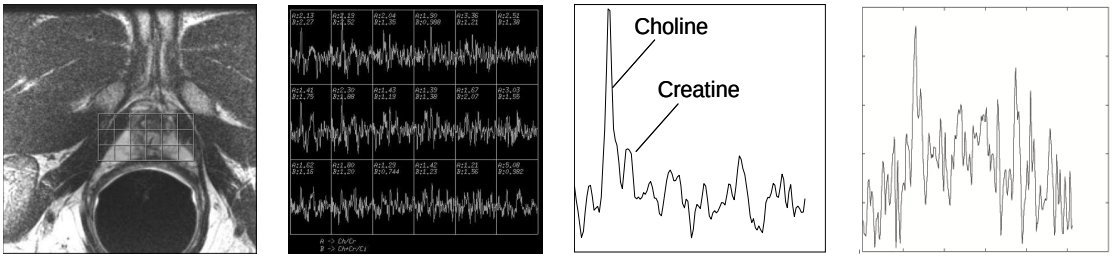
1.1 Background

Prostatic adenocarcinoma (CaP) is the most commonly occurring malignancy amongst men with 186,320 new cases and 28,660 deaths estimated to occur in the United States in 2008 (*American Cancer Society, 2008*). Early detection of CaP offers the best hope of curing it; however early prostate cancer is usually asymptomatic [1]. Screening of CaP is based on digital rectal examination (DRE) and monitoring elevated levels of the blood serum prostate specific antigen (PSA). Definitive diagnosis of CaP involves histological examination of biopsy specimens obtained via a blinded sextant trans-rectal ultrasound (TRUS) directed biopsy for patients with elevated PSA levels. Since prostate ultrasound is limited in its ability to identify CaP, biopsy locations are chosen at random within the prostate sextants. Consequently the CaP detection accuracy associated with TRUS is only 20-25% in patients with elevated PSA levels between 4-10 $\mu g/ml$ [2]. Recently, *in-vivo* endorectal T2-weighted structural Magnetic Resonance (MR) Imaging (MRI) of the prostate has allowed for greater discrimination between benign and cancerous prostatic structures as compared to TRUS [3]. However, structural T2-weighted MR imaging by itself has been shown to be limited in its ability to detect small foci of carcinoma contributing to a relatively low detection specificity [3].

1.2 Magnetic Resonance Spectroscopy

Over the past few years, Magnetic Resonance Spectroscopic (MRS) Imaging (MRSI) has emerged as a useful complement to structural MR imaging for potential screening of CaP [4, 5]. MRSI is a non-invasive technique used to obtain the metabolic concentrations of specific molecular markers and biochemicals in the prostate including citrate, creatine and choline, changes in concentration of which have been shown to be linked to presence of CaP [6, 7]. Spectra are obtained at either single or multiple locations from a rectangular spectral grid placed on a corresponding T2-weighted MR image. It has been demonstrated previously that the relative concentrations of choline, citrate and creatine are significantly different in CaP and normal regions within the prostate [6].

The relative concentrations of choline, creatine, and citrate are obtained by calculating the area under the peak for these metabolites to assess presence of CaP at a specific prostate location. Identification of precise location of characteristic metabolites on the MR spectra is a difficult task for radiologists due to (a) a low signal to noise (SNR) ratio and (b) presence of biomedical signal artifacts associated with MR spectra. Figure 1.1 (a) shows a MRS grid superimposed on the corresponding T2-w MRI slice, Figure 1.1 (b) shows a sample MRS grid with spectra acquired from the spectral metavoxels in the grid shown in 1 (a), while 1.1 (c) shows an idealized CaP spectrum (Note that the spectra in 1.1 (c) is not associated with the spectra in (b)). Figure 1.1 (d) shows a typical noisy spectra obtained from one of the locations in the MRS grid in 1 (b) and subsequently magnified. Note the amount of noise in the MR spectrum in Figure 1.1 (d) which severely limits the ability of radiologists to visually identify metabolite peaks. The limitation of peak detection based approaches has meant that MRS as a potential CaP screening modality (in conjunction with MRI) is yet to gain acceptance and its role remains controversial [8].



(a) (b) (c) (d)

Figure 1.1: (a) 2D section of representative T2 weighted MRI scene with spectral grid overlaid, (b) Individual MR spectra acquired from different locations on the spectral grid, (c) an idealized CaP spectra showing the locations of choline, creatine and, (d) a typical noisy spectra magnified from (b) where the spectral noise in (d) makes visual identification of pertinent metabolic peaks difficult. Note that the prostate is normally contained in a 3×6 or a 3×7 spectral grid in the X-Y plane. Note also that the idealized, noise free spectra shown in (c) does not actually correspond to any of the spectra in the grid in (b).

1.3 Computer aided diagnosis (CAD)

Computer-aided detection (CAD) refers to the use of computers to assist a physician in rendering a diagnostic decision through the quantitative analysis of biomedical data [9]. The low detection accuracy associated with current TRUS prostate biopsies points to a need for developing an image guided decision support system to direct needle placement in the prostate to increase CaP detection accuracy. While some researchers [10], [11] [12] have begun to develop CAD systems for CaP detection from structural and functional MR imaging, corresponding developments in MRS have not been as forthcoming. This is in spite of evidence to suggest that integration of structural and metabolic imaging [7] could boost diagnostic yield over that which could be obtained from any individual modality. Previous CAD methods for CaP detection from MRS [13, 14, 15, 16, 17], focused on automated metabolite peak detection have been hindered by presence of noise in the spectral data (Figure 1.1 (d)).

Dimensionality reduction (DR) refers to the projection of high dimensional data into

a reduced dimensional feature space without a significant loss in class discriminatory information. The low dimensional representation of the data is easier to visualize and DR algorithms aim to preserve object relationships, so that objects that are close to one another in the high dimensional ambient space are mapped to adjacent locations in the resulting low dimensional embedding space. However, linear DR schemes such as principal component analysis (PCA) assume the original high dimensional data to be inherently linear and hence employ linear projection methods to reduce data dimensionality. It has been demonstrated previously that linear DR schemes are inherently limited for purposes of reducing and classifying high dimensional biomedical data [18] owing to the non-linear manifold on which the data resides. Recently several non-linear dimensionality reduction (NLDR) algorithms have been proposed for analysis and visualization of non-linear data [19, 20, 21]. The objective behind NLDR methods is to non-linearly map objects, c, d belonging to the same object class and characterized by M dimensional feature vectors $\boldsymbol{F}(c), \boldsymbol{F}(d)$ to adjacent locations $\boldsymbol{S}(c), \boldsymbol{S}(d)$ in the low dimensional embedding, where $\boldsymbol{S}(c), \boldsymbol{S}(d)$ represent the m-dimensional dominant Eigen vectors corresponding to c, d ($m << M$). Unlike linear DR schemes that typically employ the Euclidean distance measure to estimate object distances, most NLDR schemes employ some method for estimating geodesic distances between objects in the high dimensional (M) space.

1.4 Brief outline of the work

In this work, a fully automated CAD system for detecting prostate cancer from 1.5 T endorectal *in vivo* MRS studies is presented. Figure 1.3 shows how the proposed CAD scheme works. The methodology comprises of two specific aims;

- Aim 1: To automatically identify the prostate region of interest (ROI) from out of

the large MRS grid

- Aim 2: To automatically identify suspicious regions on the prostate ROI obtained from Aim 1.

In Aim 1, a novel hierarchical classification scheme is employed to recursively distinguish informative from non-informative spectra via Graph Embedding (GE) [22, 31], a well known NLDR scheme, in order to hone in on the region of interest (ROI) corresponding to the prostate. As may be seen in Figure 1.1 (a) the prostate occupies approximately 10% of the total volume within a 3D *in-vivo* endorectal prostate MRI scene. An iterative algorithm is employed to automatically identify each of the MR spectra as belonging to either inside (informative) or outside (non-informative) the prostate. Since the non-informative spectra (outside the prostate) are most populous, the largest cluster is identified at each iteration and eliminated. Figure 1.2 shows the difference in prostate (Figure 1.2 (b)) and non-prostate (Figure 1.2 (c), (d))) MR spectra. This process is repeated until the number of MR spectra remaining is approximately equal to the number (Θ) usually contained in the prostate (Figure 1.1 (a)). The removal of non-informative spectra in Aim 1 makes it easier to discriminate between cancerous and normal spectra within the prostate. Following automated localization of the prostate spectra in Aim 1, in Aim 2, a NLDR scheme is applied to non-linearly embed the informative spectra into a reduced dimensional space. The individual prostate spectra now characterized by their low dimensional embedding coordinates are clustered into distinct classes via a "replicated clustering scheme". Replicated clustering overcomes the limitation of traditional k means clustering algorithm which is known to be dependent on randomly initialized centroids. A different clustering result,

based on the local minimum, is usually obtained each time k means is performed. In replicated clustering, multiple weak clusterings are generated by repeatedly clustering the data. The optimal clustering solution is then chosen among the various weak clusterings as the one with the least intra-cluster variance. All spectra are aggreagated into 3 classes based on the assumption that they correspond to normal, CaP, and intermediate tissue classes, where the intermediate tissue cluster may correspond to spectra from tissue classes such as Benign Hyperplexia (BPH), high-grade prostatic inter-epithelial neoplasia (HGPIN) with characteristics between normal and malignant.

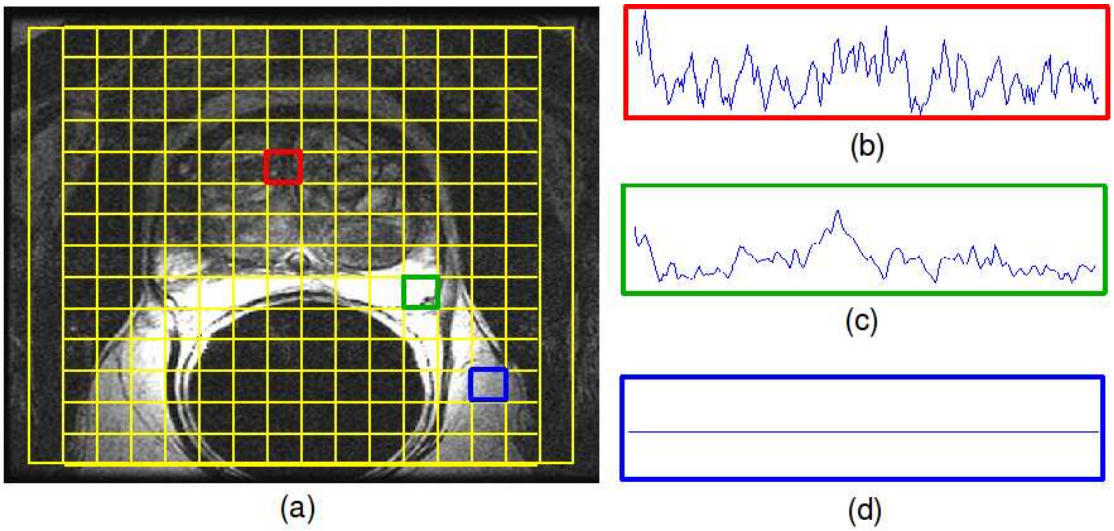


Figure 1.2: (a) 2D section of T2-weighted MR image intensity scene with a spectral grid of metavoxels overlaid. MRS spectra for a metavoxel (b) within the prostate, (c) outside the prostate, and (d) a MRS spectra for a zero-padded metavoxel near the periphery of the MRI scene.

