Published in final edited form as:
Med Image Anal. 2013 February ; 17(2): 219–235. doi:10.1016/j.media.2012.10.004.

Multi-kernel graph embedding for detection, Gleason grading of prostate cancer via MRI/MRS

Pallavi Tiwari[a], **John Kurhanewicz**[b], and **Anant Madabhushi**[a,1]

Pallavi Tiwari: pallavi.tiwari@case.edu; Anant Madabhushi: anant.madabhushi@case.edu

[a]Department of Biomedical Engineering, Case Western Reserve University, 10900 Euclid Avenue Cleveland, OH 44106-7207, United States

[b]Department of Radiology and Biomedical Imaging, University of California, San Francisco, 1700 4th Street, San Francisco, CA, United States

Abstract

Even though 1 in 6 men in the US, in their lifetime are expected to be diagnosed with prostate cancer (CaP), only 1 in 37 is expected to die on account of it. Consequently, among many men diagnosed with CaP, there has been a recent trend to resort to active surveillance (wait and watch) if diagnosed with a lower Gleason score on biopsy, as opposed to seeking immediate treatment. Some researchers have recently identified imaging markers for low and high grade CaP on multi-parametric (MP) magnetic resonance (MR) imaging (such as T2 weighted MR imaging (T2w MRI) and MR spectroscopy (MRS)). In this paper, we present a novel computerized decision support system (DSS), called Semi Supervised Multi Kernel Graph Embedding (SeSMiK-GE), that quantitatively combines structural, and metabolic imaging data for distinguishing (a) benign versus cancerous, and (b) high- versus low-Gleason grade CaP regions from *in vivo* MP-MRI. A total of 29 1.5 Tesla endorectal pre-operative *in vivo* MP MRI (T2w MRI, MRS) studies from patients undergoing radical prostatectomy were considered in this study. Ground truth for evaluation of the SeSMiK-GE classifier was obtained via annotation of disease extent on the preoperative imaging by visually correlating the MRI to the *ex vivo* whole mount histologic specimens. The SeSMiK-GE framework comprises of three main modules: (1) multi-kernel learning, (2) semi-supervised learning, and (3) dimensionality reduction, which are leveraged for the construction of an integrated low dimensional representation of the different imaging and non-imaging MRI protocols. Hierarchical classifiers for diagnosis and Gleason grading of CaP are then constructed within this unified low dimensional representation. Step 1 of the hierarchical classifier employs a random forest classifier in conjunction with the SeSMiK-GE based data representation and a probabilistic pairwise Markov Random Field algorithm (which allows for imposition of local spatial constraints) to yield a voxel based classification of CaP presence. The CaP region of interest identified in Step 1 is then subsequently classified as either high or low Gleason grade CaP in Step 2. Comparing SeSMiK-GE with unimodal T2w MRI, MRS classifiers and a commonly used feature concatenation (COD) strategy, yielded areas (AUC) under the receiver operative curve (ROC) of (a) 0.89 ± 0.09 (SeSMiK), 0.54 ± 0.18 (T2w MRI), 0.61 ± 0.20 (MRS), and 0.64 ± 0.23 (COD) for distinguishing benign from CaP regions, and (b) 0.84 ± 0.07 (SeSMiK),0.54 ± 0.13 (MRI), 0.59 ± 0.19 (MRS), and 0.62 ± 0.18 (COD) for distinguishing high and low grade CaP using a leave one out cross-validation strategy, all evaluations being performed on a per voxel basis. Our results suggest that following further rigorous validation, SeSMiK-GE could be developed into a powerful diagnostic and prognostic tool for detection and grading of

Correspondence to: Pallavi Tiwari, pallavi.tiwari@case.edu.
[1]Tel: +1 732 445 4500x6216.

CaP *in vivo* and in helping to determine the appropriate treatment option. Identifying low grade disease *in vivo* might allow CaP patients to opt for active surveillance rather than immediately opt for aggressive therapy such as radical prostatectomy.

Keywords

Prostate cancer; Grading; Data integration; Graph embedding; Semi-supervised

1. Introduction

Gleason grading is the most widely used grading scheme for prostate cancer (CaP) (Gleason, 1966), with low Gleason scores being associated with better patient outcome and high Gleason scores tending to be correlated with more biologically aggressive disease and worse prognoses for long-term, metastasis-free survival (Epstein, 2006; Bostwick, 1994). Currently, over one million prostate biopsies are performed annually in the US, of which approximately 60–70% are negative for CaP presence (Welch et al., 2007). 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 (Potosky et al., 2004). Most of the otherwise correctly diagnosed CaP cases are identified as low grade disease, and are not destined to metastasize (Klein, 2004). Such patients are now opting for a “wait and watch policy” involving active surveillance, as opposed to opting for immediate aggressive therapy (Klotz, 2005).

Multi-parametric (MP) Magnetic Resonance (MR) Imaging (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 MR Spectroscopy (MRS) for staging and possible screening of CaP (Carroll et al., 2006; Kurhanewicz et al., 2002; Kurhanewicz et al., 1996) with a view to reduce unnecessary biopsies in men, with elevated PSA but without CaP, who might otherwise have a significant risk of sexual, urinary, and bowel related symptoms caused due to biopsy (Potosky et al., 2004). MRS, a non-imaging MRI protocol used in conjunction with T2-weighted (T2w) MRI to improve specificity of CaP detection, quantifies the metabolic concentrations of specific molecular markers such as choline (A_{ch}), creatine (A_{cr}), and citrate (A_{cit}) in the prostate (Heerschap et al., 1997). 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 (Heerschap et al., 1997; Zakian et al., 2003). However, the utility of MRS metabolic features for detecting, localizing, and characterizing disease is highly dependent on the quality of MR 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) (Wetter et al., 2006).

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 (Langer et al., 2010; Shukla-Dave et al., 2007, 2009). It has been qualitatively demonstrated in clinical studies that high Gleason grade is associated with elevated ratios of A_{ch+cr}/A_{cit} (Zakian et al., 2005). Hypo-intense signal intensities on T2w MRI are also found to be significantly correlated with CaP aggressiveness (Wang et al., 2008). In Shukla-Dave et al. (2007), qualitatively combining T2w MRI and MRS allowed for accurately predicting the presence of low grade CaP. In a similar related MP study, Shukla-Dave et al. (2009) 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. 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.

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 (May et al., 2001; McLean et al., 1997). It is hence desirable to build a decision support system (DSS) that can (a) quantitatively integrate relevant MP MRI data to create meta-classifiers that can identify disease presence *in vivo* (Chan et al., 2003; Vos et al., 2010; Liu et al., 2009; Langer et al., 2009; Ozer et al., 2010) and, (b) can subsequently characterize the Gleason grade of areas ascertained by the meta-classifier, most likely to be CaP. Such a DSS 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 (for instance, a MR spectral signature at a single voxel may be comprised of concentrations from few hundred metabolites, while simultaneously characterized by a scalar T2w MR image intensity).

In this paper we present a novel computerized DSS called Semi-Supervised Multi-Kernel Graph Embedding (SeSMiK-GE), that quantitatively combines structural, and metabolic information from MP MRI for distinguishing (a) benign versus CaP, and (b) high versus low-grade CaP regions. To the best of our knowledge, this work represents the first quantitative DSS system to characterize the extent of high grade CaP using MP MRI data. Additionally, the data fusion and meta-classifier framework presented in this paper is extensible and applicable to other modalities and disease systems.

The rest of the paper is organized as follows. In Section 2, we provide a detailed description of previous related work. Section 3 provides a brief overview of SeSMiK-GE and the novel contributions of this work. A detailed mathematical formulation of SeSMiK-GE framework is provided in Section 4. In Section 5, we leverage SeSMiK-GE for integration of T2w MRI-MRS for distinguishing (a) CaP and benign, and (b) high grade versus low grade voxels. Results of CaP detection and grading are presented and discussed in Section 6, followed by concluding remarks in Section 7.

2. Previous related work

Broadly speaking, MP data fusion strategies, may be categorized as combination of data (COD) (where the information from each channel is combined prior to classification), or combination of interpretations (COI) (where independent classifications based on the individual channels are combined) (Rohlfing et al., 2005). COD involves combining the data prior to classification, while COI involves training individual data classifiers (uni-modal classification) and combining the outputs from each classifier. 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 (Lee et al., 2009). Consequently, several COD strategies with the express purpose of building integrated quantitative meta-classifiers have been proposed (Lee et al., 2009; Tiwari et al., 2012; Liu et al., 2009; Ozer et al., 2010).

Below we discuss previous related work in developing individual unimodal classifiers for T2w MRI and MRS, and multi-modal COD meta-classifiers for CaP detection using MP MRI. To the best of our knowledge, no work in combining MP data for identifying high versus low grade CaP with regards to a DSS scheme exists.

2.1. DSS strategies for CaP detection via T2w MRI

Madabhushi et al. (2005) presented a supervised DSS system for detection of CaP from 4 Tesla (T) *ex vivo* prostate T2w MRI where 33 3D texture features (statistical, gradient, and Gabor) were quantitatively extracted at each voxel (T2w 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. Viswanath et al. (2012) developed a classifier strategy to obtain distinct quantitative imaging signatures for central and peripheral CaP on 25 patients who underwent T2w endorectal *in vivo* MRI.

2.2. DSS strategies for CaP detection via MRS

Kelm et al. (2007) presented a comparative study of classification techniques for prostate MRS data based on pattern recognition methods such as Principal Component Analysis (PCA) and Independent Component Analysis against quantification based feature extraction methods using Support Vector Machines, Random Forests, and Gaussian process classifiers. Tiwari et al. (2009) presented a DSS 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 locations in the prostate.

2.3. DSS strategies for CaP detection via MP-MRI channels

In (Chan et al., 2003), a statistical classifier which integrated texture features from 1.5 T *in vivo* MP MRI was presented to generate a statistical probability map representing likelihoods of cancer for different regions within the prostate. Liu et al. (2009) examined *in vivo* MP MRI maps (T2w, DCE, DWI) within a fuzzy Markov Random Field (MRF) framework. The maps were generated via curve fitting of data from each of the protocols with the region of interest (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. (2008) reported a statistically significant improvement in detection accuracy in the combined use of image intensity features from both DCE and T2w MRI data for the classification of CaP, compared to the use of individual modalities. A comparison of different supervised and unsupervised methods for CaP segmentation using MP MRI was presented in (Ozer et al., 2010). However, most of these methods involve integrating imaging-imaging protocols, and it is not apparent that these schemes can be extended to accommodate imaging and non-imaging parameters (e.g. MRS), especially when the different channels to be combined reside in different dimensional spaces.

Recently, Tiwari et al. (2012) presented a DSS method combining information from the T2w MRI and MRS channels for CaP diagnosis via a combination of wavelets and PCA. However, PCA which is a linear DR scheme is typically unable to handle situations where the overlapping classes are embedded on non-linear manifolds – a situation that is more amenable to multi-kernel learning (MKL) (Lanckriet et al., 2004) and graph embedding (GE) (Shi and Malik, 2000) schemes. Kernels (Scholkopf and Smola, 2001) 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 (Lanckriet et al., 2004) involves computing similarity matrices for kernels derived from the individual modalities being combined, so that fused classifiers (within the fused kernel space) can be built. Lanckriet et al. (2004) 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 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 (Duda et al., 2000).

To avoid the curse of dimensionality, several DR schemes have been proposed (Roweis and Saul, 2000, Tenenbaum et al., 2000, Shi and Malik, 2000). One such DR strategy is GE (Shi and Malik, 2000), which accounts for the non-linearities in the data by constructing a similarity graph $G = (V, W)$, where V corresponds to the set of vertices that connect pairwise points and W is a $n \times n$ weight matrix, reflecting the feature similarity between the n data points. The similarity graph is then reduced to lower dimensions by solving a simple Eigenvalue decomposition problem. However GE, like most other DR schemes, is unsupervised and often leads to overlapping clusters resulting in poor class discriminability. Recently several semi-supervised DR (SSDR) schemes have been proposed where some class label information is employed in the construction of a pairwise similarity matrix. Sugiyama et al. (2008) applied semi supervised-learning (SSL) to Fisher's discriminant analysis in order to find projections that maximize class separation. Sun and Zhang (2009) implemented a semi-supervised version of PCA by exploiting between-class and within-class scatter matrices. Semi-supervised GE (Zhao, 2006) involves constructing 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.

3. Brief overview of SeSMiK-GE and novel contributions

Our SeSMiK-GE strategy 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 (Shi and Malik, 2000), 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 multimodal data integration strategies for biomedical applications (Lanckriet et al., 2004; Damoulas and Girolami, 2008; Shi et al., 2010), none of these methods have been employed in conjunction with DR. It is worth noting that although other NLDR schemes such as LLE (Roweis and Saul, 2000) and Isomaps (Tenenbaum et al., 2000) are also popular for DR purposes, GE was our method of choice for SeSMiK framework since unlike LLE (Roweis and Saul, 2000) and Isomaps (Tenenbaum et al., 2000), 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 (Lin and Zha, 2008; Zhang et al., 2011).

