Multi-modal Integration of Magnetic Resonance Imaging and Spectroscopy for Detection of Prostate Cancer

Pallavi Tiwari, Satish Viswanath, Anant Madabhushi*
Rutgers University, NJ, USA
anantm@rci.rutgers.edu

Mark Rosen,
University of Pennsylvania,
Philadelphia, PA, USA

Galen Reed, John Kurhanewicz
University of California,
San Fransisco, CA, USA

***Abstract*— Magnetic Resonance (MR) Spectroscopy (MRS) and T2-weighted (T2-w) MR Imaging (MRI) have emerged as promising tools for screening and detection of prostate cancer (CaP). T2-w MRI provides structural description of CaP while MRS provides metabolic information in terms of the concentrations of biochemicals present in the gland. While qualitative integration of both modalities has been shown to improve CaP detection, quantitative integration of MRI and MRS for building an improved (meta) classifier has been a challenge. While some researchers have adopted a decision level integration approach, which involves individually classifying the modalities and then combining the decisions, the strategy is sub-optimal since the data-sources are treated independently. The alternative, more appropriate approach, which involves data level integration by some form of combination of the MRI and MRS signals to construct a meta classifier is encumbered by differences in dimensionalities and the physical meaning of the data channels (MRS is a signal vector, while MRI yields a single scalar value at each voxel). In this paper, we present a novel multi-modal classifier which integrates texture features derived from T2-w MRI and metabolic information from MRS to automatically detect CaP. To deal with the differences in dimensionalities and physical meaning of the MRS and T2-w MRI modalities, Graph Embedding (GE), a non-linear dimensionality reduction (NLDR) scheme is used to obtain a homogeneous data representation of the two disparate modalities. Multi-modal data fusion is performed by concatenating the low dimensional transformed coordinates of the two individual data modalities. Classification was performed following integration using a Probabilistic Boosting Tree (PBT) classifier. Area under the curve (AUC) via randomized cross validation on a total of 15 T2-w MRI, MRS studies demonstrates that (a) integration of T2-w MRI and MRS data performs significantly better (0.8743 ± 0.0141) compared to using MRI (0.7327 ± 0.0236) and MRS (0.8617 ± 0.0191) individually, and (b) data level integration (0.8743 ± 0.0141) is superior to combining decisions from individual modalities (0.8592 ± 0.0161) for heterogeneous data fusion.**

I. INTRODUCTION

Prostate cancer (CaP) is the second leading cause of cancer related deaths in United States (Source: *American Cancer Society*). The current gold standard for CaP detection is a blinded sextant trans-rectal ultrasound examination which due to the poor resolution of ultrasound, has a low detection sensitivity (20-25%) [1]. Recently, T2-weighted Magnetic Resonance Imaging (T2-w MRI) and Spectroscopy (MRS) have emerged as promising modalities for CaP detection with several studies suggesting their potential as a screening tool [2], [3]. While T2-w MRI captures structural information regarding CaP, prostate MRS provides metabolic information in terms of relative concentrations of specific biochemicals (choline, creatine, citrate). Changes in the concentrations of these metabolites relative to each other at a specific location in the gland is indicative of the presence of disease at that location.

Qualitative integration of structural and metabolic MR information has been shown to yield greater CaP detection accuracy compared to using either modality individually [2]. However quantitative integration of heterogeneous modalities, such as T2-w MRI and MRS, is a challenging task. Structural information is obtained in the form of a scalar image intensity at a voxel while MRS data provides metabolic information in the form of a high dimensional metabolic vector. Consequently, both data sources first need to be homogeneously represented in the same dimensional space, prior to data integration and subsequent meta classification [4].

Currently, information fusion algorithms are categorized as either involving data level or decision level integration [5]. In data level integration, original features $(\boldsymbol{F}_A(c))$ and $(\boldsymbol{F}_B(c))$ from two modalities A and B may be combined at the feature/data level via a simple strategy like vector concatenation $\boldsymbol{F}_{AB}(c) = [\boldsymbol{F}_A(c), \boldsymbol{F}_B(c)]$ prior to classification. Such integration has the advantage of retaining any inter-source dependencies between features. However aggregating data in this manner from very different sources without accounting for differences in the number of features (dimensionality) and their relative scalings can negatively impact classification. Perhaps, on account of these data integration challenges, only one work that we are aware of [6] has previously attempted to quantitatively combine T2-w MRI intensities with metabolic spectral peaks from MRS for classification of brain tumors.