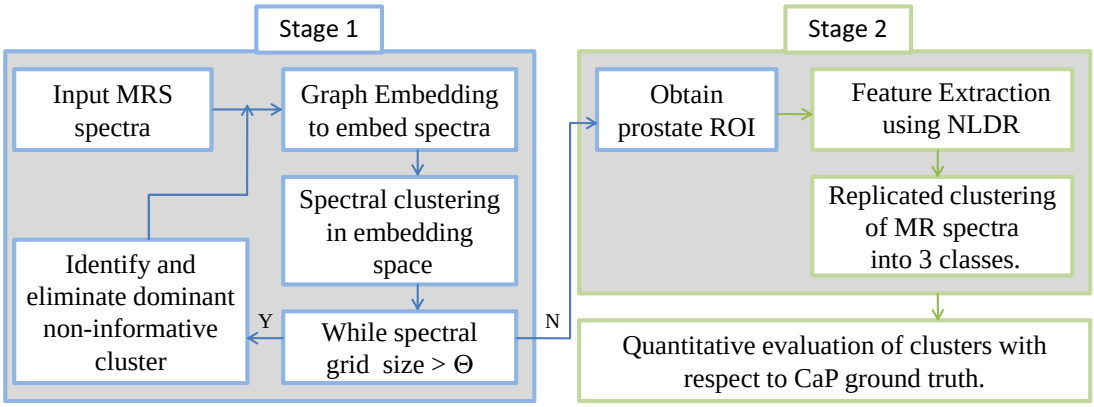


Figure 1.3: Flowchart showing various system components and methodological overview of the CAD scheme. Hierarchical spectral clustering is performed in Aim 1 to automatically obtain the prostate grid, followed in Aim 2 by identification of different tissue classes via NLDR and replicated clustering.

1.5 Organization of the Thesis

The rest of this thesis work is organized as follows. In Chapter 2 previous related work in MRS CAD is discussed. A detailed description of the dimensionality reduction methods employed in this work are described in Chapter 3, while the details of the methodology are explained in Chapter 4. In Chapter 5 results of qualitative and quantitative evaluation of the scheme on a total of 18 prostate MRI-MRS studies are presented. Two specific applications of the proposed scheme are explained briefly in Section 6. Concluding remarks and direction of future research are presented in Chapter 7.

Chapter 2

Related work

Previous attempts at computerized analysis of MRS can be classified into two broad categories, (a) signal quantification and (b) model independent techniques.

2.1 Signal quantification techniques

As mentioned previously some researchers [8] have suggested using the ratio of (choline + creatine)/citrate, as an index to help predict tumor aggressiveness. This ratio is considered sensitive and peak detection algorithms used currently are based on automated calculation of this ratio to identify suspicious regions. The approach is called peak quantification where the aim is to obtain the peak areas or relative concentrations of different metabolites like choline, creatine and citrate as accurately as possible. Figure 2.1 shows the characteristic metabolites (choline, creatine and citrate) in a typical benign MR spectrum.

Signal quantification methods can be further subdivided into two categories: time domain fitting (TDF) and frequency domain analysis (FDA) methods. Commonly used TDF methods include VARPRO [13], AMARES [14] and QUEST [15] 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. TDF techniques are considered flexible with respect to the choice of model function as signal quantification is performed in the same domain as the acquired signals.

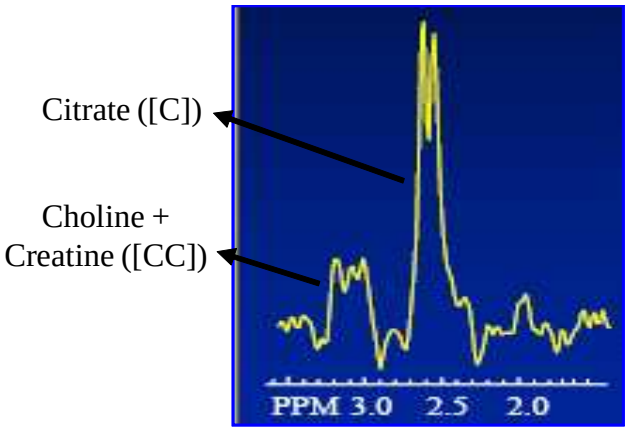


Figure 2.1: A typical artifact free benign spectrum in which choline, creatine and citrate peaks could be clearly identified. Note that in case of a cancer spectrum (choline+ creatine) peak is elevated while the citrate peak is depressed.

However, choosing the number of components and the best fit model are critical issues to be considered for accurate quantification. With FDA methods, the aim is to automatically and accurately calculate the ratio, (choline + creatine)/citrate; an index to help predict tumor aggressiveness in the frequency domain [16]. In [17], a FDA method based on peak integration was compared with a TDF method, AMARES, in terms of the ability to estimate the (choline + creatine + polyamines) to citrate ratio (CCP:C) for CaP detection. Both TDF and FDA methods were found to be successfully able to identify CaP voxels on MRS for an artifact free study using CCP:C ratio. However, performance of both the TDF and FDA models are drastically affected by (i) optimal choice of prior knowledge (model function), (ii) presence of noise and contributions from non-prostate spectra, (iii) peak overlap owing to contributions from multiple metabolites, and (iv) baseline distortion and line broadening [23]. In order to avoid the limitations of model based and peak detection approaches for MRS, recently some researchers have begun to explore domain independent techniques

such as z-score and Principal Component Analysis (PCA). An excellent comprehensive comparison of quantification and pattern recognition schemes used for MRS quantification is provided in [23].

2.2 Model Independent Techniques

Devos *et al.* [26] presented an automated CAD system for classification of brain MRS using Linear Discriminant Analysis (LDA) and a least squares support vector machine (LS-SVM) classifier to discriminate between different tumor classes. Ma *et al.* [27] and Simonetti *et al.* [28] have explored other linear dimensionality reduction methods such as Independent Component Analysis (ICA) and Principal Component Analysis (PCA) in conjunction with SVM, to differentiate between tissue classes via brain MRS.

However, Recent work in classification [18], has suggested that biomedical data is inherently non-linear. Consequently, methods such as PCA and LDA that are predicated on the fact that the underlying structure of the data is inherently linear and are limited in their ability to embed data in a space where object classes are discriminable.

2.3 Novel Contributions of this work

In this work, a computer aided diagnostic (CAD) system for detection of prostate cancer from 1.5 T MRS based on NLDR and unsupervised replicated clustering is presented. A comprehensive quantitative evaluation is performed based on 3 fold and 5 fold cross validation via partial CaP ground truth extent obtained from accompanying histology. The proposed NLDR, replicated clustering based scheme for CaP detection is also quantitatively compared against z-score and PCA in terms of detection sensitivity, and specificity. The major novel contributions of this work are,

1. A fully automated scheme for elimination of non-informative prostate spectra and localization of the prostate ROI,

2. An accurate NLDR based clustering scheme that distinguishes cancerous from normal MR spectra by exploiting the totality of the spectral data in a reduced dimensional representation, avoiding the limitations associated with model, peak based, and linear DR analysis schemes.

Chapter 3

Overview of Linear and non-linear DR methods

3.1 Notation

The 3D prostate T2 weighted scene is represented by $\mathcal{G} = (G, f)$, where G is a 3D grid of voxels $g \in G$ and $f(g)$ is a function that assigns an intensity value to every $g \in G$. A spectral scene is defined as $\mathcal{C} = (C, \boldsymbol{F})$ where C is a 3D grid of metavoxels, $c \in C$. Each metavoxel c is associated with a corresponding M-dimensional spectral vector $\boldsymbol{F}(c) = [\boldsymbol{f}_u(c) \mid u \in \{1, ..., M\}]$ where $\boldsymbol{f}_u(c)$ represents the concentration of different biochemicals (such as creatine, citrate, and choline). Figure 3.1 shows the spatial relationship between the MR spectral metavoxel $c \in C$ and MR image voxel $g \in G$. Note that the distance between any two adjacent metavoxels $c_i, c_j \in C$, $||c_i - c_j||_2$, where $|| \cdot ||_2$ denotes the L_2 norm, $i, j \in \{1, ..., |C|\}$, and $|C|$ is the cardinality of C, is roughly 16 times the distance between any two adjacent spatial voxels $g_a, g_b \in G$, where $a, b \in \{1, ..., |G|\}$. A list of commonly used symbols and notations is given in Table 3.1.

3.2 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 [29]. These orthogonal eigenvectors corresponding to the largest eigenvalues are called '*principal components*'. To obtain these principal components each

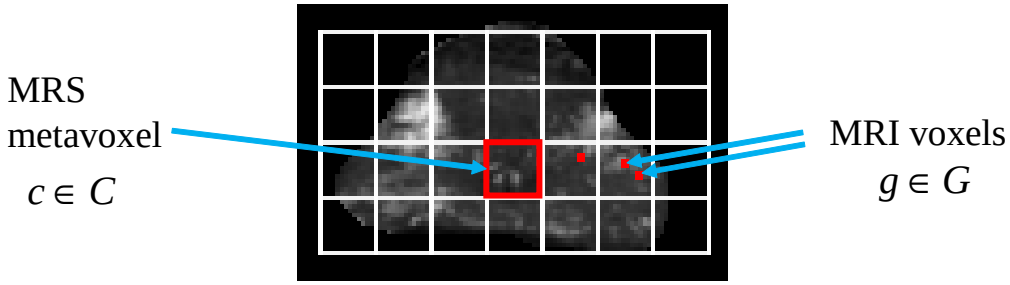


Figure 3.1: Figure showing the relationship between MRS metavoxels $c \in C$ and MRI voxels $g \in G$. The spectral grid C comprising 28 metavoxels has been overlaid on a T2-weighted MRI prostate slice and is shown in white. Note the region outlined in red on C corresponds to the area occupied by a metavoxel, but will contain multiple MRI voxels (highlighted in red).

data point $c \in C$ is first centered by subtracting the mean of all the features for each observation c from its original M dimensional feature value $\boldsymbol{f}(c)$ as shown in Equation 3.1.

$$\bar{\boldsymbol{f}}_u(c) = \boldsymbol{f}_u(c_i) - \frac{1}{|C|}\sum_{c \in C} \boldsymbol{f}_u(c_i), u \in \{1, .., M\}. \tag{3.1}$$

From feature values $\bar{\boldsymbol{f}}(c)$ for each $c \in C$, a new $|C| \times M$ matrix $\mathcal{Y}$ is constructed. The matrix $\mathcal{Y}$ is then decomposed into corresponding singular values as shown in Equation 3.2.

$$\mathcal{Y} = UW_{PCA}V^T, \tag{3.2}$$

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 Eigenvector $\boldsymbol{S}_v^{PCA}$, $v \in \{1, 2, ..., m\}$, in matrix V^T and are used to rank the corresponding eigenvectors in the order of greatest variance. Thus the first m Eigenvectors are extracted that represent a pre-specified percentage of the variance in the data while the remaining

Symbol	Description
$\mathcal{G}$	3D MRI scene
$\mathcal{C}$	3D MR spectral scene
G	3D grid of MRI voxels
C	3D grid of metavoxels
g	Voxel location in $G, g \in G$
c	A metavoxel in $C, c \in C$
$f(g)$	Intensity value at g
$\boldsymbol{F}(c)$	Vector of spectral content at c
ϕ	DR method, $\phi \in \{PCA, LLE, ISO, GE\}$
$\boldsymbol{S}^{\phi}(c)$	Output vector obtained at metavoxel c
M	Number of original high dimensions
m	Reduced dimensions, $m << M$
$\mathcal{R}$	Maximum CaP diameter
K_s	Contiguous slices with CaP presence
C_P	Potential cancer space, $C_P \subset C$
N_g	Number of CaP voxels in C_P, $N_g \in C_P$
$\hat{V}^{\phi,q}$	Stable clusters obtained after clustering for $q \in \{1, 2, 3\}$
V_T	Prostate grid
Θ	Actual number of voxels contained in prostate
θ_z	z-score threshold
κ	Nearest neighborhood parameter for $\phi \in \{LLE, ISO\}$
v	Reduced dimensions for $\phi \in \{LLE, ISO, GE, PCA\}$

Table 3.1: List of commonly used notation and symbols in this thesis work.

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 it being a linear DR scheme, an advantage of PCA is that it allows for specification of the number of Eigen vectors required to explain a pre-specified percentage of the variance in the data.

3.3 Non-Linear Dimensionality Reduction Methods

In this work, 3 popularly used NLDR schemes are considered, Isometric mapping (Isomaps), Locally Linear Embedding (LLE), and Graph Embedding (GE) for MRS analysis. The reason for considering Isomaps, LLE and graph embedding is to motivate the use of NLDR for spectral analysis over linear DR schemes such as PCA.