Fig. 1 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 $\widehat{K}=\sum_{m=1}^{M}\beta_m K_m$, where $K_m, m \in \{1,2,\ldots,M\}$, 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 a semi-supervised variant of graph embedding (Zhao, 2006), 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, $\widehat{W}=\frac{1}{M}\sum_{i=1}^{M}\tilde{W}$.
3. *Module 3: Dimensionality reduction:* $\hat{K}$ and $\hat{W}$ are employed in a generalized kernel graph embedding (KGE) framework to obtain the integrated low dimensional data representation.

The preliminary results on application of SeSMiK-GE to distinguishing CaP versus benign and high versus low Gleason grade CaP regions by integrating T2w MRI and MRS were previously presented in (Tiwari et al., 2010). The new work builds on and extends the previously presented framework (Tiwari et al., 2010) by (a) incorporating a hierarchical framework for selectively identifying high grade CaP regions, only within the regions first identified as CaP, (b) employing a MRF classifier to employ spatial constraints to the CaP classifier output, and (c) extensive evaluation of our hierarchical classification scheme over 29 patient studies using both leave-one-out (LOO), and threefold cross-validation (CV) strategies. Fig. 2 provides an overview of the hierarchical scheme that leverages SeSMiK-GE for CaP detection and grading. In Step 1, a random forest (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 (Monaco et al., 2010) is then leveraged to impose spatial constraints to the RF classifier result, yielding a lesion segmentation. CaP lesions identified in Step 1, are then further distinguished in Step 2 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 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).

4. Theory 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.

4.1. Notation

Let $\mathcal{F} = [\mathbf{F}(c_1), \mathbf{F}(c_2),\ldots,\mathbf{F}(c_n)] \in \mathbb{R}^D$ be a data matrix of n objects, c_i, $i \in \{1,\ldots, n\}$, 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_i, $i,j \in \{1,\ldots, n\}$. The diagonal matrix $\mathcal{D}$ and Laplacian matrix L of a graph G is defined as: $\mathcal{D}_{ii} = \Sigma_j w_{ij}$, where $L = \mathcal{D} - W$. A kernel gram matrix defining similarities between n points is given as K_m for protocol m, $m \in \{1,\ldots,M\}$, where M is the total number of protocols (or data channels).

4.2. Graph embedding framework

The aim of GE (Shi and Malik, 2000) is to reduce the data matrix $\mathcal{F} \in \mathbb{R}^D$ into a low-dimensional space $\mathbf{y} \in \mathbb{R}^d (D \gg d)$, 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 $\mathbf{y} = [y_1, y_2,\ldots,y_n]$ can be obtained by solving an Eigenvalue decomposition problem as,

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

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, $i,j \in \{1,\ldots,n\}$ such that $w_{ij} = \frac{-\| F(c_i) - F(c_j) \|}{\sigma}$, where σ is a scaling parameter.

4.3. Kernel Graph Embedding (KGE) framework

KGE is a technique to extend linear projections of data to a nonlinear dot product space using the kernel trick (Muller et al., 2001), 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 $K_m = [K(\mathbf{F}(c_i),\mathbf{F}(c_j))]$, $\forall i,j \in \{1,\ldots, n\}$, where K_m may be expressed as

$$K_m = \begin{bmatrix} K(\mathbf{F}(c_1),\mathbf{F}(c_1)) & K(\mathbf{F}(c_1),\mathbf{F}(c_2)) & \cdots & K(\mathbf{F}(c_1),\mathbf{F}(c_n)) \\ K(\mathbf{F}(c_2),\mathbf{F}(c_1)) & K(\mathbf{F}(c_2),\mathbf{F}(c_2)) & \cdots & K(\mathbf{F}(c_2),\mathbf{F}(c_n)) \\ \vdots & \vdots & \ddots & \vdots \\ K(\mathbf{F}(c_n),\mathbf{F}(c_1)) & K(\mathbf{F}(c_n),\mathbf{F}(c_1)) & \cdots & K(\mathbf{F}(c_n),\mathbf{F}(c_n)) \end{bmatrix}_m . \quad (2)$$

According to the Representer Theorem (Schlkopf et al., 2001), 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,

$$\mathbf{y} = \sum_{j=1}^{n} \alpha_j K(\mathcal{F}(.), \mathbf{F}(c_j)) = K(\mathcal{F})^T \boldsymbol{\alpha}, \quad (3)$$

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

$$\mathbf{y} = K_m \boldsymbol{\alpha}, \quad (4)$$

where K_m is the kernel matrix $K_{ij} = K(\mathbf{F}(c_i), \mathbf{F}(c_j))$. Using Eq. (4), the objective function for KGE can be reduced to,

$$\boldsymbol{\alpha} = \underset{\alpha}{\operatorname{argmax}} [\boldsymbol{\alpha}^T K \mathcal{D} K \boldsymbol{\alpha}], \quad (5)$$

where K is a valid positive semi-definite kernel and $\boldsymbol{\alpha}$ is the d dimensional Eigenvector of the objective kernel function in Eq. (5). As shown in (Schlkopf et al., 2001), optimization function of KGE can again be solved by an Eigenvalue decomposition problem as,

$$KWK\boldsymbol{\alpha}=\lambda K\mathscr{D}K\boldsymbol{\alpha}, \quad (6)$$

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

4.4. Semi-Supervised Multi-Kernel Graph Embedding (SeSMiK-GE)

4.4.1. 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, $K_m, m \in \{1,\ldots, M\}$, with corresponding individual kernel weights β_m, the combined kernel function may be expressed as,

$$\widehat{K}(\mathbf{F}(c_i),\mathbf{F}(c_j))=\sum_{m=1}^{M}\beta_m K_m(\mathbf{F}(c_i),\mathbf{F}(c_j)),\beta_m \geq 0, \forall i,j \in \{1,\ldots,n\}, =\sum_{m=1}^{M}\beta_m K_m, \beta_m \geq 0. \quad (7)$$

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

4.4.2. Module 2. Semi-Supervised Graph Embedding (SSGE)—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 $\mathscr{N}$ nearest neighbor graph, $\mathscr{N} > 0$, is created to obtain W such that pairwise points in a $\mathscr{N}$ neighborhood with same labels are given high weights and points with different class labels are given low weights (Zhao, 2006). If the points are not in $\mathscr{N}$ the corresponding edges are not connected. Thus the weight matrix may be expressed as,

$$\tilde{w}_{ij}=\begin{cases} \gamma(1+\gamma), & \text{if} c_i \in \mathscr{N}_j \quad \text{or} c_j \in \mathscr{N}_i \quad \text{and} \omega_i=\omega_j, \\ \gamma(1-\gamma), & \text{if} c_i \in \mathscr{N}_j \quad \text{or} c_j \in \mathscr{N}_i \quad \text{and} \omega_i \neq \omega_j, \\ \gamma, & \text{if} c_i \in \mathscr{N}_j \quad \text{or} c_j \in \mathscr{N}_i, \quad i>l \text{or} j>l, \\ 0, & \text{otherwise}, \end{cases} \quad (8)$$

where $\gamma = e^{\frac{-\| F(c)_i - F(c)_j \|^2}{\sigma}}$ The Gaussian weight matrix $\tilde{W}_m = [\tilde{W}_{ij}]$, is normalized by σ such that $\sigma = \max(\| F(c_i) - F(c_j) \|^2) \; \forall i,j$ for each individual data channel m, $m \in \{1,\ldots, M\}$. 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, $\tilde{W}_m$, $m \in \{1,\ldots, M\}$ can then be averaged to obtain,

$$\widehat{W}=\frac{1}{M}\sum_{m=1}^{M}\widehat{W}_m, \quad (9)$$

where $\tilde{W}_m$ is the modified weight matrix obtained using Eq. (8) 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.

4.4.3. 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 Eq. (7) and $\hat{W}$ from Eq. (9), Eq. (6) is reduced to a multi-kernel Eigenvalue decomposition problem as,

$$\widehat{K}\widehat{W}\widehat{K}^T\overline{\alpha}=\lambda\widehat{K}\widehat{\mathscr{D}}\widehat{K}^T\overline{\boldsymbol{\alpha}}, \quad (10)$$

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

4.5. SeSMiK-GE optimization

In Eq. (10), 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.

4.5.1. Optimizing the low dimensional representation ($\bar{\boldsymbol{\alpha}}$)—The optimal d dimensional Eigenvectors $\bar{\boldsymbol{\alpha}} = \{\bar{\alpha}_1, \ldots, \bar{\alpha}_d\}$, $d \ll D$ are obtained from Eq. (10) using standard kernel ridge regression optimization as described in (Cai et al., 2007). 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 $\bar{\alpha}$, we make use of the regularized solution of Eq. (4). The optimization of Eq. (10) can then be solved via a two step process:

- **(1.1)** Solve the Eigenvalue decomposition problem as given in Eq. (1) for the combined data matrix $\hat{W}\boldsymbol{y} = \lambda\hat{\mathscr{D}}\boldsymbol{y}$.
- **(1.2)** If $\hat{K}$ is non-singular for any given $\boldsymbol{y}$, unique Eigenvalues can be obtained as $\bar{\boldsymbol{\alpha}} = \hat{K}^{-1}\boldsymbol{y}$ (Eq. (4)). 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, δ = 0.1, as suggested in (Cai et al., 2007).

4.5.2. Optimizing weights (β) for MKL—To obtain $\bar{\boldsymbol{\alpha}}$, optimal set of weights $\hat{\boldsymbol{\beta}} = [\hat{\beta}_1, \ldots, \hat{\beta}_M]$, $\hat{\beta}_m \in [0,1]$, have to be obtained for each modality m, $m \in \{1,\ldots M\}$ such that $\sum_{m=1}^{M}\widehat{\beta}_m=1$. A hierarchical brute force optimization strategy is employed to optimize weights, $\boldsymbol{\beta} = [\beta_1,\ldots, \beta_M]$, which iteratively optimizes $\boldsymbol{\beta}$ based on the classification accuracy of training data. Once initial values for $\hat{\boldsymbol{\beta}}$ $\forall m$, $m \in \{1,\ldots, M\}$, 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.

4.6. SeSMiK-GE algorithm

Algorithm. SeSMiK-GE

Input: $\mathscr{F}_m$, M, $\mathscr{N}$ d, $\hat{\boldsymbol{\beta}} = [\hat{\beta}_1, \hat{\beta}_2,\ldots, \hat{\beta}_M]$, $\omega_l \in [0,1]$

Output: $\bar{\alpha}$

begin

0. *for* $m = 1 : M$

1. Obtain K_m for each data channel $\mathcal{F}_m$

2. Use ω_l to obtain $\tilde{W}_m$ using $\mathcal{N}$ from Eq. (8)

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 Eq. (6)

7. Obtain d-dimensional $\bar{\alpha}$ by solving Eq. (10)

end

K_m and $\tilde{W}_m$ (via ω_l, where ω_l are the exposed labels while computing $\tilde{W}_m$) are obtained for each of the data channels $\mathcal{F}_m$, $m \in \{1,\ldots,M\}$ 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 $\bar{\alpha}$ for subsequent object classification. Note that when v new samples are introduced in the original data matrix $\mathcal{F}$ of size n, the kernel matrix K_m, and the weight matrix $\tilde{W}_m$ (with original ω_l exposed) are recomputed to obtain matrices of size $n + v \times n + v$, and the classifiers are retrained using the recomputed low dimensional representation $\bar{\alpha}$ for samples $n + v$.

5. Experimental design

5.1. Data description and preliminaries

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. MRI was performed by using a 1.5-T whole-body MRI 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. Data sets were acquired as 16 × 8 × 8 phase-encoded spectral arrays (1024 voxels) by using a nominal spectral resolution of 0.24−0.34 cm^3, 1000/130, and a 17-min acquisition time. Three-dimensional, MR spectroscopic imaging data were processed and aligned with the corresponding T2w imaging data using a combination of in-house software and Interactive Display Language (Research Systems, Boulder, Colorado) software tools (Tran et al., 2000). 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 non-linear 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.

The 3D prostate T2w MRI scene is represented by $\hat{\mathscr{C}} = (\hat{C}, \hat{f})$, 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 $\hat{c} \in \hat{C}$. We also define a MR spectral scene $\mathscr{C} = (C, \mathbf{F})$ where C is a 3D grid of MRS metavoxels, $c \in C$, and $\mathbf{F}$ is a spectral vector associated with each $c \in C$. 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 $\mathbf{h}(c)$. The corresponding 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 paper are listed in Table 1, and the major acronyms are listed in Table 2.

5.2. 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 (Fig. 3). An expert pathologist manually delineated CaP and grade extent on whole mount histological sections. The CaP and grade extent on the MRI were further delineated and graded by an expert prostate imaging expert (JK) (more than 25 year experience in the field of MP MRI for CaP detection and grading) by visually registering corresponding histological (Fig. 3a) and radiological sections (Fig. 3b); 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. Based on the CaP extent on visually registered histology, hypo intense appearance on T2w MRI, and reduced choline concentrations on MRS, MR spectral voxels were annotated as CaP, benign, or excluded (as described in Section 5.2.1). Similarly, correspondences across pathological scores from visually registered histology, hypo intense appearance on T2w MRI, and reduced choline concentrations on MRS, were used to further label the CaP MR spectral voxels as low- or high-grade (as described in Section 5.2.2). The class labels thus obtained for each of the individual MR spectral voxels were used as the surrogate ground truth for CaP detection and grading on the MRI/MRS, and were employed for subsequent training and evaluation of the SeSMiK-GE classifier.

5.2.1. Annotations for CaP vs. Benign—Since digitized images of the radical prostatectomy histologic sections were not available for this study (Fig. 3a only represents one single illustrative example), deformable image registration methods for spatially aligning *ex vivo* histology and pre-operative MRI (Chappelow et al., 2011; Xiao et al., 2011) could not be employed to more precisely map the extent of CaP onto MRI.

The MRS voxels were annotated on a 5-point scale (Fig. 4) adapted by the standardized 5-point scale developed by Jung et al. (2004), where each spectrum is defined as being either (a) definitely benign (scale 1), (b) probably benign (scale 2), (c) equivocal (scale 3), (d) probably malignant (scale 4), and (e) definitely malignant (scale 5). While the scale

described by Jung et al. (2004) was based on metabolic ratio of MR spectra alone, the spectral annotations in this work were performed by incorporating following information: (1) expert pathologists delineated CaP annotations on step-section histology, (2) presence and strength of hypointensities on T2w MRI within the corresponding CaP region after visual alignment with histology, (3) changes in metabolic concentrations on MR spectrum that were identified within the tumor boundary on visually registered histology, and (4) visual similarity of each MR spectrum to representative scale as described in Jung et al. (2004). MRS voxels annotated as 'likely benign' or 'probably benign' were labeled as 'benign', while the voxels annotated as 'probably malignant' or 'likely malignant' were labeled as CaP. MRS voxels annotated as 'equivocal' or seen to be atrophic on histology were excluded from further analysis.

5.2.2. Annotations for high vs. low Gleason score—Prostate lesions were further assigned a pathologic score based on the step section histopathology findings. Based on the correspondence of tumor annotations on step section histopathology with T2w MRI and MRS, the prostate-imaging expert binned the pathologic scores into 5 groups and annotated each spectrum as belonging to one of these classes: (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). The voxels labeled (1, 2) were expert-annotated as low grade and voxels labeled (3, 4, 5) were expert-annotated as high grade, and were subsequently used for training and evaluation of machine learning classifiers.

5.3. Feature extraction from MRI and MRS

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

5.3.2. Feature extraction from T2w MRI—T2w MR images were first corrected for bias field (due to the inhomogeneous endorectal coil reception) using a post-processing correction strategy as given in (Noworolski et al., 2010). 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 (Madabhushi et al., 2005). Fig. 5 shows an example of a texture feature (Fig. 5b) extracted from the original T2w MR image (Fig. 5a) that captures some of the subtle differences between low and high Gleason grade CaP regions, both of which appear as hypointense on the original T2w MR image. A brief summary of T2w MRI features extracted in this work is provided in Table 3.

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 Φ_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 (gray 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 $g_u(c) = \frac{1}{|R_{cd}|} \sum_{\hat{c} \in R_{cd}} \left[\widehat{f_u(c)} \right]$. 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 $\mathcal{G} = [\boldsymbol{G}_1; \boldsymbol{G}_2; \ldots; \boldsymbol{G}_n] \in \mathbb{R}^{n \times 34}$. 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.

5.4. Hierarchical classification of high grade CaP via MP SeSMiK-GE signature

5.4.1. Random Forest (RF) classifier—The RF classifier uses a majority voting rule for class assignment by combining decisions from an ensemble of bagged (bootstrapped aggregated) (Breiman, 2001) decision trees. The C4.5 decision tree (Quinlan, 1993) is a multi-stage 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 (Breiman, 2001). RF further combines the weak decisions obtained from multiple decision trees trained on bootstrapped training subsets to provide a more optimal and stable solution. Advantages of RF over other classifier schemes (such as the Support Vector machine (SVM) classifier) include, (1) ability to seamlessly integrate a large number of input variables, (2) robustness to noise in the data, and (3) relatively few tuning-parameters. 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 (Quinlan, 1993) classifier $\mathbf{h}_j$, $j \in \{1,\ldots, N\}$ is constructed. The class label $\mathbf{h}_j(c)$ for each meta-voxel $c \in C$, based on the feature vector $\bar{\alpha}$, is then obtained using the decision trees $\mathbf{h}_j$, $j \in \{1,2,\ldots,N\}$; $\mathbf{h}_j(c) = 1$ if c is classified as CaP and $\mathbf{h}_j(c) = 0$ otherwise. The final class likelihood that c belongs to CaP, via the RF classifier, is obtained by aggregating the decisions of individual weak learners as $\frac{1}{N}\sum_{j=1}^{N}\mathbf{h}_j(c)$.

5.4.2. 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 (Liu et al., 2009). 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 (Monaco et al., 2010). 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 $\bar{\alpha}$.

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

Algorithm. *HierarchHighGradeCaP*

Input: $\mathcal{F}$, $\mathcal{G}$ $\mathcal{N}$ d, $\hat{\boldsymbol{\beta}} = [\hat{\beta}_1, \hat{\beta}_2]$, ω_l

Output: $\tilde{\mathbf{h}}(c)$, $\hat{\mathbf{h}}(c)$

begin

0. Obtain $\boldsymbol{F}^{T2MRS} = SeSMiK\text{-}GE([\mathcal{F}, \mathcal{GN}d, \hat{\boldsymbol{\beta}}, \omega_l])$

Training phase

1 Obtain classifier training data : $\mathcal{F}_{tr}^{T2MRS}$

2 Identify subset of voxels with no labels exposed, $U = (B - l)$, where $B = \left| \mathcal{F}_{tr}^{T2MRS} \right|$

3 $\rho = 0$

4 while $\rho \leq 1$

5 Compute $\mathbf{h}(c)$ for each $c \in U$ using RF

6 Apply PPMM to obtain $\tilde{\mathbf{h}}(c)$ for each $c \in U$

7 obtain $\gamma_{(\rho)}^{tr} = [SN_{(\rho)}^{tr} + SP_{(\rho)}^{tr}]$

8 $\rho = \rho + 0.1$

9 endwhile

10 Obtain $\nu = \arg\max_{\rho}[\gamma_{(\rho)}^{tr}]$

Test phase

1 Obtain classifier test data $\mathcal{F}_{test}^{T2MRS}$

2 Compute $\mathbf{h}(c)$ for each $c \in \mathcal{F}_{test}^{T2MRS}$ using RF

3 Apply PPMM to obtain $\tilde{\mathbf{h}}(c)$ for each $c \in \mathcal{F}_{test}^{T2MRS}$

4 $A = \{c | \tilde{\mathbf{h}}(c) \geq \nu\}$, $A \subset \mathcal{F}_{test}^{T2MRS}$, $\nu \in [0, 1]$

5 $\forall c \in A$, compute $\hat{\mathbf{h}}(c)$ using RF

6 **return** $\tilde{\mathbf{h}}(c)$, $\hat{\mathbf{h}}(c)$

end

In Step 0, fused MP T2w MRI/MRS signature $\boldsymbol{F}^{T2MRS}$ for every $c \in C$ is obtained by calling the *SeSMiK-GE* algorithm. A total of l samples that are exposed for obtaining the low dimensional semi-supervised representation $\boldsymbol{F}^{T2MRS}$, are also used to obtain the optimal threshold ν that maximizes sensitivity and specificity of classifying high versus low grade CaP. In Step 1 of the training phase, training samples $\mathcal{F}_{tr}^{T2MRS} = [\boldsymbol{F}_{tr}^{T2MRS}(c_1), \boldsymbol{F}_{tr}^{T2MRS}(c_2), \ldots, \boldsymbol{F}_{tr}^{T2MRS}(c_B)]$ are identified of which l randomly selected samples have their labels exposed. The l samples are used for training the classifier and the remaining U, $U = (B - l)$ samples are used to obtain the optimal threshold parameter for separating high versus low grade CaP on training data. In Steps 4–7, classifier detection sensitivity $(SN_{(\rho)}^{tr})$ specificity $(SP_{(\rho)}^{tr})$, and the sum $\gamma_{(\rho)} = [SN_{(\rho)}^{tr} + SP_{(\rho)}^{tr}]$ are calculated at every threshold ρ, $\rho \in [0,1]$ for the training data. The thresholds are increase in intervals of 0.1 to obtain 11 $SN_{(\rho)}^{tr}$, $SP_{(\rho)}^{tr}$ and $\boldsymbol{\gamma}$ values. In Step 10, ν is obtained as the threshold that maximizes $\boldsymbol{\gamma}$ and is subsequently used in the test phase.

In Step 1 of the test phase, the set of voxels corresponding to 3D volumetric images, $\mathcal{F}_{\text{test}}^{T2MRS} = [F_{\text{test}}^{T2MRS}(c_1), F_{\text{test}}^{T2MRS}(c_2), \ldots, F_{\text{test}}^{T2MRS}(c_P)]$ from a study (kept independent from the set of training voxels) are used to obtain the probability of high versus low grade CaP using the classifier model trained in the training phase. $\mathbf{h}(c)$ is obtained for every

$c \in \mathscr{F}_{\text{test}}^{T2MRS}$ using a RF classifier in Step 2, followed by applying PPMM on $\mathbf{h}(c)$ to obtain $\tilde{\mathbf{h}}(c)$. In Step 4, the threshold ν obtained in Step 10 of the training phase is used to identify a subset of voxels, $A \subset \mathscr{F}_{\text{test}}^{T2MRS}$, that have a CaP probability greater than ν. All the voxels in A are then further distinguished, in Step 5, as belonging to either high or low grade CaP using a RF classifier.

5.5. Implementation details and classifier training

5.5.1. Parameter selection for SeSMiK-GE—For the SSL module, 40% of the total training samples were randomly selected to obtain $\tilde{W}_{MRS}$ and $\tilde{W}_{MRI}$, and train the classifier to obtain ν. 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 4.5), and subsequently used to obtain F^{T2MRS}. The interval resolution for termination of β was set to 0.001.

5.5.2. Classifier training—Two independent CV strategies, leave-one-out (LOO) and threefold CV, were used for evaluation of CaP classifiers. LOO is known to be an unbiased predictor and is capable of creating sufficient training data for studies with a small sample size. Given the relatively small sample size for datasets available for high versus low-grade classification, LOO was our first method of choice for CV (Efron, 1983). In LOO, each classifier was trained on 28 CaP studies, while one study was used for testing (Kelm et al., 2007). This process was repeated until all 29 studies were classified once within a single run of LOO CV on a per voxel basis. LOO however is known to have a high variance (Efron, 1983). A threefold CV strategy was hence additionally employed for an unbiased evaluation of the efficacy of SeSMiK-GE across the two CV strategies. For three-fold-CV, 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. Note that the RF classifier is reconstructed with different training and test sets at every CV run (training and test sets are maintained independent of each other) for both LOO and threefold CV.

This process was repeated until all 29 studies were classified once within a single run of CV. The threefold randomized CV process was then repeated 25 times for different training and test sets.

A total of 12 out of the 29 CaP patient studies were found to have high grade CaP. Both LOO, and threefold CV 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 threefold-CV, 8 studies were used for training and 4 for testing, until all 12 patient studies are classified within a single CV run. Threefold-CV was similarly repeated 25 times over different training and test sets for classifying patients with high Gleason grade CaP on a per-voxel basis.

5.5.3. Performance evaluation—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 LOO and threefold CV strategies. Each point on the curve corresponds to the CaP detection sensitivity ($SN_{(\rho)}$) and specificity ($SP_{(\rho)}$) 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%. To generate a ROC curve across multiple CV runs, we first interpolate the values of SN and SP for individual

ROC curves (per CV run) at equal intervals between 0 and 1 using a cubic-spline polynomial, to obtain an average ROC curve.

5.6. 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 (Tiwari et al., 2009), T2w MRI (Madabhushi et al., 2005) 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 (Rohlfing et al., 2005). A summary of all the comparative strategies used for CaP detection and grading in this work is given in Table 4.

5.6.1. 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 (Tiwari et al., 2009) 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 $\hat{\mathbf{h}}^{MRS}$.