3.3.1 Isometric Mapping (Isomap):

The Isomap algorithm [19] handles nonlinearities in the data through the use of a neighborhood mapping. By creating linear connections from each point $c \in C$ to its κ closest neighbors in Euclidean space, a manifold representation of the data is constructed, κ being a user-defined neighborhood parameter. Non-linear connections between points outside of the κ neighborhood are approximated by calculating the shortest distance between c_i, $c_j \in C, i, j \in \{1, ..., |C|\}$, along the paths in the neighborhood map. Thus new geodesic distances (distances measured along the surface of the manifold) are calculated and arranged in a pairwise distance matrix W_{ISO}, where $W_{ISO}(i, j)$ contains the nonlinear geodesic distances between c_i, $c_j \in C$. The matrix W_{ISO} is then given as an input to the classical Multi Dimensional Scaling (MDS) algorithm [30] which projects each $\boldsymbol{F}(c_i), \forall c_i \in C$, to a nonlinear lower dimensional space, $\boldsymbol{S}^{ISO}(c_i), i \in \{1, \ldots, |C|\}$. MDS is implemented as a linear method that preserves the Euclidean geometry between each pair of metavoxels $c_i, c_j \in C$ 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 W_{ISO} matrix.

3.3.2 Graph Embedding (GE):

The aim of Graph Embedding [31] is to find an embedding vector $\boldsymbol{S}^{GE}(c_i)$, $\forall c_i \in C$, $i \in \{1, \ldots, |C|\}$, such that the relative ordering of the distances between objects in high dimensional space is maximally preserved in the lower dimensional space. Thus, if locations $c_i, c_j \in C, i, j \in \{1, \ldots, |C|\}$, are adjacent in the high dimensional feature space, then $||\boldsymbol{S}^{GE}(c_i) - \boldsymbol{S}^{GE}(c_j)||_2$ should be small, where $||.||_2$ represents the Euclidean norm. This will only be true if the distances between all $c_i, c_j \in C$ are preserved in the low dimensional mapping of the data. To compute the optimal embedding, an adjacency matrix $W_{GE} \in$

$\Re^{|C|\times|C|}$ is first defined as

$$W_{GE}(i,j) = e^{-||\boldsymbol{F}(c_i)-\boldsymbol{F}(c_j)||_2}, \forall c_i, c_j \in C, i, j \in \{1, \ldots, |C|\}. \tag{3.3}$$

$\boldsymbol{S}^{GE}(c_i)$ then obtained from the maximization of the function:

$$E(\mathcal{X}_{GE}) = 2\gamma \times trace\left[\frac{\mathcal{X}_{GE}(D - W_{GE})\mathcal{X}_{GE}^{\mathsf{T}}}{\mathcal{X}_{GE}D\mathcal{X}_{GE}^{\mathsf{T}}}\right], \tag{3.4}$$

where $\mathcal{X}_{GE} = \left[\boldsymbol{S}^{GE}(c_1); \boldsymbol{S}^{GE}(c_2); \ldots; \mathbf{S}^{GE}(c_n)\right], n = |C|$, and $\gamma = |C| - 1$. Additionally, D is a diagonal matrix where $\forall c_i \in C$, $i \in \{1, ..., |C|\}$, the diagonal element is defined as $D(i,i) = \sum_j W_{GE}(i,j)$. The embedding space is defined by the Eigen vectors corresponding to the smallest m Eigen values of $(D - W_{GE})\ \mathcal{X}_{GE} = \lambda D W_{GE}$. The matrix $\mathcal{X}_{GE} \in \Re^{|C|\times m}$ of the first m Eigen vectors is constructed, and $\forall c_i \in C, \boldsymbol{S}^{GE}(c_i)$ is defined as row i of $\mathcal{X}_{GE}$.

3.3.3 Locally Linear Embedding (LLE):

LLE [20] operates by assuming that objects within a local neighborhood in a high dimensional feature space are linearly related. Consider the set of spectral vectors $\mathcal{F} = \{\boldsymbol{F}(c_1), \boldsymbol{F}(c_2), \ldots, \boldsymbol{F}(c$ $\forall c_i \in C, i \in \{1, ..., n\}$. LLE aims to map the set $\mathcal{F}$ to the corresponding set $\mathcal{X}_L LE = \{\boldsymbol{S}^{LLE}(c_1), \boldsymbol{S}^{LLE}(c_2), \ldots, \boldsymbol{S}^{LLE}(c_n)\}$ of embedding co-ordinates. Let $d^{(1)}, ..., d^{(k)}$ be the k nearest neighbors of c_i and let $\eta^k(c_i)$ is the index of the location of its k-nearest neighbor (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 local linearly, 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 kNN. 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

$$E_1(W_{LLE}) = \sum_{i=1}^{n}\left\|\boldsymbol{F}(c_i) - \sum_{r=1}^{k} W_{LLE}(i, \eta^r(c_i))\, \boldsymbol{F}\left(d^{(r)}\right)\right\|_2. \tag{3.5}$$

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 kNN, its projection $\boldsymbol{S}^{LLE}(c_i)$ will be the weighted combination of the projections of these same kNN. The optimal $\mathcal{X}_{LLE}$ in the least squares sense minimizes

$$E_2(\mathcal{X}_{LLE}) = \sum_{i=1}^{n} \left\| \boldsymbol{S}^{LLE}(c_i) - \sum_{j=1}^{n} W_{LLE}(i,j)\, \boldsymbol{S}^{LLE}(c_j) \right\|_2 = trace\left(\mathcal{X}_{LLE} L \mathcal{X}_{LLE}^{\mathsf{T}}\right), \quad (3.6)$$

where $\mathcal{X}_{LLE} = \left[\boldsymbol{S}^{LLE}(c_1), \boldsymbol{S}^{LLE}(c_2), \ldots, \boldsymbol{S}^{LLE}(c_n)\right]$, $L = (I - W_{LLE})\left(I - W_{LLE}^{\mathsf{T}}\right)$, and I is the identity matrix. The minimization of Equation 3.6 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 Eigenvectors of the Laplacian matrix L. Since the rank of L is $n-1$, the first Eigenvector is ignored and the second smallest Eigenvector represents the best one-dimensional projection of all the samples. The best two-dimensional projection is given by the Eigenvectors with the second and third smallest eigenvalues, and so forth.

Chapter 4

Methodology

In Section 4.1, a brief description of the data sets considered in this work is presented. Details concerning determination of ground truth for spatial extent of CaP on MRI are given in Section 4.2. Replicated unsupervised clustering employed in the CAD system is explained in Section 4.3. Experimental details of the two specific aims of the CAD system for CaP detection on MRS are explained in Sections 4.4 and 4.5 respectively.

4.1 Data Description

A total of 18 *in-vivo* endorectal MRI and MRS studies were obtained from the American College of Radiology Imaging Network (ACRIN) multi-site prostate trial[1]. These studies were acquired with 1.5 Tesla GE Medical Systems through the PROSE (PROstate Spectroscopy and imaging Examination)[2] package (voxel width 0.4 x 0.4 x 3 mm). For each patient, MR (T2 imaging protocol) and MRS data were acquired prior to radical prostatectomy. Following resection, the gland was sectioned and stained for CaP. In most cases, only partial histological tissue sections were obtained due to issues with gland sectioning. In some cases, some tissue sections and individual tissue pieces were discarded. The individual

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

[2] *www.gehealthcare.com/usen/mr/applications/products/prose.html*

tissue pieces were then manually examined ans CaP extent delineated on the histopathology. Following CaP annotation on the histological tissue sections, CaP size (diameter) and sextant locations were recorded for each study. The MRS spectral grid was contained in DICOM images from which the 16×16 grid containing N=256 spectra was obtained using IDL 6.4 (ITT Visual Information Systems).

4.2 Determining approximate ground truth for spatial extent of CaP on prostate

As mentioned above only partial ground truth for spatial extent for CaP on the 1.5 T MR studies in the ACRIN database is available in the form of sextant location and diameter size of CaP. Due to the limitations associated with having only partial tissue sections, an approximate ground truth is defined for spatial extent of CaP on MRI in the following manner. The prostate is visually divided into two symmetrical regions: Left (L) and Right (R), and further divided into 3 regions during imaging: Base (B), Midgland (M), and Apex (A). Approximate space within which the CaP extent is contained is then determined by identifying within which of six regions (Left Base (LB), Left Midgland (LM), Left Apex (LA), Right Base (RB), Right Midgland (RM) and Right Apex (RA)) CaP was identified during needle biopsies. The maximum diameter of CaP is also recorded in each of the 6 candidate locations and is denoted as $\mathcal{R}$.

In order to quantitatively evaluate CAD performance in terms of quantitative metrics such as sensitivity and specificity, the precise spatial location of the target class within C is required and unfortunately not available for these studies. Hence, a probabilistic ground truth is defined for CaP that involves first defining a *" potential cancer space, (C_P)"*. To appreciate the need for C_P, let us assume the ideal case scenario (Figure 4.1) where the

precise spatial location of CaP is known *a priori* and is denoted by the set C_s. If C_a denotes the set of spatial locations corresponding to CaP identified by the CAD system, the true positive (TP) area could be calculated as $|C_s \cap C_a|$. Similarly, the false positive (FP) area for the CAD system is $|C_a - (C_s \cap C_a)|$ and false negative (FN) area is $|C_s - (C_s \cap C_a)|$. For the problem considered in this work, C_s is unavailable and hence the need for C_P which is defined as the set of spatial locations within which a total number of N_g CaP locations are contained. The true number of metavoxels c within C_P that represent CaP can be calculated as,

$$N_g = K_s \times \left\lceil \frac{(\mathcal{R}^2)}{\Delta X \Delta Y} \right\rceil, \tag{4.1}$$

where K_s represents the number of contiguous MR sections containing CaP, $\lceil\rceil$ refers to the ceiling operation and $\Delta X, \Delta Y$ refer to the size of the metavoxel c in the X and Y dimensions. Thus for a MRS scene $\mathcal{C}$, with known cancer in left midgland (LM), the prostate being contained in a 3×6 grid and the prostate midgland region extending over 2 contiguous slices, the total number of CaP metavoxels $|C_P|$ is 18 $(3 \times 3 \times 2)$. The 3×6 prostate grid is divided into two equal right and left halves. Given that the tumor has a maximum diameter of 13.75 mm in LM, with $\Delta X, \Delta Y = 6.875$, N_g =8 metavoxels corresponding to CaP within C_P. Both C_P and N_g are integral to defining probabilistic estimates of CAD sensitivity, specificity and positive predictive value, details of which are provided in Section 5.1.

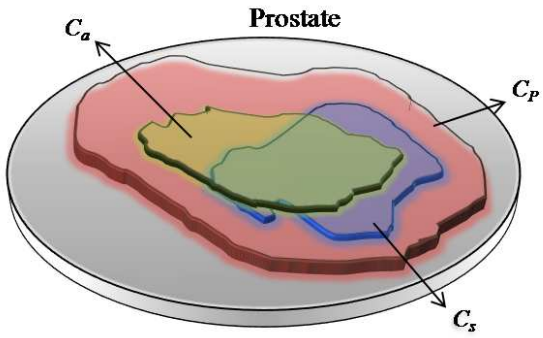


Figure 4.1: An illustration of the precise ground truth location (C_s) on the prostate and spatial location of the class (C_a) identified as CaP by a CAD system. Note that in such a case sensitivity, specificity of CaP detection via CAD can be determined precisely since C_s is known exactly. C_P represents the potential cancer space that needs to be defined when C_s is not available and contains within it N_g CaP metavoxels.

4.3 Replicated k-means clustering in the reduced feature space

For the DR schemes, $\phi \in \{PCA, LLE, GE, ISO\}$, unsupervised replicated clustering is used to stratify all metavoxels $c \in C$ into one of the 3 classes (cancer, benign, intermediate) based on $\boldsymbol{S}^{\phi}(c)$, the low dimensional representation of $\boldsymbol{F}(c)$. Replicated clustering is a variant of the popular k-means [37] clustering scheme. k-means however is known to be sensitive to the choice of initial cluster centers randomly chosen by the algorithm which has a significant bearing on the final clustering of the data. The k-means algorithm is initialized by randomly partitioning the data into k clusters and computing the cluster center for each partition. The distance of each point from each of the k centroids is computed and each object is reassigned to the closest centroid to minimize the intra-class variance. This random initialization may lead to local minima leading in turn to different clustering results. The

motivation behind replicated clustering is to make the final aggregation results from k-means more deterministic. Below the various steps involved in this algorithm are briefly described.