5.6.2. 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 $c \in C$.

5.6.3. Data combination (COD)—A combined feature vector $F^{Int}(c) = [F^{MRS}(c), F^{T2}(c)]$ is obtained by concatenating MRS and T2w MRI reduced Eigenfeatures for each metavoxel $c \in C$. A RF classifier is then trained using $F^{Int}(c)$ 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 $c \in C$.

5.7. Experimental evaluation

5.7.1. 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 identifying CaP regions defined via the "surrogate ground truth" determination of which was described in Section 5.2.1. This was done via (a) area (AUC) under the ROC (Zweig and Campbell, 1993) curve (φ^{AUC}), and (b) classification accuracy (φ^{Acc}) at the operating point on the ROC curve.

5.7.2. Experiment 2. High grade CaP detection via SeSMiK-GE—ROC analysis across both LOO and threefold CV was independently performed across a total of 12 high grade CaP studies 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}. Since the remaining 17 studies (out of the total of 29 CaP studies) were determined to only have low grade CaP, quantitative evaluation (in terms of true positives, false positives) was not feasible, and hence was not reported.

6. Results and discussion

6.1. Experiment 1: CaP detection via SeSMiK-GE

ROC curves were generated by thresholding CaP probabilities at different thresholds for every cross-validation run (29 CV runs for LOO and 25 runs accumulated over threefold CV runs). AUC and accuracy values were reported at the operating point of each ROC curve for both LOO and threefold CV. Fig. 6a and c show AUC results (φ^{AUC}), while Fig. 6b and d show accuracy results (φ^{Acc}) obtained across 29 iterations of LOO CV, and 25 iterations of threefold CV respectively for various feature extraction and classifier strategies ($\tilde{\mathbf{h}}^{T2}$,$\tilde{\mathbf{h}}^{MRS}$,$\tilde{\mathbf{h}}^{Int}$,$\tilde{\mathbf{h}}^{IntD}$,$\tilde{\mathbf{h}}^{T2MRS}$) at the operating point using box-and-whiskers plots. The mean (φ_{μ}^{AUC}) and standard deviation (ζ^{AUC}) of AUC values, and classifier accuracy (φ_{μ}^{Acc}) across 29 iterations for LOO, and 25 iterations of threefold CV were also recorded and are shown in Table 5.

The results across the two CV strategies reflect that although the LOO CV strategy yielded higher φ^{AUC} and φ^{Acc} values, variance across φ^{AUC} and φ^{Acc} was significantly reduced with a threefold CV strategy. It is worth noting however that φ^{AUC} and φ^{Acc} across both CV strategies were found to be very similar, and superior to using individual modalities and other comparative strategies ($\tilde{\mathbf{h}}^{T2}$,$\tilde{\mathbf{h}}^{MRS}$,$\tilde{\mathbf{h}}^{Int}$, and $\tilde{\mathbf{h}}^{IntD}$).

Both Table 5 and Fig. 6 suggest 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 (Chan et al., 2003, Vos et al., 2010, Liu et al., 2009) and a COI based data integration classifier.

As suggested in (Demsar, 2006), to overcome the issue of multiple comparisons and biases with CV and parametric paired t-test, we incorporated a non-parametric Wilcoxon signed ranked test while comparing different classifier strategies across the two CV schemes. To further make the statistical significance test more stringent, the p-value was appropriately adjusted at $p = 0.00125$ via a Bonferroni test to account for Type 1 errors due to multiple testing. Table 6 shows the p values of statistical significance obtained using a non-parametric Wilcoxon signed test (Demsar, 2006) 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 threefold-CV strategies at $d = 15$. 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 $\tilde{\mathbf{h}}^{T2MRS}$ ($\mathrm{varphi}_{\mu}^{Acc}$=0.72 ± 0.16) significantly improved CaP detection accuracy (~20% improvement for $d = 15$) over not imposing any spatial constraints $\mathbf{h}^{T2MRS}$(φ_{μ}^{Acc}=0.89 ± 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 threefold-CV (Table 5), and resonates with similar findings in (Liu et al., 2009).

6.2. Experiment 2: CaP grading via SeSMiK-GE

Fig. 7 shows qualitative results for the hierarchical classification strategy for identifying high grade CaP. Fig. 7a, d and g show CaP and high grade CaP ground truth as annotated by an expert. Yellow outline in Fig. 7a, d and 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 Fig. 7b, e and 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 Fig. 7b, e and h, a probability of high grade CaP is further assigned, the corresponding probability heat maps for which are shown in Fig. 7c, f and i. Here, the spatial locations shown in red (Fig. 7c, f and 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 sections.

Fig. 8a shows φ^{AUC} while Fig. 8b 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-CV strategy. Corresponding results for φ^{AUC} and φ^{Acc} obtained via a threefold CV strategy are shown in Fig. 8c and d respectively. The mean (φ_{μ}^{AUC}) and standard deviation (ζ^{AUC}) of AUC values at the operating point of the ROC curve for LOO and threefold CV are shown in Table 7. Table 8 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 threefold-cross-validation strategies at $d = 15$.

Fig. 9 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 (Fig. 9a) 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 (Fig. 9b) 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 ($\mathbf{h}^{T2MRS}$) (Fig. 9a).

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 (Tiwari et al., 2009). 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) 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 (Langer et al., 2010; Shukla-Dave et al., 2007; Shukla-Dave et al., 2009).

Fig. 10b shows the classification result obtained for a single 2D section from a 3D volume for a T2w MRI/MRS sections (Fig. 10a), the classifier results having been plotted back onto the corresponding MRS grid Fig. 10b. Fig. 10b 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 detection of high grade CaP. Note that the spectra identified as low 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.

7. Concluding remarks

In this work we presented a novel semi-supervised multi-kernel Graph Embedding (SeSMiK-GE) scheme, a generalized data fusion framework that could potentially be used for integrating any combination of imaging and non-imaging data channels, independent of

the dimensionality of the individual channels. Specifically, we applied SeSMiK-GE for integrating structural and metabolic information from MP MRI/MRS for (a) identifying high probability CaP regions, and (b) further classifying the detected CaP regions as high or low grade. To the best of our knowledge, this is the first application of a computerized decision support classifier for identifying high grade CaP using MP MRI. SeSMiK-GE uniquely combines 3 very new, exciting machine learning methodologies – (1) multi-kernel learning, (2) semi-supervised learning, and (3) dimensionality reduction within a single unified framework. 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. Most previous DSS approaches for CaP detection on MRI (Chan et al., 2003; Vos et al., 2010; Liu et al., 2009; Langer et al., 2009; Ozer et al., 2010) have only reported results for identifying CaP versus benign using MP MRI. Our classifier results for high grade CaP (AUC = 84%) via SeSMiK-GE were found to be superior in most cases to results reported in clinical studies (Kobus et al., 2011; Zakian et al., 2005; Shukla-Dave et al., 2007).

Although the results obtained via SeSMiK-GE significantly outperformed other classifier strategies for both CaP diagnosis, as well as grading, 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. These intermediate grade spectra might provide some clinical insights about the biology of prostate cancer, an area we intend to explore in future work. (2) Obtaining annotations is known to be expensive, time consuming, and requiring highly specialized training; the prostate-imaging expert hence only annotated studies that were known to have high grade CaP, and the false positive rate for low grade CaP on the remaining studies could not be evaluated. (3) Ground truth labels for grade and extent of disease on MRI were determined manually by visually correlating the corresponding *ex vivo* histologic and pre-operative *in vivo* MRI sections and served as a “surrogate ground truth” in the absence of digitized histological samples. However, several high impact technical and clinical papers in the literature have previously employed a similar surrogate of ground truth for CaP extent (Chan et al., 2003; Vos et al., 2010; Liu et al., 2009; Langer et al., 2009; Ozer et al., 2010; Madabhushi et al., 2005; Tiwari et al., 2012; Westphalen et al., 2008; Costouros et al., 2008). (4) While only 29 studies were considered in this study, the total number of datasets is still more than similar classifier strategies for prostate cancer diagnosis from MRI (Chan et al., 2003; Liu et al., 2009; Langer et al., 2009; Ozer et al., 2010) (where the dataset size ranged from $N = 10–25$). (5) Due to the relatively small sample size available in this work, the training set does not change much while training the classifier during each run of CV and hence the independence assumption when computing AUC and accuracy values may not be valid. However, the training and test sets are kept completely independent of each other during each CV run to preserve as much degree of independence as possible during classification. (6) While protocols other than T2w MRI and MRS (for e.g. DWI, DCE), and vendor platforms other than GE were not considered in this study, SeSMiK GE was developed as a general purpose platform for enabling fusion of different kinds of data both imaging and non-imaging and across scanners, protocols. While an exhaustive validation of this platform across vendor platforms and protocols has not yet been performed, it is our intent to apply SeSMiK-GE in the context of different problems and for fusing different types of imaging parameters.

The methodology described in this work 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. With the recent US Preventive Task Force recommendation against PSA based screening for CaP,[2] a tool like SeSMiK-GE becomes even more critical for affording patients with low grade, indolent disease to opt for watchful waiting. With further larger scale clinical validation, we hope to be able to employ SeSMiK-GE in the context of a clinical trial to assist in identifying patients with high grade CaP against patients with no or low grade CaP.

Acknowledgments

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, R03CA143991-01), and Department of Defense (Predoctoral fellowship, W81XWH-09). The authors would also like to thank Dr. Mark Rosen, and Dr. Paul Wileyto for their advice and suggestions during the preparation of this manuscript.

References

Ampeliotis D, Antonakoudi A, Berberidis K, Psarakis EZ, Kounoudes A. A computer-aided system for the detection of prostate cancer based on magnetic resonance image analysis. Communications, Control and Signal Processing. 2008:1372–1377.

Bostwick DG. Grading prostate cancer. American Journal of Clinical Pathology. 1994; 102:S38–S56. [PubMed: 7524306]

Breiman L. Random forest. Machine Learning. 2001; 451:5–32.

Cai D, He X, Hu Y, Han J, Huang T. Learning a spatially smooth subspace for face recognition. Proc. IEEE Conf. Computer Vision and Pattern Recognition Machine Learning (CVPR'07). 2007:1–7.

Carroll P, Coakley F, Kurhanewicz J. Magnetic resonance imaging and spectroscopy of prostate cancer. Reviews in Urology. 2006; 8(1):S4–S10. [PubMed: 17021625]

Chan I, Wells W III, Mulkern RV, Haker S, Zhang J, Zou KH, Maier SE, Tempany CMC. Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel statistical classifier. Medical Physics. 2003; 30:2390–2398. [PubMed: 14528961]

Chappelow J, Bloch N, Rofsky N, Genega E, Lenkinski R, DeWolf W, Madabhushi A. Elastic registration of multimodal prostate MRI and histology via multiattribute combined mutual information. Medical Physics. 2011; 38(4):2005–2018. [PubMed: 21626933]

Damoulas T, Girolami MA. Probabilistic multi-class multi-kernel learning: on protein fold recognition and remote homology detection. Bioinformatics. 2008; 24(10):1264–1270. [PubMed: 18378524]

Demsar J. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research. 2006; 7:1–30.

Duda, R.; Hart, P.; Stork, D. Pattern Classification. second ed.. New York: Wiley-Interscience; 2000. p. 680

Epstein JI. An update of the Gleason grading system. Journal of Eurology. 2006; 13(1):57–59.

Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. Journal of American Statistical Association. 1983; 78:316–330.

Gleason D. Classification of prostatic carcinomas. Cancer Chemotherapy Reports. 1966; 50(3):125–128.

Heerschap A, Jager GJ, Van der Graaf M, Barentsz JO, Rosetten J, De La Oosterhof G, Ruijter E, Ruijs J. In vivo proton MR spectroscopy reveals altered metabolite content in malignant prostate tissue. Anticancer Research. 1997; 17:1455–1460. [PubMed: 9179183]

Jung JA, Coakley FV, Vigneron DB, Swanson MG, Qayyum A, Weinberg V, Jones KD, Carroll PR, Kurhanewicz J. Prostate depiction at endorectal MR spectroscopic imaging: investigation of a standardized evaluation system. Radiology. 2004; 233:701–708. [PubMed: 15564406]

[2] http://www.uspreventiveservicestaskforce.org/prostatecancerscreening.htm.

Kelm BM, Bjoern HM, Christian MZ, Klaus TB, Fred AH. Automated estimation of tumor probability in prostate magnetic resonance spectroscopic imaging: pattern recognition vs. quantification. Magnetic Resonance in Medicine. 2007; 57:150–159. [PubMed: 17191229]

Klein EA. What is insignificant prostate carcinoma. Cancer. 2004; 101:1923–1925. [PubMed: 15372479]

Klotz L. Active surveillance for prostate cancer: for whom? Journal of Clinical Oncology. 2005; 23(32):8165–8169. [PubMed: 16278468]

Kobus T, Hambrock T, Hulsbergen-van de Kaa C, Wright A, Barentsz J, Heerschap A, Scheenen T. In vivo assessment of prostate cancer aggressiveness using magnetic resonance spectroscopic imaging at 3 T with an endorectal coil. European Radiology. 2011; 60:1074–1080.

Kurhanewicz J, Vigneron J, Hricak H, Narayan P, Carroll P, Nelson S. Three dimensional H-1 MR Spectroscopic Imaging of the in situ human prostate with high (0.24−0.7-cm^3) spatial resolution. Radiology. 1996; 198:795–805. [PubMed: 8628874]

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. 2002; 16:451–463. [PubMed: 12353259]

Lanckriet GRG, Deng M, Cristianini N, Jordan MI, Noble WS. Kernel-based data fusion and its application to protein function prediction in yeast. Pacific Symposium on Biocomputing. 2004; 9:300–311. [PubMed: 14992512]

Langer D, Kwast T, Evans A, Trachtenberg J, Wilson B, Haider M. 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. 2009; 30(2):327–334. [PubMed: 19629981]

Langer DL, van der Kwast TH, Evans AJ, Plotkin A, Trachtenberg J, Wilson BC, Haider MA. Prostate tissue composition and MR measurements: investigating the relationships between ADC, T2, Ktrans, ve, and corresponding histologic features[1]. Radiology. 2010; 255:485–494. [PubMed: 20413761]

Lee G, Doyle S, Monaco J, Madabhushi A, Feldman MD, Master SR, Tomaszewski JE. 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. Proc. ISBI. 2009:77–80.

Lin T, Zha H. Riemannian manifold learning. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2008; 30(5)

Liu X, Langer DL, Haider MA, Yang Y, Wernick MN, Yetik IS. Prostate cancer segmentation with simultaneous estimation of Markov random field parameters and class. IEEE Transactions on Medical Imaging. 2009; 28(6):906–915. [PubMed: 19164079]

Madabhushi A, Feldman MD, Metaxas DN, Tomaszeweski J, Chute D. Automated detection of prostatic adenocarcinoma from high-resolution ex vivo MRI. IEEE Transactions on Medical Imaging. 2005; 24:1611–1625. [PubMed: 16350920]

May F, Treumann T, Dettmar P, Hartung R, Breul J. Limited value of endorectal magnetic resonance imaging and transrectal ultrasonography in the staging of clinically localized prostate cancer. BJU International. 2001; 87(1):66–69. [PubMed: 11121995]

McLean M, Srigley J, Banerjee D, Warde P, Hao Y. Interobserver variation in prostate cancer Gleason scoring: are there implications for the design of clinical trials and treatment strategies? Clinical Oncology. 1997; 9(4):222–225. [PubMed: 9315395]

Monaco J, Tomaszewski JE, Feldman MD, Hagemann I, Moradi M, Mousavi P, Boag A, Davidson C, Abolmaesumi P, Madabhushi A. High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models. Medical Image Analysis. 2010; 14(4):617–629. [PubMed: 20493759]

Muller KR, Mika S, Ratsch G, Tsuda K, Scholkopf B. An introduction to kernel-based learning algorithms. IEEE Transactions on Neural Networks. 2001; 12(2):181–201. [PubMed: 18244377]

Noworolski SM, Reed GD, Kurhanewicz J, Vigneron DB. Post-processing correction of the endorectal coil reception effects in MR spectroscopic imaging of the prostate. Journal of Magnetic Resonance in Imaging. 2010; 32(3):654–662.

Ozer S, Langer D, Liu X, Haider M, Kwast T, Evans A, Yang Y, Wernick M, Yetik I. Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI. Medical Physics. 2010; 37:1873–1883. [PubMed: 20443509]

Potosky A, Davis W, Hoffman R, Stanford J, Stephenson R, Penson D, Harlan L. Five-year outcomes after prostatectomy or radiotherapy for prostate cancer: the prostate cancer outcomes study. Journal of the National Cancer Institute. 2004; 96(18):1358–1367. [PubMed: 15367568]

Quinlan, J. C4.5: Programs for Machine Learning. San Francisco, CA: Morgan Kaufmann Publishers Inc; 1993.

Rohlfing T, Pfefferbaum A, Sullivan EV, Maurer CR. Information fusion in biomedical image analysis: combination of data vs. combination of interpretations. Information Processing in Medical Imaging. 2005; 3565:150–161. [PubMed: 17354692]

Roweis S, Saul L. Nonlinear dimensionality reduction by locally linear embedding. Science. 2000; 290:2323–2326. [PubMed: 11125150]

Scholkopf, B.; Smola, A. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press; 2001.

Schlkopf B, Herbrich R, Smola A. A generalized representer theorem. Computational Learning Theory. 2001; 2111:416–426.

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

Shi Y, Tillmann F, Anneleen D, Leon-Charles T, Johan S, Bart DM, Yves M. L2-norm multiple kernel learning and its application to biomedical data fusion. BMC Bioinformatics. 2010; 11(1):309–333. [PubMed: 20529363]

Shukla-Dave A, Hricak H, Kattan MW, Pucar D, Kuroiwa K, Chen HN, Spector J, Koutcher JA, Zakian KL, Scardino PT. The utility of magnetic resonance imaging and spectroscopy for predicting insignificant prostate cancer: an initial analysis. BJU International. 2007; 99:786–793. [PubMed: 17223922]

Shukla-Dave A, Hricak H, Ishill NM, Moskowitz CS, Drobnjak M, Reuter VE, Zakian KL, Scardino PT, Cordon-Cardo C. Correlation of MR imaging and MR spectroscopic imaging findings with Ki-67, Phospho-Akt, and androgen receptor expression in prostate cancer. Radiology. 2009; 250:803–812. [PubMed: 19244047]

Sugiyama M, Idé T, Nakajima S, Sese J. Semi-supervised local Fisher discriminant analysis for dimensionality reduction. Advances in Knowledge Discovery and Data Mining. 2008:333–344.

Sun D, Zhang D. A new discriminant principal component analysis method with partial supervision. Neural Processing Letters. 2009; 30(2):103–113.

Tenenbaum JB, De Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. Science. 2000; 290:2323–2326. [PubMed: 11125150]

Tiwari P, Rosen M, Madabhushi A. A hierarchical spectral clustering and nonlinear dimensionality reduction scheme for detection of prostate cancer from magnetic resonance spectroscopy (MRS). Medical Physics. 2009; 36(9):3927–3939. [PubMed: 19810465]

Tiwari P, Kurhanewicz J, Madabhushi A. Semi Supervised Multi Kernel (SeSMiK) graph embedding: identifying aggressive prostate cancer via magnetic resonance imaging and spectroscopy. MICCAI. 2010; 6363(3):666–673. [PubMed: 20879458]

Tiwari P, Viswanath S, Kurhanewicz J, Shridhar A, Madabhushi A. Multimodal wavelet embedding representation for data combination (MaWERiC): integrating magnetic resonance imaging and spectroscopy for prostate cancer detection. NMR in Biomedicine. 2012; 25:607–619. [PubMed: 21960175]

Viswanath S, Bloch N, Chappelow J, Toth R, Rofsky N, Genega E, Lenkinski R, Madabhushi A. Central gland and peripheral zone prostate tumors have significantly different quantitative imaging signatures on 3 Tesla endorectal, in vivo T2-weighted MR imagery. Journal of Magnetic Resonance Imaging. 2012; 36:213–224. [PubMed: 22337003]

Vos PC, Hambrock T, Barenstz J, Huisman H. Computer-assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI. Physics in Medicine and Biology. 2010; 55:1719. [PubMed: 20197602]

Wang L, Mazaheri Y, Zhang J, Ishill NM, Kuroiwa K, Hricak H. Assessment of biologic aggressiveness of prostate cancer: correlation of MR signal intensity with Gleason grade after radical prostatectomy. Radiology. 2008; 246:168–176. [PubMed: 18024440]

Tran TK, Vigneron DB, Sailasuta N, Tropp J, Roux P, Kurhanewicz J, Nelson S, Hurd R. Very selective suppression pulses for clinical MRSI studies of brain and prostate cancer. Magnetic Resonance in Medicine. 2000; 43(1):23–33. [PubMed: 10642728]

Welch H, Fisher E, Gottlieb D, Barry M. Detection of prostate cancer via biopsy in the Medicare-SEER population during the PSA era. Journal of the National Cancer Institute. 2007; 99(18):1395–1400. [PubMed: 17848671]

Wetter A, Engl T, Nadjmabadi D, Fliessbach K, Lehnert T, Gurung J, Beecken W, Vogl J. Combined MRI and MR spectroscopy of the prostate before radical prostatectomy. AJR, American Journal of Roentgenology. 2006; 187:724–730. [PubMed: 16928937]

Xiao G, Bloch N, Chappelow J, Genega E, Rofsky N, Lenkinski R, Tomaszewski J, Feldman M, Rosen M, Madabhushi A. Determining histology-MRI slice correspondences for defining MRI-based disease signatures of prostate cancer. Computerized Medical Imaging and Graphics. 2011; 35:568–578. [PubMed: 21255974]

Zakian KL, Eberhardt S, Hricak H, Shukla-Dave A, Kleinman S, Muruganandham M, Sircar K, Kattan MW, Reuter VE, Scardino PT, Koutcher JA. Transition zone prostate cancer: metabolic characteristics at H MR spectroscopic imaginginitial results. Radiology. 2003; 229:241–247. [PubMed: 12920178]

Zakian KL, Sircar K, Hricak H, Chen HN, Shukla-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 septsection pathologic analysis after radical prostatectomy. Radiology. 2005; 234:804–814. [PubMed: 15734935]

Zhang Z, Wang J, Zha H. Adaptive manifold learning. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2011 Pre-print.

Zhao H. Combining labeled and unlabeled data with graph embedding. Neurocomputing. 2006; 69(16–18):2385–2389.

Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry. 1993; 39(4):561–577. [PubMed: 8472349]

Westphalen AC, Coakley FV, Qayyum A, Swanson M, Simko JF, Lu Y, Zhao S, Carroll PR, Yeh BM, Kurhanewicz J. Peripheral zone prostate cancer: accuracy of different interpretative approaches with MR and MR spectroscopic imaging. Radiology. 2008; 246(1):177–184. [PubMed: 18024434]

Costouros NG, Coakley FV, Westphalen AC, Qayyum A, Yeh BM, Joe BM, Kurhanewicz J. Diagnosis of prostate cancer in patients with an elevated prostate-specific antigen level: role of endorectal MRI and MR spectroscopic imaging. American Journal of Roentgenology. 2008; 188:812–816. [PubMed: 17312072]

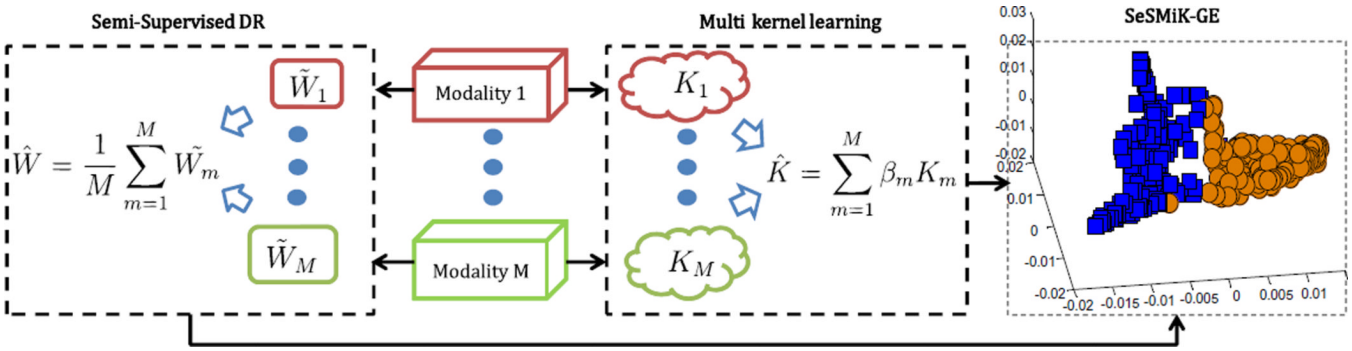

Fig. 1.