Step 1: At each of T iterations, k-means is applied to cluster all objects $c \in C$ into one of the three classes $V_t^1, V_t^2, V_t^3, t \in \{1, ..., T\}$. For each $c \in V_t^q$, $q \in \{1, 2, 3\}$, $t \in \{1, 2, ..., T\}$ the centroid of each cluster is determined as,

$$c_t^q = \frac{1}{|V_t^q|} \sum_{c \in V_t^q} c \tag{4.2}$$

Step 2: At each iteration $t \in \{1, ..., T\}$, the average Euclidean distance between each $c \in V_t^q$ and corresponding cluster center c_t^q, $t \in \{1, ..., T\}$, $q \in \{1, 2, 3\}$, is then determined as,

$$d_t^q = \frac{1}{|V_t^q|} \sum_{c \in V_t^q} ||c - c_t^q||. \tag{4.3}$$

The average intra-cluster distance over all 3 clusters is then obtained as,

$$\mu_t^d = \frac{1}{3} \sum_q d_t^q \tag{4.4}$$

Step 3: Finally, the clusterings $\hat{V}^q$, $q \in \{1, 2, 3\}$ within a specific iteration $t \in \{1, ..., T\}$ are identified as the stable clustering result for which μ_t^d is minimum over all t.

Note that replicated clustering identifies stable clusterings as those that minimize intra-class variance. Note further, that while the aim is not to explicitly seek to increase inter-cluster distance, the empirical results suggest that replicated k-means clustering tends to also push the cluster centers farther apart.

4.4 Aim 1: Localization of Prostate using Hierarchical clustering

Initially, most of the locations $c \in C$ correspond to zero padded or non informative spectra that lie outside the prostate (Figure 1.4). Hence, in Aim 1 the algorithm aims to eliminate

the dominant cluster which comprises non-prostate spectra. The non-linear DR scheme Graph Embedding [31] is employed to project all spectra into a reduced dimensional embedding $\boldsymbol{S}^{GE}(c)$ followed by replicated k-means clustering to aggregate all $c \in C$ into two clusters $\hat{V}^1$, $\hat{V}^2$ corresponding to informative and non-informative spectral classes. At each iteration $t \in \{1, ..., T\}$, a subset of voxels $\tilde{C}_t$ in C is obtained by eliminating the non-informative spectra identified as the dominant cluster ($\hat{V}^{dom}$). The approximate number of prostate spectra (Θ) of a MRS grid is learnt during the offline training phase. The automatic cascaded scheme stops when the number of remaining spectra in the MRS spectra is approximately equal to Θ. The result of the $HierarclustMRS$ algorithm is a spectral grid ($\tilde{C}_T$) containing all the prostate spectra.

Algorithm $HierarclustMRS$

Input: $\boldsymbol{F}(c)$ for all $c \in C$, Θ, C.

Output: $\tilde{C}_T$.

begin

0. Initialize $\tilde{C}_0 = C, t = 0$;
1. *while* $|\tilde{C}_t| > \Theta$ *do*
2. Apply Graph Embedding [31], to $\boldsymbol{F}(c)$, for all $c \in \tilde{C}_t$ to obtain $\boldsymbol{S}_t^{GE}(c)$;
3. Apply replicated k-means clustering on $\boldsymbol{S}_t^{GE}(c)$ to obtain two stable clusters $\hat{V}_t^1$, $\hat{V}_t^2$;
4. Identify larger cluster $\hat{V}_t^{dom} = arg\{\max_w[V_t^w]\}$, where $w \in \{1, 2\}$;
5. Create set $\tilde{C}_{t+1} \subset \tilde{C}_t$ by eliminating all $c \in \hat{V}_t^{dom}$ from $\tilde{C}_t$;
6. $t = t + 1$;
7. *endwhile*;
8. $\tilde{C}_T = \tilde{C}_t$;

end

Note that in general the algorithm is terminated when the total number of spectra is marginally greater than (or equal to) Θ, which usually occurs within 2 to 3 iterations.

4.5 Aim 2: CaP identification via MRS

Having obtained $\tilde{C}_T$, the next aim is to apply more sophisticated analysis to the spectra in $\tilde{C}_T$ to be able to discriminate between different tissue classes in the prostate. Apart from NLDR schemes (Isomaps, LLE, Graph Embedding) that were considered, two other feature extraction schemes (z-score and PCA) were also evaluated for discrimination of the prostate spectra. Following feature extraction, replicated k-means clustering for the DR schemes (LLE, Isomaps, Graph Embedding, PCA) and thresholding for z-score was applied to obtain hard classification of the spectra into CaP and non CaP categories.

4.5.1 Feature Extraction

(a) z-score:

For a set of objects, $\phi^{tr} \subset C$, the mean spectral vector $\boldsymbol{F}^{\mu} = [\boldsymbol{f}_u^{\mu} | u \in \{1, ..256\}]$ and the corresponding standard deviation vector $\boldsymbol{F}^{\sigma} = [\boldsymbol{f}_u^{\sigma} | u \in \{1, ...256\}]$, where $\boldsymbol{f}_u^{\mu} = \frac{1}{|\phi^{tr}|} \sum_{c \in \phi^{tr}} \boldsymbol{f}_u(c)$ and $\boldsymbol{f}_u^{\sigma} = \sqrt{\frac{1}{|\phi^{tr}|} \sum_{c \in \phi^{tr}} [\boldsymbol{f}_u(c) - \boldsymbol{f}^{\mu}(c)]^2}$, are obtained. z-score at each $c \in C$ is then defined as,

$$S^z(c) = \frac{||\boldsymbol{F}(c) - \boldsymbol{F}^{\mu}||_2}{||\boldsymbol{F}^{\sigma}||_2}, \tag{4.5}$$

z-score model gives the degree to which the value of an object deviates from the normal based on a statistical linear model. A predefined threshold θ^z is used to classify each $c \in \tilde{C}_T$

as cancer or benign based on whether $S^z(c) > \theta^z$ or $S^z(c) < \theta^z$.

(b) Dimensionality Reduction:

4 commonly employed DR methods (explained previously in Chapter 3) are applied to the MR spectra in $\tilde{C}_T$ so that for any $c \in \tilde{C}_T$, the high dimensional ambient feature vector $\boldsymbol{F}(c)$ is mapped to $\boldsymbol{S}^{\phi}(c)$, where $\phi \in \{PCA, LLE, ISO, GE\}$. Replicated clustering is then employed to cluster each $\boldsymbol{S}^{\phi}(c)$,$\forall c \in \tilde{C}_T$, into one of the three possible classes, $\hat{V}^{\phi,1}, \hat{V}^{\phi,2}, \hat{V}^{\phi,3}$ corresponding to cancer, benign, or an intermediate tissue class (benign prostatic hyperplasia (BPH), high grade prostatic intraepithelial neoplasia (HGPIN)) with characteristics between that of the cancer and benign classes.

Chapter 5

Results

In Section 5.1, the probabilistic metrics employed for evaluation of the CAD scheme based on the partial knowledge of ground truth extent of CaP are briefly described. In Section 5.2 the qualitative results of Aim 1 (hierarchical clustering) and Aim 2 (CaP detection) of the MRS CAD scheme are presented. Quantitative results for both Aims are explained in Section 5.C.

5.1 Probabilistic Quantitative metrics used for assessing CAD performance in absence of precise CaP ground truth

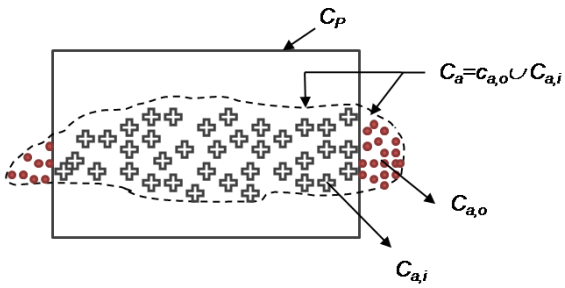


Figure 5.1: Illustrations of the potential ground truth C_P space containing N_g metavoxels corresponding to CaP. C_a represents the CaP segmentation obtained by the MRS CAD analysis scheme and $C_{a,o}$ and $C_{a,i}$ represent those regions of C_a that lie outside and within C_P respectively.

As mentioned previously in Section 4.2, a potential CaP space C_P is defined within which the number of CaP locations N_g is determined. Following replicated clustering of $\boldsymbol{S}^{\phi}(c)$,$c \in \tilde{C}_T$, which of the 3 clusters $\hat{V}^1, \hat{V}^2, \hat{V}^3 \in \tilde{C}_T$ corresponds to the CaP class needs to be identified. The corresponding sets of spatial locations for clusters $\hat{V}^1, \hat{V}^2, \hat{V}^3$ are represented as C_a^1, C_a^2 and C_a^3 respectively. With respect to C_P (as illustrated in Figure 5.1), some part of C_a^1, C_a^2 or C_a^3 may be within ($C_{a,i}^q$, $q \in \{1,2,3\}$) or outside ($C_{a,o}^q$, $q \in \{1,2,3\}$) C_P. Thus $C_a^q = C_{a,o}^q \cup C_{a,o}^q$, $q \in \{1,2,3\}$. The true positive (TP), false positive (FP), true negative (TN) and false negative (FN) ratios associated with each class C_a^q, $q \in \{1,2,3\}$ with respect to C_P and N_g are then obtained. The following heuristic algorithm is then used to identify which of C_a^q, $q \in \{1,2,3\}$ represents the CaP class (C_a^{CaP}).

Algorithm $IdentifyCaPCluster$

Input: $C_a^q, q \in \{1,2,3\}$, C_P, N_g, C.

Output: C_a^{CaP}.

begin

0. *for* $q = 1$ *to 3 do;*
1. $C_{a,i}^q = C_a^q \cap C_P$;
2. $C_{a,o}^q = C_a^q - C_{a,i}^q$;
3. *If* $|C_{a,i}^q| \geq N_q$ *then*

   $TP^q = N_g$, $FP^q = |C_a^q| - N_g$, $FN^q = 0, TN = |C - C_a^q|$;
4. *else* $TP^q = |C_{a,i}^q|$, $FP = |C_{a,o}^q|$, $FN = N_g - |C_{a,i}^q|$, $TN = |C - C_{a,o}^q| - N_g$;
5. *endif;*
6. *endfor;*
7. $C_a^{CaP} = arg\max_q\{[\frac{TP^q}{FP^q}]\}$;

end

As Algorithm $IdentifyCaPCluster$ indicates, the CaP cluster is determined as the one that maximizes true positive fraction while simultaneously minimizing the false positive fraction. Having identified C_a, corresponding sensitivity and specificity values are determined as,

$$SN(C_a^{CaP}) = \frac{TP_a}{TP_a + FN_a} \times 100, \tag{5.1}$$

$$SP(C_a^{CaP}) = \frac{TN_a}{TN_a + FP_a} \times 100, \tag{5.2}$$

Note that the confidence estimate (η) associated with $SN(C_a^{CaP})$ and $SP(C_a^{CaP})$ is a function of $|C_P|$ and N_g and is determined as,

$$\eta = \frac{N_g}{|C_P|} \times 100. \tag{5.3}$$

5.2 Qualitative Results

5.2.1 Aim 1: Qualitative evaluation of the hierarchical clustering scheme

Figure 5.2 shows the qualitative results of the hierarchical cascade scheme for distinguishing informative from non-informative spectra. Figures 5.2 (a), (g) represent spatial maps of the spectral grid $\tilde{C}_0$ (16×16 spectral voxels) superimposed on the corresponding T2-weighted MRI scene for two different patient studies. Every $c \in \tilde{C}_0$ in Figures 5.2 (a), (g) is assigned one of two colors (blue and red), in turn corresponding to informative and non-informative (background) class and determined via replicated clustering on the embedding $\boldsymbol{S}^{GE}(c)$. Note that the dominant cluster (spatial locations in red in Figures 5.2 (a), (g)) has been eliminated in the second iteration ($\tilde{C}_1$ (16×8 spectral voxels)) (Figures 5.2 (b), (h)) in both studies. The final spectral grid ($\tilde{C}_2$ in Figures 5.2 (c), (i)) is obtained after elimination of non-informative peripheral spectra (red locations) during the third iteration, third level of the cascade. Figures 5.2 (d)-(f) and Figures 5.2 (j)-(l) represent the embedding plots

(where each location is plotted in 3D eigen vector space using the 3 dominant embedding values as co-ordinates) from $\tilde{C}_0$ (16×16 spectral voxels) to $\tilde{C}_2$ (7×4 spectral voxels) in (c) and 6×4 spectral voxels in (i)) for two different studies at the 3 different levels of the cascade. Note that for both studies at the end of the third iteration, the spectral grid has been accurately overlaid on the prostate. Further note in Figures 5.2 (a)-(c), and 5.2 (g)-(i) that the spectral grid with the pronounced boundary indicates the region of interest during the current iteration.