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). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

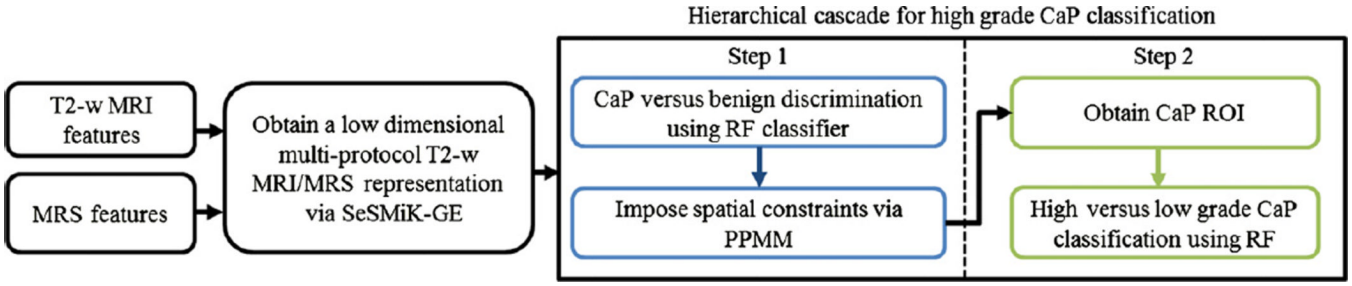

Fig. 2.

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.

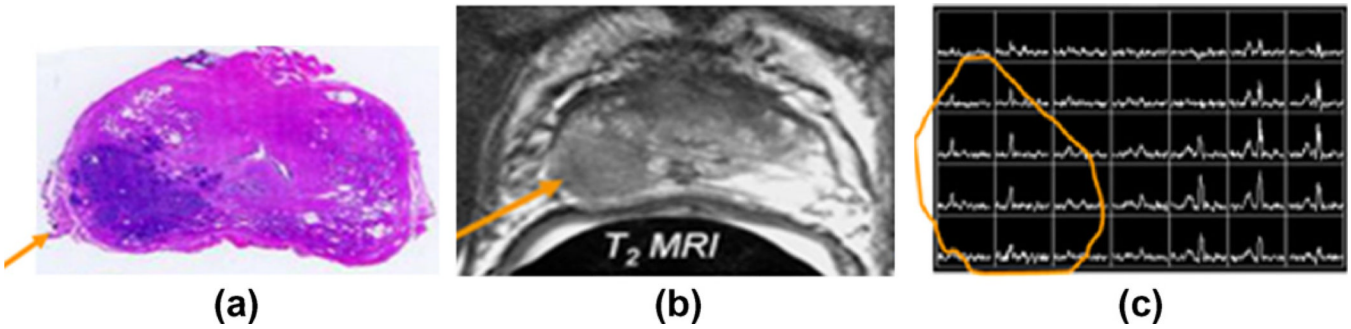

Fig. 3.
(a) Ground truth for CaP extent as defined through the histopathological analysis of hemotoxylin and eosin stained tissue section. The histological CaP extent in (a) is then visually registered onto the corresponding T2w MRI (b) and MRS sections (c) by an expert using histology as a visual reference.

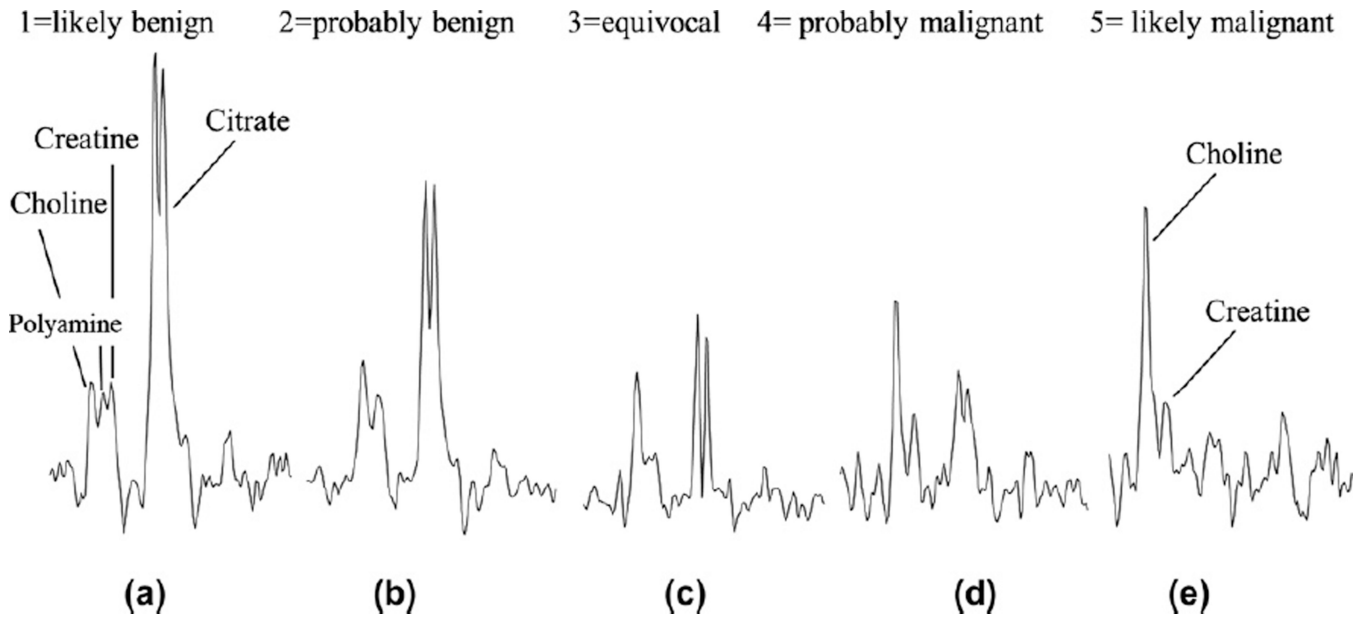

Fig. 4.

Illustration of the standardized five point scale spectra classified as (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) (Fig. 3 reproduced from Jung et al. (2004) with permission of the author).

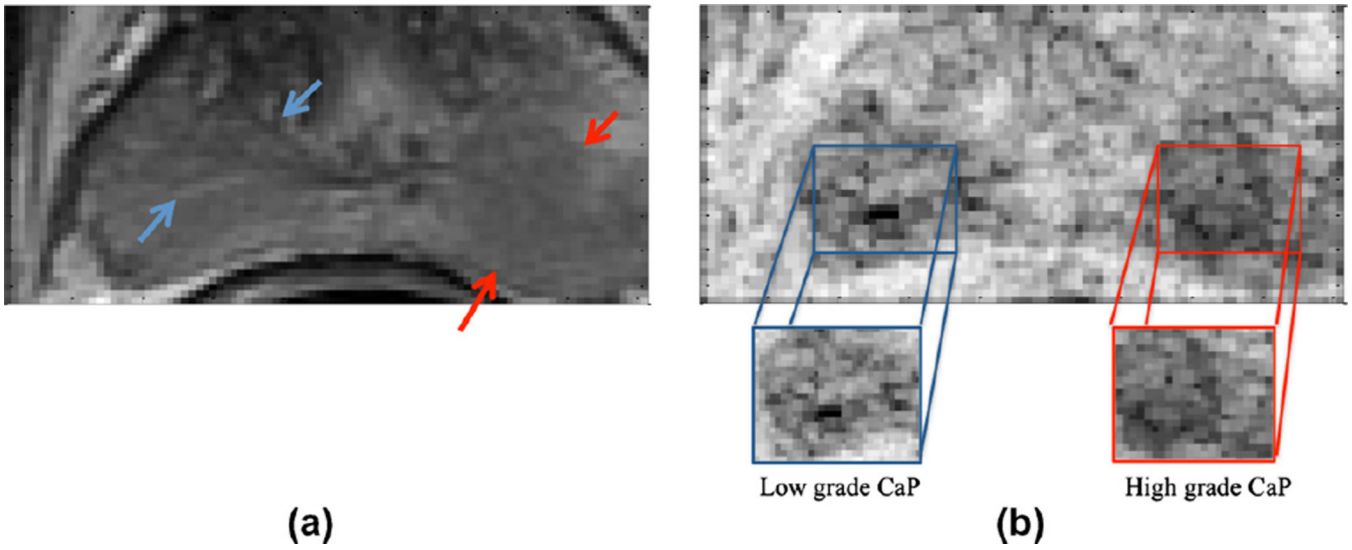

Fig. 5.
(a) Ground truth for high (red arrows) and low grade (blue arrows) CaP extent on a single T2w MRI section. (b) A grey-level texture feature for the corresponding section used to illustrate subtle, yet existing texture differences for low and high grade CaP regions on the same section. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

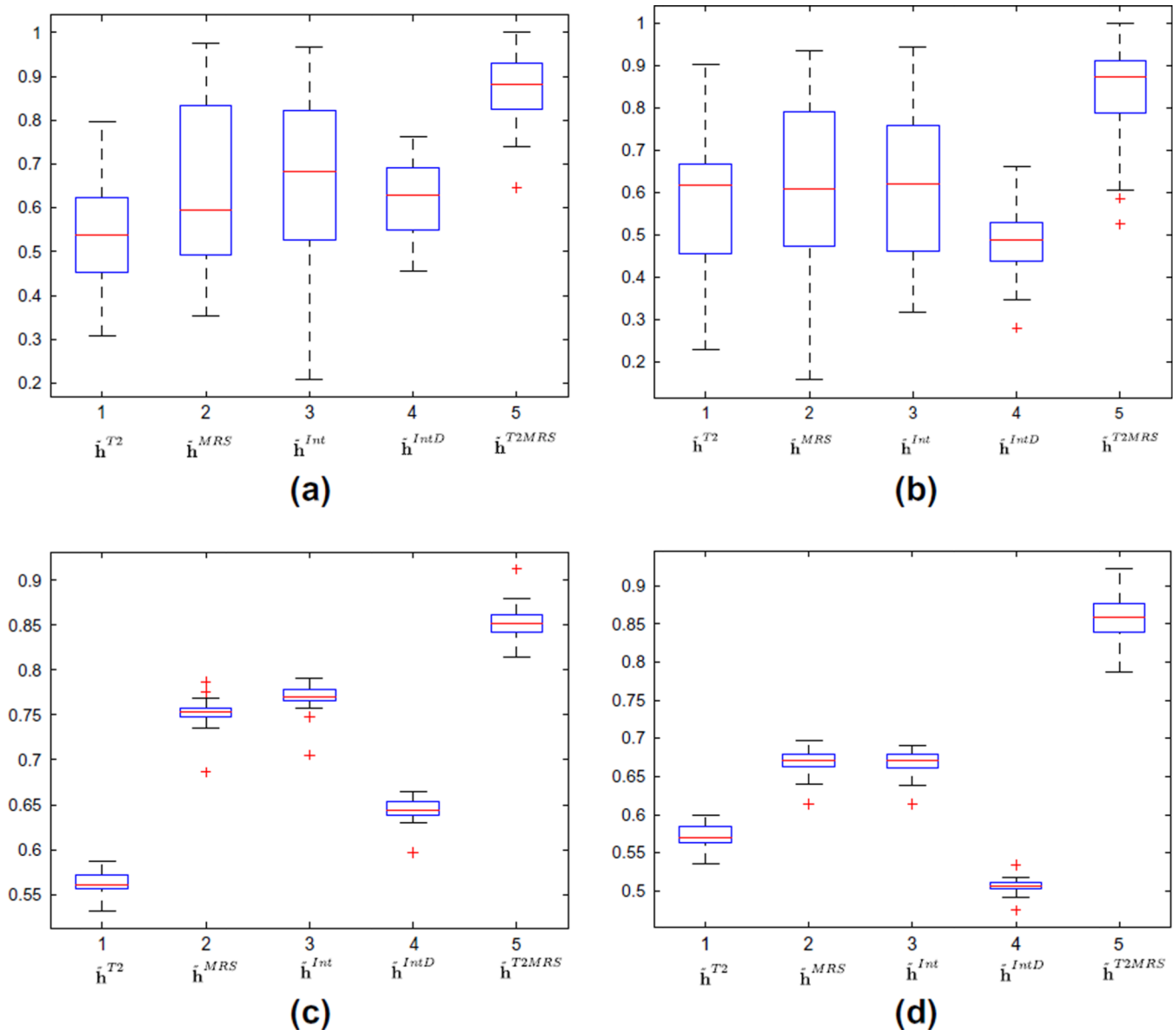

Fig. 6.

Box-and-whisker plot results of AUC (a) and accuracy (b) obtained over 29 studies via a LOO CV strategy for $\tilde{\mathbf{h}}^{T2}$, $\tilde{\mathbf{h}}^{MRS}$, $\tilde{\mathbf{h}}^{Int}$, $\tilde{\mathbf{h}}^{IntD}$, and $\tilde{\mathbf{h}}^{T2MRS}$. (c and d) The box-whisker-plots for threefold CV strategy over 25 CV 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 (a and c) and accuracy (b and 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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

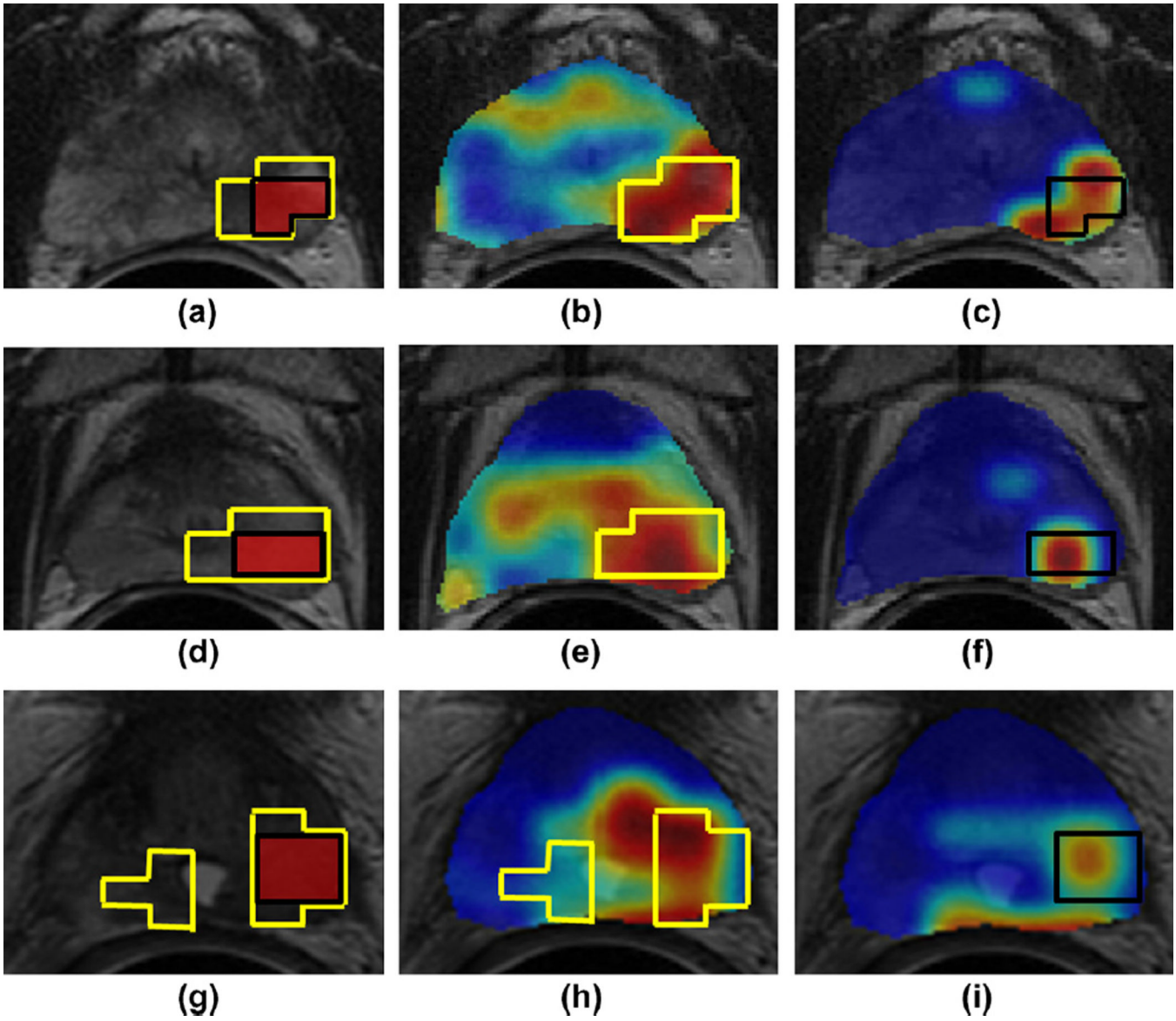

Fig. 7.

(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. (b), (e), and (h) show the probability heat map results corresponding to CaP classification on T2w MRI sections (by interpolating the CaP probabilities at MRS resolution to a pixel level T2w MRI resolution using Gaussian smoothing) in (a), (d), and (g) respectively for three different T2w MRI studies. (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, obtained by interpolating the high grade CaP probabilities at MRS resolution to a pixel level T2w MRI resolution using Gaussian smoothing) in (b), (e) and (h) respectively. Note that locations shown in red in (b), (e), and (h) correspond to those identified by $\mathbf{h}^{T2MRS}$ as CaP while in (c), (f), and (i) as those identified as high grade CaP by $\hat{\mathbf{h}}^{T2MRS}$. Similarly the spatial locations shown in blue in (b), (e), and (h) correspond to spatial locations classified as benign and as low grade CaP in (c),

(f), and (i). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

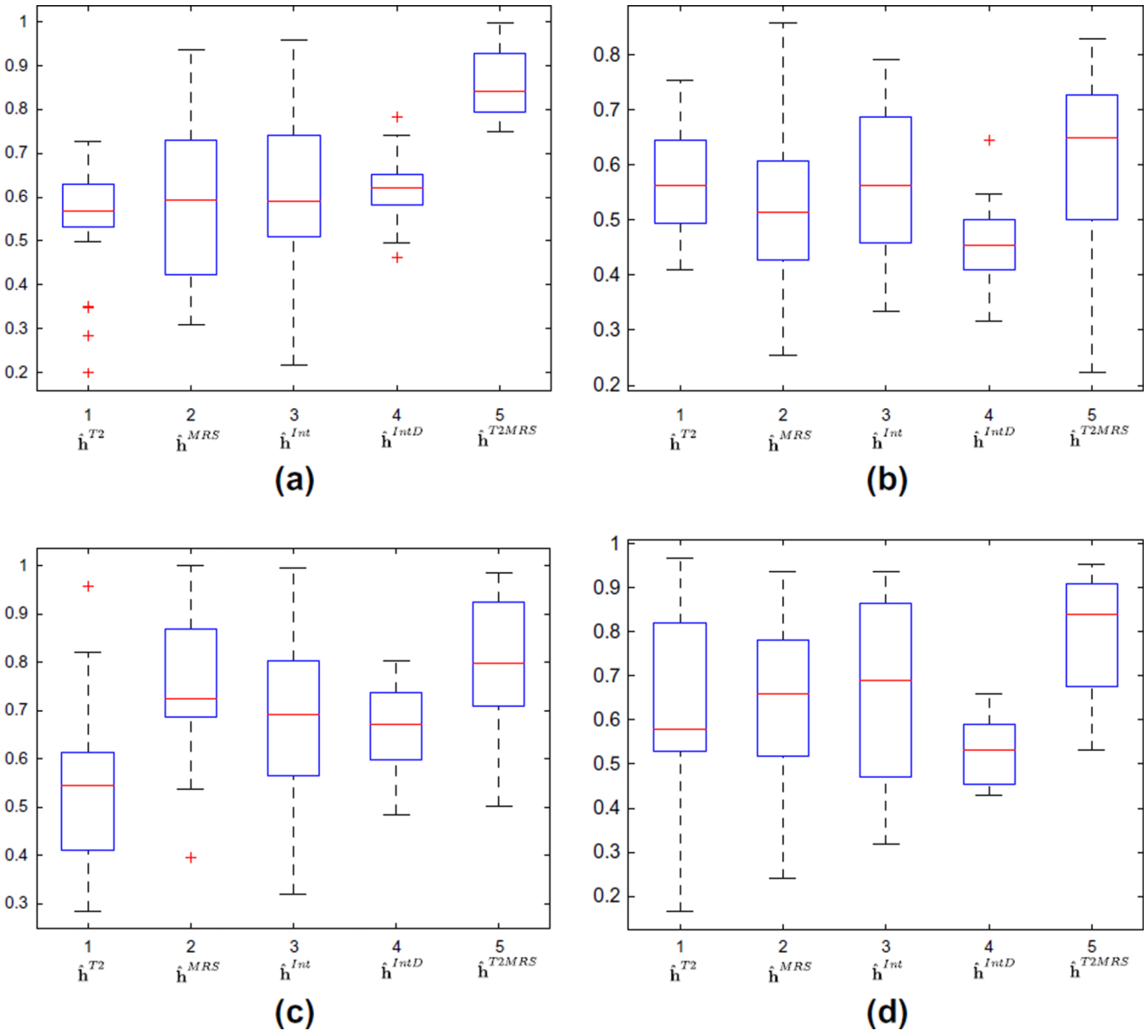

Fig. 8.

Box-and-whisker plot results of AUC (a) and accuracy (b) obtained over 12 studies via a LOO CV 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 threefold-CV are shown in (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 (a and c) and accuracy (b and 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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

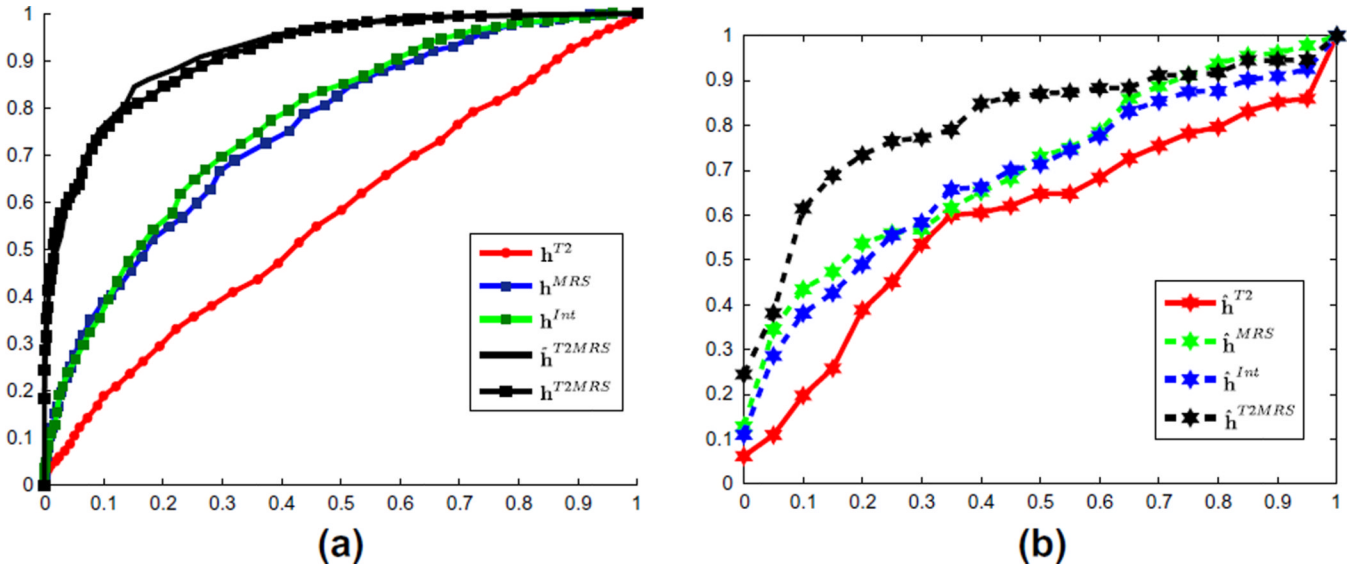

Fig. 9.

Average 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 vs. low grade CaP using $\hat{\mathbf{h}}^{T2}$, $\hat{\mathbf{h}}^{MRS}$, $\hat{\mathbf{h}}^{Int}$, and $\hat{\mathbf{h}}^{T2MRS}$ for d = 15.

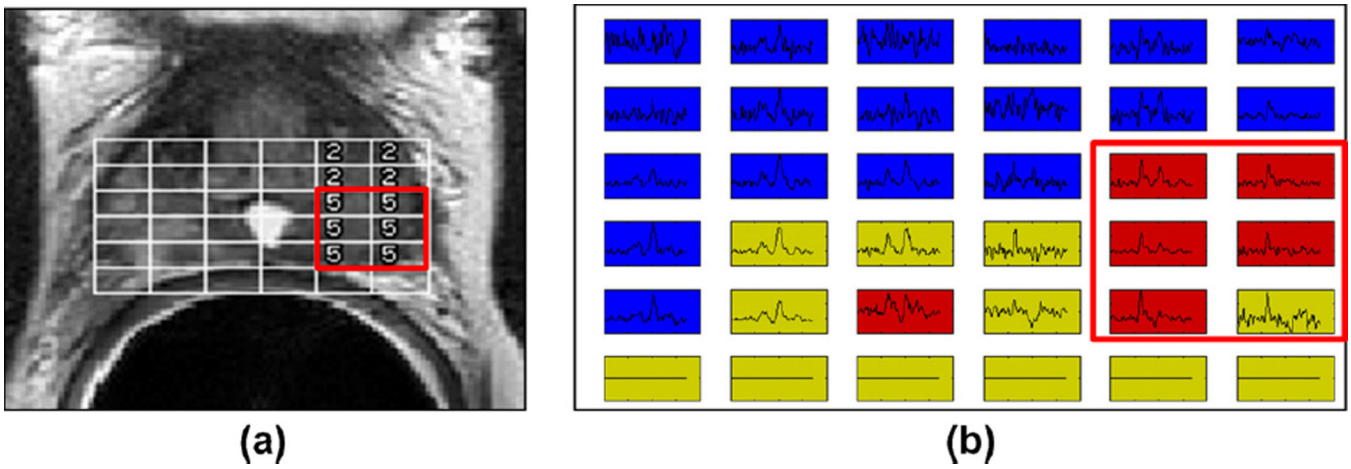

Fig. 10.