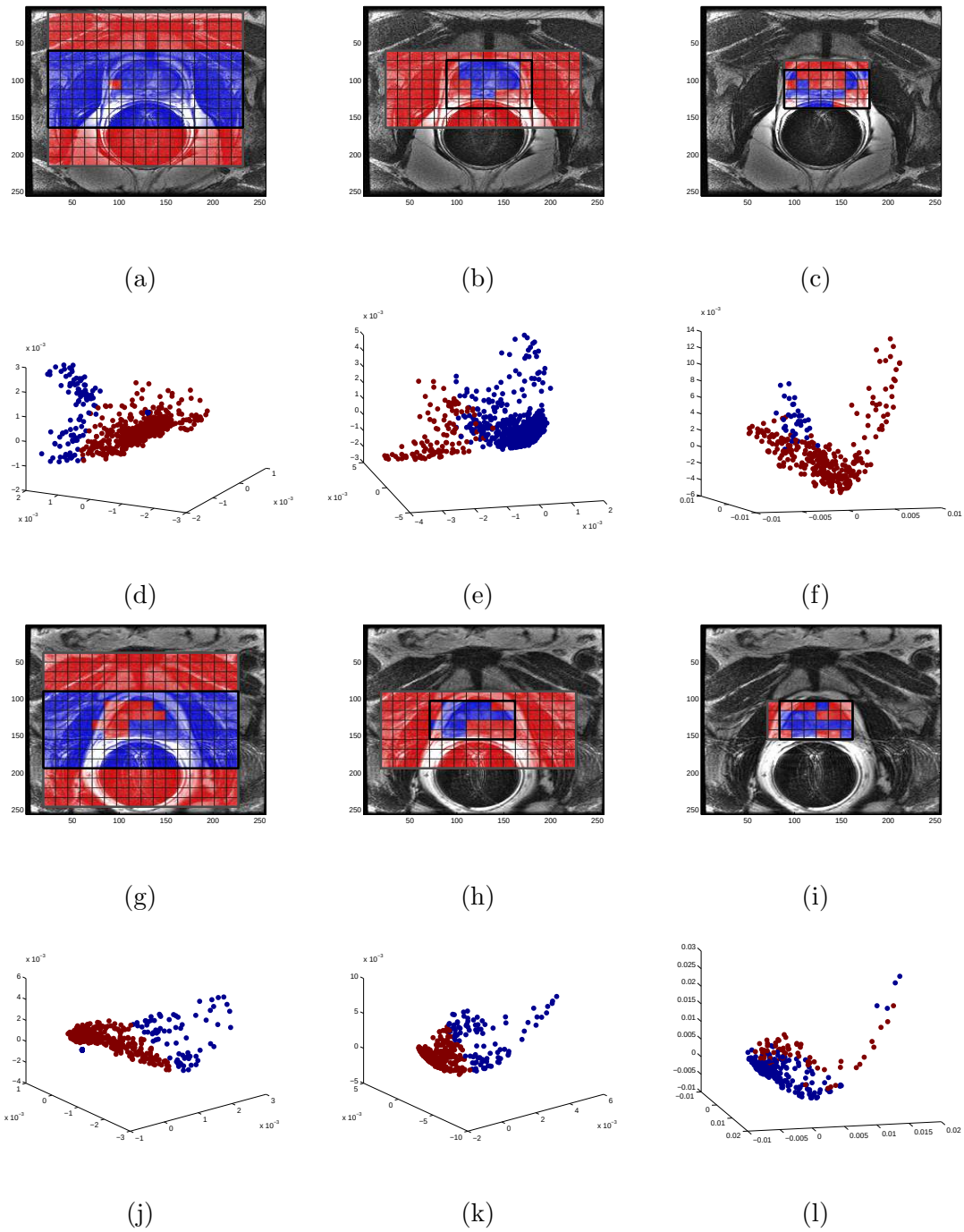


Figure 5.2: Spectral grids for a single slice T2-w MRI scenes for 2 patients at (a), (g) the first cascade level $\tilde{C}_0$, (b), (h) second cascade level $\tilde{C}_1$, and (c), (i) third cascade level $\tilde{C}_2$. Note that the size of the grid reduces from 16×16 metavoxels in (a), (g) to 7×4 in (c) and to 6×4 metavoxels in (i) by elimination of non-informative spectra in the dominant cluster (red). The corresponding clustered embedding plots at each of the cascaded levels are also shown in (d) - (f) and (j) - (l), which correspond to the metavoxel grids shown in (a) - (c) and (g) - (i).

5.2.2 Aim 2: Evaluation of feature extraction schemes for CaP detection

The identification of the prostate grid in Aim 1 allows for the resolvability of the 3 tissue classes (one of which is cancer). The differences between these 3 spectral classes within the informative cluster at the higher levels in the cascade ($\tilde{C}_1, \tilde{C}_2, \tilde{C}_3$) become discriminable only after removal of non-informative spectra. Figure 5.3 shows the qualitative results of the 5 feature extraction schemes employed in this work for CaP detection in for 3 different patient studies, each row in Figure 5.3 corresponding to a different study. The three colors assigned to the spectral voxels in Figure 5.3 correspond to the three clusters obtained via replicated clustering on the reduced dimensional spectra $\boldsymbol{S}^{\phi}(c)$, $\phi \in \{PCA, LLE, GE, ISO\}$. For the z-score scheme, each metavoxel was classified as belonging to one of 2 classes (red and blue in Figures 5.3 (a), (d), (g)). The white box superposed on Figure 5.3 (a)-(i) shows the potential cancer space for corresponding slices. In each of Figures 5.3 (a)-(i) the red cluster was identified as CaP using the $IdentifyCaPCluster$ algorithm (Section 5.A) and following feature extraction and replicated clustering. Figures 5.3(a), (d), (g) show results obtained via z-score; while Figures 5.3 (b), (e), (h) show the corresponding results for PCA. Figures 5.3(c), (f), (l) show similar results when employing Graph Embedding, LLE and Isomaps respectively. For the first study (first row in Figure 5.3), Graph Embedding (Figure 5.3(c)) yields a near perfect CaP detection in terms of sensitivity and specificity as only CaP voxels are identified within the white cancer grid (C_P). The corresponding results for z-score 5.3 (a) and PCA 5.3 (b) both yield poor detection sensitivity and specificity. Similarly for the 2 other studies shown in Figure 5.3, Isomaps (5.3(f)) and LLE (5.3(i)) yield higher CaP detection sensitivity and specificity compared to z-score (5.3(d), (g)) and PCA (5.3(e), (h)). Figure 5.4 shows an example of the MR spectral grid with classification labels obtained from Graph Embedding replicated clustering plotted back on the individual spectra.

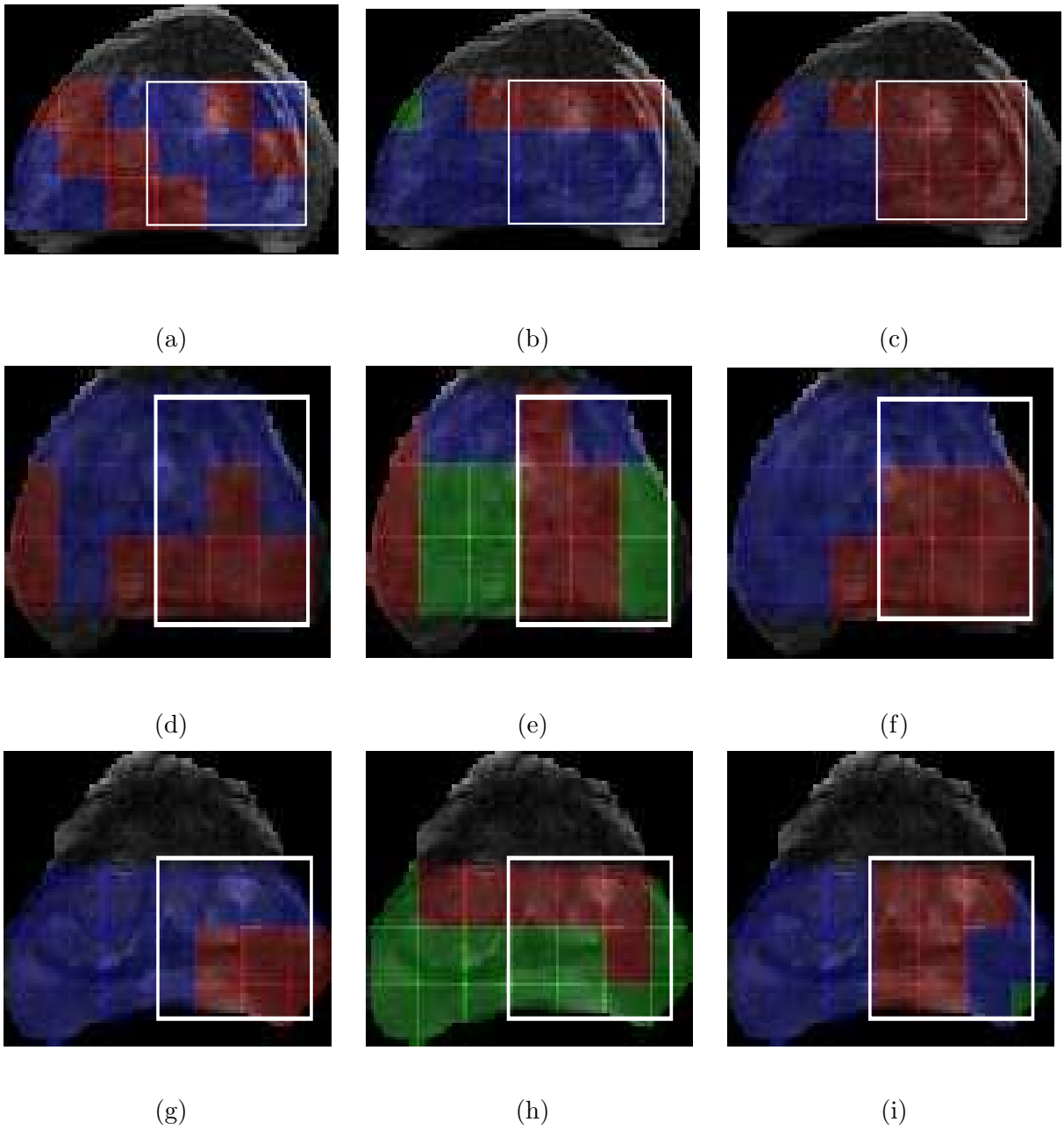


Figure 5.3: CaP detection results via feature extraction schemes employed for three different studies; different rows corresponding to different studies. Figures 5.3 (a), (d), (g) represent the CaP (red voxels) and benign clusters (blue voxels) obtained via z-score, and (b), (e), (h) demonstrate the corresponding results obtained via PCA (red, blue, green (intermediate)). Results for the NLDR schemes (c) GE, (f) LLE and (i) Isomaps are also shown. The white box superposed on Figures 5.3 (a)-(i) shows the locations of the potential cancer space (C_P). In each of Figures 5.3 (a)-(i) the cluster with the red metavoxels was the one identified as the CaP class based on the $IdentifyCaPCluster$ algorithm (Section 5.1).