(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. (b) The corresponding classification result obtained by thresholding the probability values at the operating point v 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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1

List of commonly used notation and symbols in this paper.

Symbol	Description	Symbol	Description
$\hat{\mathscr{C}}$	3D MRI scene	$\mathscr{C}$	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, $c \in C$
$f_a(c)$	MRS signal intensity at c	a	Frequency index
$\hat{\mathscr{G}}$	T2w MRI feature scene	Φ_u	Feature operator, $u \in \{1,\ldots, 34\}$
$g_u(c)$	Mean feature value at metavoxel c	G	Undirected weighted graph
K	Kernel matrix	ϕ	Pairwise kernel mapping
$\mathscr{F}_m$	Original data matrix for protocol m	$F(c)$	Feature vector at every c
D	High dimensional feature space	d	Low dimensional feature space
$\mathscr{D}$	Diagonal weight matrix	L	Laplacian matrix, $L = \mathscr{D} - W$
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,$\omega \in [0,1]$
$\mathscr{N}$	Neighborhood parameter	$\tilde{W}_m$	Modified weight matrix for every 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	$\hat{\mathscr{D}}$	Combined diagonal matrix
$\hat{K}$	Combined kernel matrix	$\hat{W}$	Combined weight matrix
K_{MRI}	MRI kernel matrix	K_{MRS}	MRS kernel matrix
δ	Optimization parameter	σ	Scaling parameter
$\hat{\boldsymbol{\beta}}$	Subset of optimal weights $[\hat{\beta}_m]$	$\bar{\boldsymbol{\alpha}}$	Integrated representation
$\mathbf{F}^{MRS}$	MRS feature vector for each $c \in C$	$\mathbf{F}^{T2MRS}$	Reduced SeSMiK-GE feature vector
$\mathbf{G}$	Average T2w feature vector	$\mathbf{G}^{T2}$	Reduced 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	v	Threshold for hierarchical classifier
$\mathscr{F}_{tr}^{T2MRS}$	Set of 3D volumetric training images	$\mathscr{F}_{\text{test}}^{T2MRS}$	Set of 3D volumetric test images
B	Number of training samples	P	Number of test samples

Table 2

List of major acronyms used in the manuscript.

Acronym	Definition	Acronym	Definition
SeSMiK	Semi-supervised multi-kernel	GE	Graph Embedding
CaP	Prostate cancer	MP	Multi-parametric
MRI	Magnetic Resonance Imaging	MRS	Magnetic Resonance Spectroscopy
ROC	Receiver Operating Characteristic curve	DSS	Decision support system
COD	Combination of Data	COI	Combination of Interpretation
DR	Dimensionality Reduction	PCA	Principal Component Analysis
NLDR	Non-linear DR	MKL	Multi-kernel learning
SSL	Semi-supervised learning	KGE	Kernel Graph Embedding
SVM	Support vector machine	RF	Random Forest
MRF	Markov Random Field	PPMM	Probabilistic pairwise MRF
LOO	Leave-one-out	CV	Cross-validation

Table 3

Summary of T2w MRI features considered in this work with associated parameter values.

Texture feature class	Individual attributes	Parameters
Kirsch filters (4)	*X*-direction	–
	Y-direction	
	XY-diagonal	
	YX-diagonal	
Sobel filters (4)	*X*-direction	–
	Y-direction	
	XY-diagonal	
	YX-diagonal	
Directional filters (5)	*x*-Gradient	–
	y-Gradient	
	Magnitude of gradient	
	Diagonal gradient (2)	
First order statistical gray level (8)	Mean	Window size = 3 × 3, 5 × 5
	Median	
	Standard deviation	
	Range	
Haralick features (13)	Contrast energy	Window size = 3 × 3 distance between c_i, $c_j = 1$
	Contrast inverse moment	
	Contrast average	
	Contrast variance	
	Contrast entropy	
	Intensity average	
	Intensity variance	
	Intensity entropy	
	Entropy	
	Energy	
	Correlation	
	Info. measure of correlation 1	
	Info. measure of correlation 2	

Table 4

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

Index	Feature extraction strategy	CaP classifier	PPMM	High-grade CaP classifier
1.	MRS classifier	$\mathbf{h}^{MRS}$	$\mathfrak{h}^{MRS}$	$\hat{\mathfrak{h}}^{MRS}$
2.	T2w MRI classifier	$\mathbf{h}^{T2}$	$\mathfrak{h}^{T2}$	$\hat{\mathfrak{h}}^{T2}$
3.	Classifier combination (COI)	$\mathbf{h}^{IntD} = \mathbf{h}^{MRS} \times \mathbf{h}^{T2}$	$\mathfrak{h}^{IntD}$	$\hat{\mathfrak{h}}^{IntD}$
4.	Data combination (COD) classifier	$\mathbf{h}^{Int} = [\mathbf{h}^{\mathrm{MRS}}, \mathbf{h}^{T2}]$	$\mathfrak{h}^{Int}$	$\hat{\mathfrak{h}}^{Int}$
5.	SeSMiK-GE classifier	$\mathbf{h}^{T2MRS}$	$\mathfrak{h}^{T2MRS}$	$\hat{\mathfrak{h}}^{T2MRS}$

Table 5

Table showing the mean and standard deviation of (a) AUC (φ_{μ}^{AUC}), and (b) accuracy (φ_{μ}^{Acc}) for CaP detection across 29 studies via LOO and 3-fold CV strategies for $\mathbf{h}^{T2}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}^{IntD}$, $\mathbf{h}^{T2MRS}$, and $\tilde{\mathbf{h}}^{T2MRS}$.

CV strategy	$\varphi_{\mathbf{h}^{T2}}^{AUC} \pm \zeta_{\mathbf{h}^{T2}}^{AUC}$	$\varphi_{\mathbf{h}^{MRS}}^{AUC} \pm \zeta_{\mathbf{h}^{MRS}}^{AUC}$	$\varphi_{\mathbf{h}^{Int}}^{AUC} \pm \zeta_{\mathbf{h}^{Int}}^{AUC}$	$\varphi_{\mathbf{h}^{IntD}}^{AUC} \pm \zeta_{\mathbf{h}^{IntD}}^{AUC}$	$\varphi_{\mathbf{h}^{T2MRS}}^{AUC} \pm \zeta_{\mathbf{h}^{T2MRS}}^{AUC}$	$\varphi_{\tilde{\mathbf{h}}^{T2MRS}}^{AUC} \pm \zeta_{\tilde{\mathbf{h}}^{T2MRS}}^{AUC}$
(*a*)						
LOO	0.54 ± 0.18	0.61 ± 0.20	0.64 ± 0.22	0.62 ± 0.07	0.86 ± 0.08	**0.89 ± 0.07**
3-Fold	0.57 ± 0.02	0.76 ± 0.01	0.77 ± 0.01	0.64 ± 0.01	0.84 ± 0.02	**0.85 ± 0.04**
	$\varphi_{\tilde{\mathbf{h}}^{T2}}^{Acc} \pm \zeta_{\tilde{\mathbf{h}}^{T2}}^{Acc}$	$\varphi_{\tilde{\mathbf{h}}^{MRS}}^{Acc} \pm \zeta_{\tilde{\mathbf{h}}^{MRS}}^{Acc}$	$\varphi_{\tilde{\mathbf{h}}^{Int}}^{Acc} \pm \zeta_{\tilde{\mathbf{h}}^{Int}}^{Acc}$	$\varphi_{\tilde{\mathbf{h}}^{IntD}}^{Acc} \pm \zeta_{\tilde{\mathbf{h}}^{IntD}}^{Acc}$	$\varphi_{\mathbf{h}^{T2MRS}}^{Acc} \pm \zeta_{\mathbf{h}^{T2MRS}}^{Acc}$	$\varphi_{\tilde{\mathbf{h}}^{T2MRS}}^{Acc} \pm \zeta_{\tilde{\mathbf{h}}^{T2MRS}}^{Acc}$
(*b*)						
LOO	0.58 ± 0.20	0.54 ± 0.20	0.61 ± 0.18	0.45 ± 0.07	0.70 ± 0.16	0.86 ± 0.14
3-Fold	0.58 ± 0.01	0.67 ± 0.01	0.66 ± 0.02	0.51 ± 0.02	0.72 ± 0.06	**0.85 ± 0.05**

The bold values correspond to significantly higher AUC and accuracy values of SeSMiK-GE compared to the other methods.

Med Image Anal. Author manuscript; available in PMC 2014 February 01.

Table 6

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

CV strategy	$\varphi^{AUC}_{\tilde{\mathbf{h}}^{T2}} - \varphi^{AUC}_{\tilde{\mathbf{h}}^{T2MRS}}$	$\varphi^{AUC}_{\tilde{\mathbf{h}}^{MRS}} - \varphi^{AUC}_{\tilde{\mathbf{h}}^{T2MRS}}$	$\varphi^{AUC}_{\tilde{\mathbf{h}}^{COD}} - \varphi^{AUC}_{\tilde{\mathbf{h}}^{T2MRS}}$	$\varphi^{AUC}_{\tilde{\mathbf{h}}^{COI}} - \varphi^{AUC}_{\tilde{\mathbf{h}}^{T2MRS}}$
(*a*)				
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}
	$\varphi^{Acc}_{\tilde{\mathbf{h}}^{T2}} - \varphi^{Acc}_{\tilde{\mathbf{h}}^{T2MRS}}$	$\varphi^{Acc}_{\tilde{\mathbf{h}}^{MRS}} - \varphi^{Acc}_{\tilde{\mathbf{h}}^{T2MRS}}$	$\varphi^{Acc}_{\tilde{\mathbf{h}}^{COD}} - \varphi^{Acc}_{\tilde{\mathbf{h}}^{T2MRS}}$	$\varphi^{Acc}_{\tilde{\mathbf{h}}^{COI}} - \varphi^{Acc}_{\tilde{\mathbf{h}}^{T2MRS}}$
(*b*)				
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}

Table 7

Table showing the mean and standard deviation of AUC (φ_{μ}^{AUC}) for high grade CaP detection across 12 studies via LOO and 3-fold CV strategies for $\hat{\mathbf{h}}^{T2}$, $\hat{\mathbf{h}}^{MRS}$, $\hat{\mathbf{h}}^{Int}$, $\hat{\mathbf{h}}^{IntD}$, and $\hat{\mathbf{h}}^{T2MRS}$.

CV strategy	$\varphi_{\widehat{\mathbf{h}}^{T2}}^{AUC} \pm \zeta_{\widehat{\mathbf{h}}^{T2}}^{AUC}$	$\varphi_{\widehat{\mathbf{h}}^{MRS}}^{AUC} \pm \zeta_{\widehat{\mathbf{h}}^{MRS}}^{AUC}$	$\varphi_{\widehat{\mathbf{h}}^{Int}}^{AUC} \pm \zeta_{\widehat{\mathbf{h}}^{Int}}^{AUC}$	$\varphi_{\widehat{\mathbf{h}}^{IntD}}^{AUC} \pm \zeta_{\widehat{\mathbf{h}}^{IntD}}^{AUC}$	$\varphi_{\widehat{\mathbf{h}}^{T2MRS}}^{AUC} \pm \zeta_{\widehat{\mathbf{h}}^{T2MRS}}^{AUC}$
LOO	0.54 ± 0.13	0.59 ± 0.19	0.62 ± 0.18	0.61 ± 0.07	**0.84 ± 0.07**
3-Fold	0.54 ± 0.15	0.72 ± 0.13	0.71 ± 0.18	0.70 ± 0.14	**0.80 ± 0.16**

The bold values correspond to significantly higher AUC and accuracy values of SeSMiK-GE compared to the other methods.

Table 8

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

CV Strategy	$\varphi^{AUC}_{\widehat{\mathbf{h}}^{T2}} - \varphi^{AUC}_{\widehat{\mathbf{h}}^{T2MRS}}$	$\varphi^{AUC}_{\widehat{\mathbf{h}}^{MRS}} - \varphi^{AUC}_{\widehat{\mathbf{h}}^{T2MRS}}$	$\varphi^{AUC}_{\widehat{\mathbf{h}}^{COD}} - \varphi^{AUC}_{\widehat{\mathbf{h}}^{T2MRS}}$	$\varphi^{AUC}_{\widehat{\mathbf{h}}^{COI}} - \varphi^{AUC}_{\widehat{\mathbf{h}}^{T2MRS}}$
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}