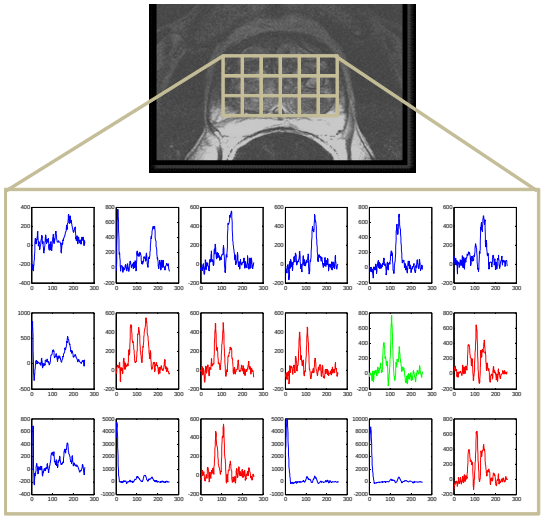


Figure 5.4: MRS spectral grid plotted with the classification labels (three colors correspond to three different clusters) obtained from Graph Embedding and replicated clustering. The red spectra correspond to those identified as CaP, the blue correspond to benign spectra and the green correspond to intermediate spectra.

5.3 Quantitative Results

5.3.1 Quantitative evaluation of Aim 1- Hierarchical clustering

Sensitivity	Specificity	PPV
97.66%	98.87%	89.29%

Table 5.1: Average sensitivity, specificity, PPV values for automated identification of prostate grid using hierarchical spectral clustering averaged over 18 studies.

Following identification of prostate spectra, the largest box C_T that contains all the prostate spectra is then overlaid on the T2-w image. Note that in qualitative evaluation of Aim 1, the precise spatial extent of the prostate is all that is needed. This ground truth is ascertained by manual placement of a spectral grid ($\tilde{C}_g^T$) on the prostate by an expert radiologist. Table 5.1 shows the average sensitivity, specificity and positive predictive value (PPV) in

automated identification of the prostate grid, $\tilde{C}_T$ with respect to $\tilde{C}_g^T$ and averaged over 18 studies.

5.3.2 Quantitative evaluation of Aim 2- CaP detection

(a) z-score analysis:

In order to define the optimal threshold θ^z for performing a z-score based classification of each $c \in C$ as normal or cancerous, a set of cancerous voxels $C^M \subset C$ is defined during an offline training phase. For each $c \in C^M$, corresponding $S^z(c)$ was obtained which was then used to define the average $\mu_{s^z}^M$ and standard deviation $\sigma_{s_z}^M$ of z score values for CaP. The threshold θ^z was then defined as $\mu_{s^z}^M \pm \alpha\sigma_{s^z}^M$, where $\alpha \in [0, 1]$. The value of α was uniformly varied between [0,1] and a corresponding values for θ^z obtained. At each value of θ^z, each metavoxel $c \in \tilde{C}_T$ is identified as belonging to CaP if $S^z(c) > \theta^z$, normal otherwise. Thus at each θ^z, the corresponding sensitivity and specificity of CaP detection via z-score is computed (Equations 5.1, 5.2). A curve is then fit to the sensitivity, specificity values to obtain the Receiver Operating Characteristic (ROC) curve. The optimal threshold θ^z was determined as the operating point on the ROC curve; the location on the ROC curve closest to 100% sensitivity, specificity. Figure 5.5 shows the ROC curve obtained by averaging sensitivity and specificity values for CaP detection as a function of α, over 18 studies.

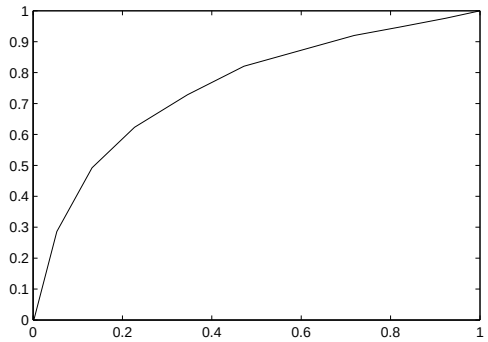


Figure 5.5: ROC curve of the z-score analysis scheme obtained by varying the threshold ($\theta_z, \alpha \in [0, 1]$) averaged over 18 studies.

(b) Evaluation of DR methods:

The ability of a classifier to distinguish between object classes in embedding spaces obtained via DR methods is known to be sensitive to the choice of number of dimensions (v) of the embedding space in which the data is represented. For the NLDR methods, the low dimensional data representations obtained via LLE [20] and Isomaps [19] are also a function of κ, the parameter controlling the size of the local neighborhood within which linearity is assumed. In order to evaluate the parameter sensitivity of different DR methods, the robustness of the DR methods over different values of κ and v was quantitatively evaluated. LLE and Isomaps were evaluated by varying $\kappa \in \{6, 7, ...15\}$ and $v \in \{3, 4, ...10\}$, a total of 80 different combinations of parameter values. PCA and Graph Embedding were evaluated for 8 different values of $v \in \{3, 4, ..., 10\}$.

3 fold and 5 fold cross validation averaged over 20 iterations was also performed for $\phi, \phi \in \{LLE, ISO, GE\}$ to obtain average sensitivity and specificity for all 18 studies. 3 fold cross validation was performed by randomly choosing 3 datasets and calculating average CaP detection sensitivity and specificity across 3 studies on the 80-dimensional parameter space.

The parameter set (v_{max}, κ_{max}) with maximum sensitivity and specificity in this space were then identified as optimal values and used for the CAD scheme on the remaining 15 studies. The average CaP detection sensitivity and specificity on these 15 studies was then recorded. On the next trial, 3 random training studies from 18 were again selected and used to optimize the parameters and evaluation again done on the remaining 15 studies. This entire process was repeated a total of 20 times. The mean and standard deviation in CaP detection sensitivity and specificity across these 20 iterations is reported in Table 5.2 (a) for the 4 DR methods. Note that for PCA and GE only v_{max} was optimized. A similar routine was employed when performing 5 fold cross validation. Corresponding results are reported in Table 5.2 (b). All NLDR schemes employed in this work were found to have higher sensitivity of CaP detection compared to PCA, Graph Embedding performing the best with a sensitivity of almost 90%. PCA, however was found to have marginally higher specificity compared to the NLDR schemes. In Figures 5.6 and 5.7 are shown barplots representing CaP detection sensitivity, specificity for each of the 18 studies considered in this work via 3-fold and 5-fold cross validation respectively over 20 iterations for GE and ISO schemes. Note that both average and standard deviation values over 20 iterations are shown. The confidence estimates (η) associated with the sensitivity, specificity measurements of each study are also reported in Figures 5.6 and 5.7. Note that while η is low for a majority of the studies, for studies 11 and 18, the confidence associated with CaP detection sensitivity, specificity was almost 90% and 80% respectively.

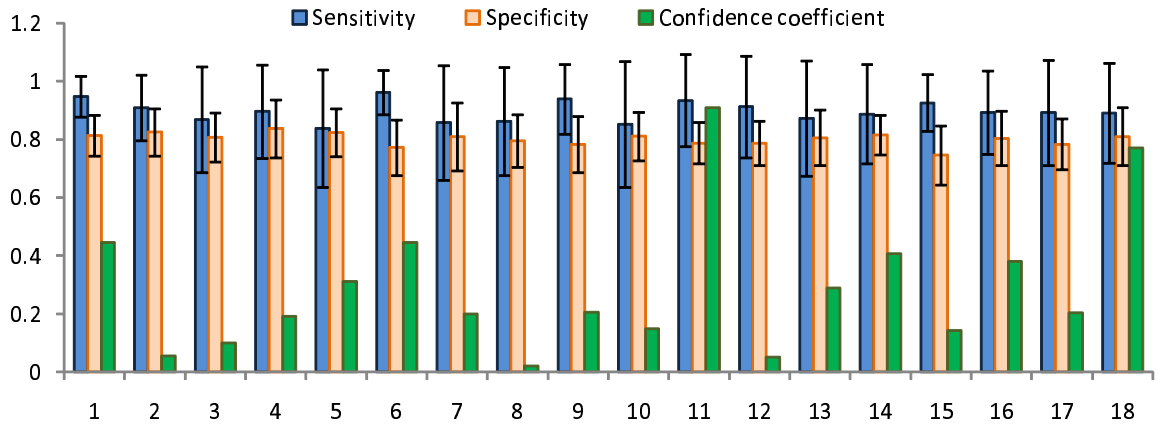


Figure 5.6: Barplot showing average and standard deviation in CaP detection sensitivity and specificity for the individual 18 studies averaged over 20 iterations of 3 fold cross validation via Graph Embedding. Confident coefficient (η) is also shown in green.

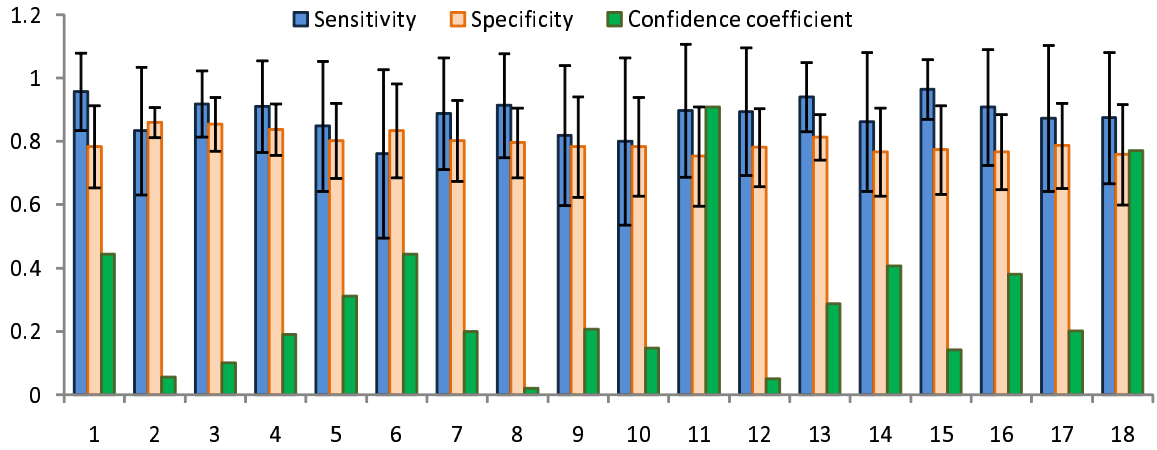


Figure 5.7: Barplot showing average and standard deviation in CaP detection sensitivity and specificity for the individual 18 studies averaged over 20 iterations for 5 fold cross validation via Isomaps. Confident coefficient (η) is also shown in green for each study.

Method	Sensitivity	Specificity
GE	89.33± 1.87	79.79± 2.24
ISO	87.94± 3.16	79.04± 1.73
LLE	83.70± 4.50	81.04± 4.60
PCA	78.53± 3.20	83.97 ± 2.81

(a)

Method	Sensitivity	Specificity
GE	89.09± 2.10	79.22± 2.10
ISO	87.96± 2.88	78.75± 1.97
LLE	84.93± 4.42	81.32± 4.05
PCA	78.78± 4.31	84.11 ± 2.46

(b)

Table 5.2: Table showing the average and standard deviation in CaP detection sensitivity and specificity for different DR methods over $\kappa \in \{5, ..., 15\}$ and $v \in \{3, ..., 10\}$ for 18 different studies via (a) 3 fold cross validation, and (b) 5 fold cross validation.

Chapter 6

Applications of this work

The work presented in this thesis is further extended to two very specific applications, (a) segmentation of the prostate and, (b) integration of metabolic and structural content present in MRI and MRS for automated CaP detection. A brief overview of the applications of the hierarchical unsupervised clustering presented in this work used for segmentation of prostate and CaP detection using integration of MRI-MRS are presented in following two sections.

6.1 Application to automated segmentation of prostate

Segmentation of the prostate boundary on clinical images is useful in a large number of applications including calculation of prostate volume during biopsy, estimating spatial location of cancer in the prostate, and treatment planning. Manual segmentation of the prostate boundary is, however, time consuming and subject to inter- and intra-reader variability. A semi-automated scheme for accurate prostate segmentation on *in-vivo* multi-modal MRI studies is performed using the popular Active Shape Model (ASM). However, ASMs require careful initialization and are extremely sensitive to model initialization. The spatial locations of the prostate spectra identified in Aim 1 are thus used as the initial ROI for the ASM for automated initialization. A bounding box is first constructed to inscribe the prostate spectra obtained from Aim 1. During the training stage of the ASM, a set of prostate

training images are manually landmarked, from which the shape and intensity information is obtained. This training stage is done off-line, and only needs to be done once to construct the trained ASM model comprising both shape and intensity information.The mean shape is transformed to fit in the bounding box previously obtained in Aim 1 from the spectral clustering of the MRS data, which serves to provide the landmark points of the model for the first iteration. The shape is updated to fit the landmark points on the border. Several qualitative results for the initialization of the prostate ASM via spectral clustering of the MRS data are shown in Figure 6.1 [34]. In each of these images, ground truth segmentation by Expert 1 is shown in green, ground truth segmentation by Expert 2 is shown in blue, and the prostate boundary segmentation using the ROI initialization resulting from the spectral clustering of MRS data is shown in red.

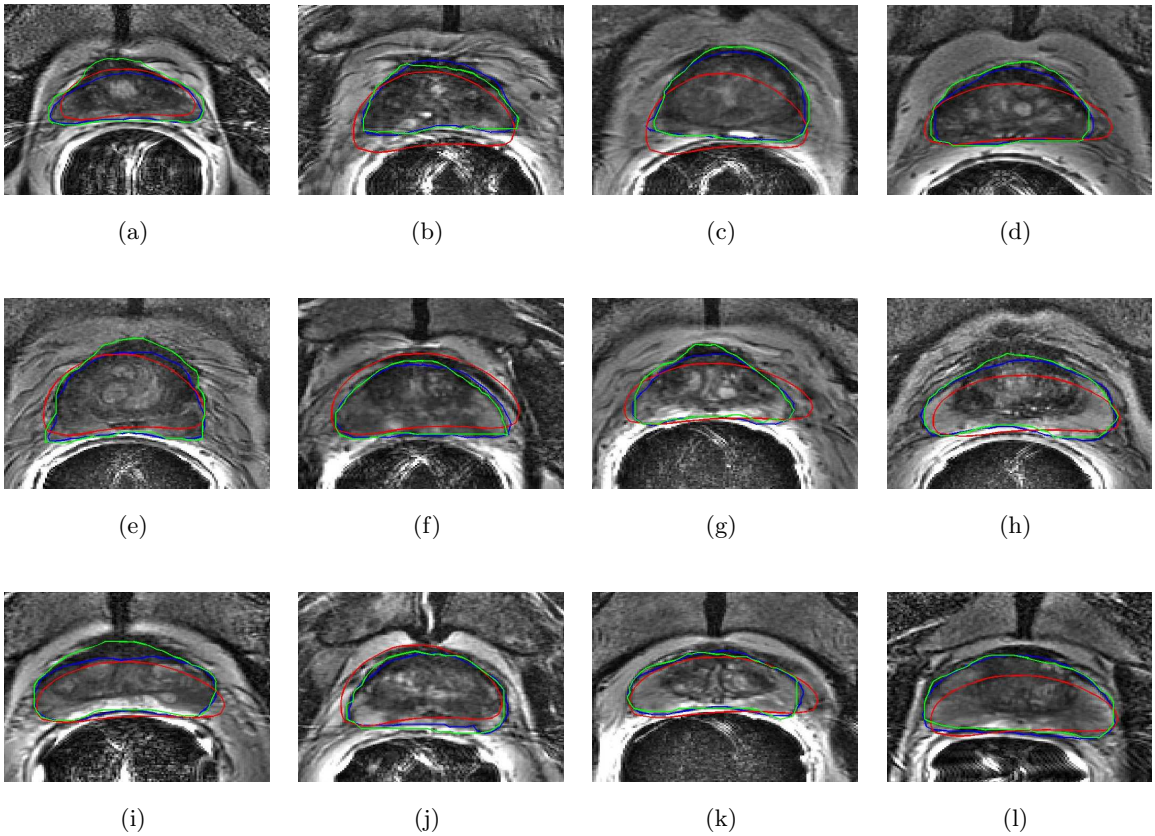


Figure 6.1: Figure (a) - (l) show the initialization results from the spectral clustering of MRS data for 12 different studies. In each image, the green line indicates Expert 1's segmentation, the blue line indicates Expert 2's segmentation, and the red line indicates the initialization from the spectral clustering.

6.2 Integration of MRI-MRS for automated CaP detection

Recently, *in vivo* Magnetic Resonance Imaging (MRI) and Magnetic Resonance Spectroscopy (MRS) have emerged as promising new modalities to aid in prostate cancer (CaP) detection. MRI provides anatomic and structural information of the prostate while MRS provides functional metabolic data pertaining to biochemical concentrations of metabolites such as creatine, choline and citrate. Similar to the CaP detection scheme presented in

this work, a meta-classifier is developed to detect CaP *in vivo* via quantitative integration of multimodal prostate MRS and MRI using non-linear dimensionality reduction (NLDR) methods including spectral clustering and locally linear embedding (LLE). Quantitative integration of multimodal image data (MRI and PET) involves the concatenation of image intensities following image registration. However data integration is non-trivial when the individual modalities include spectral and image intensity data. A data combination solution of MRI-MRS is proposed wherein the feature spaces (image intensities and spectral data) associated with each of the modalities are projected into a lower dimensional embedding space via NLDR. These methods preserve the relationships between the objects from the original high dimensional into the reduced low dimensional space. Since the original spectral and image intensity data are divorced from their original physical meaning in the reduced dimensional space, data at the same spatial location can be integrated by concatenating the respective embedding vectors. Unsupervised clustering is then used to partition objects into different classes in the combined MRS and MRI embedding space. Quantitative results of this multimodal computer-aided diagnosis scheme on 16 sets of patient data obtained from the ACRIN trial, for which corresponding histological ground truth for spatial extent of CaP is known, show a marginally higher sensitivity, specificity, and positive predictive value compared to corresponding CAD results with the individual modalities [35]. Figure 6.2 shows the qualitative results for 2 studies comparing the CAD results for the two modalities individually with the combined structural and spectral information.

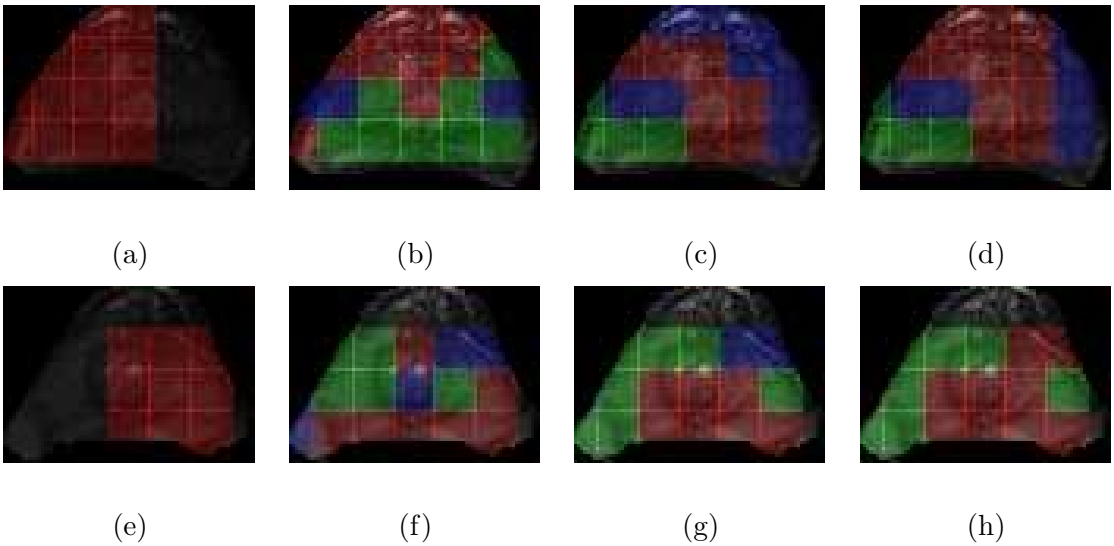


Figure 6.2: (a) and (e) show the location of *potential cancer space* (at metavoxel resolution) shaded in a translucent red on a 2D MRI slice. This is followed by the results of shading the section with different colors (red, blue and green) based on the labels of the objects in clusters: (b) and (f) based on the MRI embedding space (c) and (g) based on the MRS embedding space (d) and (h) based on the integrated embedding space. In each case, the NLDR method used was GE and the labels were obtained via unsupervised clustering. For each of the result images the red region was found to correspond most closely to the potential cancer space in (a) and (e) respectively.

Chapter 7

Concluding Remarks and Future work

In this work a novel application of non linear dimensionality reduction and hierarchical clustering is presented for automated identification of (a) the prostate ROI based on classification of MR spectral data alone, and (b) cancerous spectra within the prostate ROI. Current state of the art peak detection methods for CaP detection on MRS are generally limited in their ability to accurately identify metabolite peaks on account of the low signal to noise ratio of MRS data. The inherent non-linearity associated with the spectral data also limits the applicability of linear DR schemes such as PCA and statistical methods such as z-score. In previous work [18] the efficacy of non-linear DR schemes over linear DR schemes for classification of high dimensional non-linear biomedical data has been demonstrated. The empirical comparisons of 3 non-linear DR schemes (LLE, Isomaps, Graph Embedding) with PCA and z-score via 3 fold and 5 fold cross validation over 18 studies revealed that the non-linear DR schemes consistently had a higher classification sensitivity in terms of CaP detection. In addition, the non-linear DR schemes were found to be relatively robust to changes in the value of the system parameters (v, κ). An improved variant of the popular k-means clustering algorithm called replicated clustering was also employed which yielded consistently stable clusters. Aim 1 of the presented scheme took an average of 11.24 seconds to analyze a $256 \times 256 \times 8$ MRI/ MRS 3D 1.5 T scene on a Pentium IV, 2 GB RAM Intel processor machine, while Aim 2 took an average of 5.23 seconds for automated identification

of CaP.

Owing to the fact that only limited knowledge regarding precise spatial extent of CaP was available for the studies considered in this work a confidence coefficient was defined to assess the degree of certainty associated with the CAD sensitivity, specificity measurement reported. The high confidence estimates associated with 2 of the studies seem to suggest that the consistently high CaP detection sensitivity and specificity measurements for the other studies are not erroneous.

To summarize, the primary contributions of this work are,

1. Novel integration of non-linear dimensionality reduction methods with replicated clustering for analysis of prostate Magnetic Resonance Spectroscopy. Methods employed in this work make no assumption about the data and can be applied (without any modifications) to higher resolution (3 Tesla) MRS or other spectral data. The scheme also has an execution time of under 20 seconds.

2. A fully automated scheme for identifying the prostate region of interest based on spectral classification of the MRS data alone, and

3. An automated CaP detection scheme based on MRS data that appears to be robust to system parameters and more accurate compared to such state of the art MRS analysis schemes as PCA and z-score.

Future work will involve performing more rigorous analysis of the scheme on a larger cohort of datasets and performing similar CAD analysis using 3T MRS/ MRI studies. Aim would also be to develop more robust unsupervised and supervised classification tools to be able to classify spectral information as accurately as possible. Recently we have also proposed a new algorithm, Consensus-Locally Linear Embedding (C-LLE), as an improvement

to a commonly used NLDR scheme, Locally Linear Embedding (LLE) [36] which is known to be highly dependent on the user defined parameter, κ. Further detailed evaluation of the C-LLE algorithm as an improvement over the proposed NLDR based clustering scheme would also be a subject of future research. The aim would also be to explore different unsupervised clustering schemes such as mean shift and EM based clustering as an improvement to the currently employed k-means clustering. Refining the replicated clustering algorithm as a robust clustering scheme would also be considered in future. In order to expand the automated analysis via the hierarchical CAD system one of the aims would be to propose a scheme to be able to automatically classify different grades of CaP from MR spectroscopy based on spectral information, once CaP has been identified using the automated clustering scheme. Future work would also include applying the proposed scheme for the two mentioned applications in a more comprehensive fashion and aim to propose new schemes both for segmentation and classification of prostate based on MR spectral and structural information. Multimodal classification involving spectroscopy would also be an interesting area yet to be explored.

References

[1] Catalona, W., Smith, D., Ratliff, T., Dodds, K., Coplen, D., Yuan, J., Petros, J., Andriole, G., "Measurement of Prostate-Specific Antigen in Serum as a Screening Test for Prostate Cancer.", *New England Journal of Medicine*, vol. 324, no.17, pp. 1156-1161, 1991.

[2] Borboroglu, P., Comer, S., Riffenburgh, R., Amling, C., "Extensive Repeat Transrectal Ultrasound Guided Prostate Biopsy in Patients with Previous Benign Sextant Biopsies.", *The Journal of Urology*, vol. 163(1), pp. 158-162, 2000.

[3] Schiebler, M., Schnall, M., Pollack, H., Lenkinski, R., Tomaszewski, J., Wein, A., Whittington, R., Rauschning, W., Kressel, H., "Current role of MR Imaging in Staging of Adenocarcinoma of the Prostate.", *Radiology*, vol. 189(2), pp. 339-352, 1993.

[4] Coakley, F., Teh, Hui Seong., Qayyum, A., Swanson, M., Lu, Y., Roach, M., Pickett, B., Shinohara, K., Vigneron, D., Kurhanewicz, J., "Endorectal MR Imaging and MR Spectroscopic Imaging for Locally Recurrent Prostate Cancer after External Beam Radiation Therapy: Preliminary experience.", *Radiology*, pp. 441-448, 2004.

[5] Carroll, P., Coakley, F., Kurhanewicz, J., "Magnetic Resonance Imaging and Spectroscopy of Prostate Cancer.", *Reviews in Urology*, vol. 8(1), pp. S4-S10, 2006.

[6] Casciani, E., Gualdi, G.F., "Prostate cancer: Value of Magnetic Resonance Spectroscopy 3D Chemical Shift Imaging", *Abdominal Imaging*, pp.1-10, 2000.

[7] Kurhanewicz, J., Swanson, M., Nelson, S., Vigneron, D., "Combined Magnetic Resonance Imaging and Spectroscopic Imaging Approach to Molecular Imaging of Prostate Cancer.", *Journal of Magnetic Resonance Imaging*, vol. 16, pp. 451-463, 2002.

[8] Weinreb, J., Coakley, F., Kurhanewicz, J., Wheeler, T., Blume, J., Cormack, J., Boparai, K., "MRI and MRSI of Prostate Cancer Prior to Radical Prostatectomy: A Prospective Multi-Institutional Clinicopathological Study.", *RSNA*, Abstract SSJ05-06, 2006.

[9] Madabhushi, Anant., Metaxas, Dimitris., "Advances in Computerized Image Analysis on Breast Ultrasound.", *World Scientific Publishing Co., Inc.(In Press)*,

[10] Madabhushi, A., Feldman, M., Metaxas, D., Tomaszeweski, J., Chute, D., "Automated Detection of Prostatic Adenocarcinoma from High-Resolution Ex Vivo MRI.", *IEEE Transactions on Medical Imaging*, vol. 24(12), pp. 1611-1625, 2005.

[11] Viswanath, S., Rosen, M., Madabhushi, A., "A Consensus Embedding Approach for Segmentation of High Resolution In Vivo Prostate Magnetic Resonance Imagery.", *Society of Photo-Optical Instrumentation Engineers (SPIE 2008)*, vol. 6915, pp. 69150U, 2008.

[12] Chan, I., Wells, W., Mulkern, R., Haker, S., Zhang, J., Zou, K., Maier, S., Tempany, C., Detection of Prostate Cancer by Integration of Line-scan Diffusion, T2-mapping and T2-weighted Magnetic Resonance Imaging; a Multichannel Statistical Classifier, *Medical Physics* vol. 30(9), pp. 2390-2398, 2003.

[13] Van der Veen J.W.C., De Beer R., Luyten P.R., Van Ormondt D., "Accurate quantification of in vivo 31P NMR signals using the variable projection method and prior knowledge.", *Magnetic Resonance in Medicine*, vol. 6(1), pp. 92-98, 1988.

[14] Vanhamme, L., Boogaart, A., and Huffel, S., "Improved Method for Accurate and Efficient Quantification of MRS Data with Use of Prior Knowledge.", *Journal of Magnetic Resonance*, vol. 129(1), pp. 35-43, 1998.

[15] Ratiney, H., Sdika, M., Coenradie, Y., Cavassila, S., van Ormondt, D., Graveron-Demilly, D., "Time-domain semi-parametric estimation based on a metabolite basis set.", *NMR in biomedicine*, vol. 18(1), pp. 1-13, 2005.

[16] Zakian, K., Sircar, K., Hricak, H., Chen, H., Dave, A., Eberhardt, S., Muruganandham, M., Ebora, L., Kattan, M., Reuter, V., Scardino, P., Koutcher, J., "Correlation of proton MR Spectroscopic Imaging with gleason score based on sept-section Pathologic Analysis after Radical Prostatectomy.", *Radiology*, vol. 234, pp. 804-814, 2005.

[17] Pels, P., Ozturk-Isik, E., Swanson, M., Vanhamme, L. Kurhanewicz, J., Nelson, S., Huffel, S., "Quantification of prostate MRSI data by model-based time domain fitting and frequency domain analysis.", *NMR in biomedicine*, vol. 19, pp. 188-197, 2006.

[18] Lee, G., Madabhushi, A., and Rodriguez, C., "Investigating the Efficacy of Nonlinear Dimensionality Reduction Schemes in Classifying Gene and Protein Expression Studies." *IEEE/ACM Transactions on Computational Biology and Bioinformatics,* vol. 5, pp. 368-384, 2008.

[19] Tenenbaum, Joshua, Vin de Silva, Langford, John C., "A Global Geometric Framework for Nonlinear Dimensionality Reduction.", *Science*, vol. 290, pp. 2319-2322, 2000.

[20] S. Roweis and L. Saul, Nonlinear Dimensionality Reduction by Locally Linear Embedding, *Science* vol. 290, pp. 2323-2326, 2000.

[21] Madabhushi, A., Shi, J., Rosen, M., Tomasezweski, J., Feldman, M., Graph Embedding to Improve Supervised Classification and Novel Class Detection: Application to Prostate Cancer, *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2005)*,vol. 3749, pp. 729-737, 2005.

[22] Jianbo Shi and Jitendra Malik, "Normalized Cuts and Image Segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*,vol. 22(8), pp. 888-905, 2000.

[23] Kelm, M., Menze, B., Zechmann, C., Baudendistel, K., Hanprecht, F., "Automated estimation of tumor probability in prostate MRSI: Pattern recognition vs. quantification.", *Magnetic Resonance in Medicine*, vol. 57, pp.150-159, 2006.

[24] Quan, H., Bao, S., "z-score analysis for H-MRSI data of Glioma and Prostate Cancer.", *Asia-Oceania Federation of Organizations for Medical Physics*, 2004.

[25] Laudadio, T., Pels, P., Lathauwer, L., Hecke, P., Huffel, S., "Tissue Segmentation and Classification of MRSI Data Using Canonical Correlation Analysis.", *Magnetic Resonance in Medicine*, vol. 54, pp. 1519-1529, 2005.

[26] Devos, A., Lukas, L., Suykens, J.A.K., Vanhamme, L., Tate, A.R., Howe, F.A., Majs, C., Moreno-Torres, A., van der Graaf, M., Ars, C., Van Huffel, S. "Classification of brain tumours using short echo time 1H MR spectra.", *Journal of Magnetic Resonance*, vol. 170, pp. 164-175, 2004

[27] Ma, J., Sun, J., "MRS Classification based on Independent Component Analysis and Support Vector Machines.", *IEEE Intl. Conf. on Hybrid Intel. Syst.*, pp. 81-84, 2005.

[28] Simonetti, AW., Melssen, WJ., Szabo de Edelenyi, F., van Asten, JJ., Heerschap, A., Buydens, LM., "Combination of Feature-Reduced MR Spectroscopic and MR Imaging Data for Improved Brain Tumor Classification.", *NMR in Biomedicine*, vol. 18, pp. 34-43, 2005.

[29] Hotelling, H., "Analysis of a Complex of Statistical Variables into Principal Components.", *Journal of Educational Psychology*, vol. 24, pp. 417-441, 1933.

[30] Venna, Jarkko, Kaski, Samuel., Local multidimensional scaling. *Neural Networks*, vol. 19, pp. 889-899, 2006.

[31] Shi, J., Malik, J., "Normalized Cuts and Image Segmentation.", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.22(8), pp. 888-905, 2000.

[32] Chappelow, J., Madabhushi, A., Rosen, M., Tomaszeweski, J., Feldman, M., "Multimodal Image Registration of ex vivo 4 Tesla Prostate MRI with Whole Mount Histology for Cancer Detection.", *Proceedings of SPIE Medical Imaging (SPIE 2007)*, pp. S1-S12, 2008.

[33] McKnight, T., Noworolski, S., Vigneron, D., Nelson, S., "An Automated Technique for the Quantitative Assessment of 3D-MRSI Data from Patients with Glioma.", *Journal of Magnetic Resonance Imaging*, vol. 13, pp. 167-77, 2001.

[34] Toth, R., Tiwari, P., Rosen, M., Madabhushi, A., Kalyanpur, A., Pungavkar, S., "An Integrated Multi-modal Prostate Segmentation Scheme by Combining Magnetic Resonance Spectroscopy and Active Shape Models", *SPIE Medical Imaging*, vol. 6914(1), 2008.

[35] Viswanath, S., Tiwari, P., Rosen, M., Madabhushi, A., "A meta-classifier for detecting prostate cancer by quantitative integration of in vivo magnetic resonance spectroscopy and magnetic resonance imaging", *Proceedings of SPIE Medical Imaging*, vol. 6915, 2008.

[36] Tiwari, P., Rosen, M., Madabhushi, A., "Consensus-Locally Linear Embedding (C-LLE): Application to Prostate Cancer Detection on Magnetic Resonance Spectroscopy", *MICCAI*, 2008.

[37] MacQueen, J.B., "Some Methods for classification and Analysis of Multivariate Observations", *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, vol. 1, pp. 281-297, 1967.

Curriculum Vita

Pallavi Tiwari

Education

- M.S. in Biomedical Engineering, Rutgers University, USA, 2006-2008.
- B.E in Biomedical Engineering, S.G.S.I.T.S, Indore, India, 2002-2006.

Patents

- "Use of Magnetic Resonance Spectroscopy and Magnetic Resonance Imaging In Vivo for Computer- aided Diagnosis of Prostate Cancer", Provisional Patent Filed, RU Docket 08-007, Oct. 2007.
- "Use of Independent Component Analysis and Non-linear dimensionality reduction for detection of prostate cancer from in vivo High Resolution Magnetic Resonance Spectroscopy", Provisional Patent Filed, RU Docket 08-007A, May. 2008.

Publications

Peer-reviewed Journal paper

- Tiwari P, Rosen M, Madabhushi A, A Hierarchical Consensus Clustering Scheme for Detection of Prostate Cancer from Magnetic Resonance Spectroscopy (MRS) via Non-Linear Dimensionality Reduction, Medical Physics 2008 (Under Review).
- Toth R, Tiwari P, Rosen M, Madabhushi A, Active Shape Model For Prostate Segmentation Using Magnetic Resonance Spectroscopy For Initialization, IEEE Transactions on Medical Imaging 2008 (Under Review).