In decision level integration, individual classifications $(\mathbf{h}_A)$ and $(\mathbf{h}_B)$ from each modality A, B are combined via one of the several classifier fusion strategies such as product, averaging or majority [5]. Since each individual classifier produces a one-dimensional output, the curse-of-dimensionality may be greatly mitigated in decision level integration. Such integration strategies however tend to implicitly treat the data channels as independent and may result in sub-optimal fusion and classification.

*Work made possible via grants from Coulter Foundation (WHCF 4-29368), NJ Commission on Cancer Research, National Cancer Institute (R21CA127186-01, R03CA128081-01), Society for Imaging Informatics in Medicine (SIIM-S1), and Life Science Commercialization Award.

In this paper, we present a multi-modal MRI+MRS meta-classifier which overcomes the limitations in data-level integration by transforming the original feature spaces of the disparate sources to a lower dimensional space using Graph Embedding (GE) [7]. GE, a non-linear dimensionality reduction (NLDR) scheme offers two distinct advantages in that, (a) it provides a homogeneous low dimensional representation of T2-w MRI and MRS feature vectors, and (b) it accounts for the inherent non-linearity present in the data while reducing the dimensionality of the data. The aim of GE is to non-linearly map objects $c, d \in C$ that are adjacent in the M dimensional ambient space $(\boldsymbol{F}(c), \boldsymbol{F}(d))$ to adjacent locations in the low dimensional embedding $(\boldsymbol{S}(c), \boldsymbol{S}(d))$, where $\boldsymbol{S}(c)$, $\boldsymbol{S}(d)$ represent the m-dimensional dominant Eigen vectors corresponding to c, d $(m << M)$. Such non-linear mapping captures the inherent structure of the high dimensional data manifold such that object proximity and local geometries are preserved in the reduced Eigen space. Data level fusion can now be done, via one of several strategies (e.g. concatenation, averaging) in the transformed space. This fused data representation can then be used to train a classifier. In this paper we consider the Probabilistic Boosting Tree (PBT) to develop a meta-classifier for CaP detection by fusing MRI, MRS.

II. Methods

A. Data Description

A total of 15 1.5 Tesla (T) T2-w MRI, MRS studies were obtained prior to radical prostatectomy from University of California, San Francisco. We represent the 3D prostate T2-w scene by $\hat{\mathcal{C}} = (\hat{C}, \hat{g})$, where $\hat{C}$ is a 3D grid of voxels $\hat{c} \in \hat{C}$ and $\hat{g}(\hat{c})$ is a function that assigns an intensity value to every $\hat{c} \in \hat{C}$. We also define a spectral scene $\mathcal{C} = (C, \boldsymbol{F})$ where C is a 3D grid of MRS metavoxels, $c \in C$ and $\boldsymbol{F}$ is a spectral vector associated with each $c \in C$. Note that multiple voxels are present within the region R_{cd} between any two adjacent metavoxels $c, d \in C$. For the sake of convenience we represent R_{cd} as $R(c)$, where $\hat{c} \in R(c)$.

B. Feature extraction for MRI-MRS

1) Feature extraction from MRS: Figure 1(a) shows a MRS spectral grid superposed on a T2-w MRI slice with expert annotated class labels $Y(c) \in \{1, 2, 3, 4, 5\}$ based on a clinical standardized 5-point scale [8] which classifies each spectra as definitely benign (1), probably benign (2), equivocal (3), probably cancer (4) and definitely cancer (5). In this work, all spectra labeled (4, 5) were assumed to be CaP and all spectra labeled as (1, 2) were assumed as benign. The 15 studies comprised 1331 class 1, 2 and 407 class 4, 5 spectra. Clinically, identifying CaP is based on calculating the ratios of choline, creatine and citrate peaks (CC/C) and identifying spectra with abnormal ratios. However, peak detection algorithms are often constrained in their ability to deal with the poor signal to noise ratio of non-invasive MRS [9]. Instead of extracting features derived from spectral peaks, we consider the spectra in its totality. Thus, each $c \in C$, $\boldsymbol{F}(c) = [f_\alpha(c) | \alpha \in \{1, ...U\}]$ represents the MR spectral vector, reflecting the frequency component of each of the U metabolites.

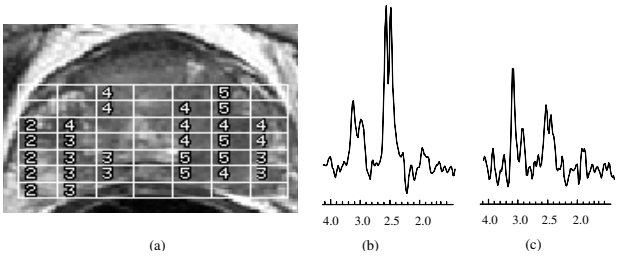


Fig. 1. (a) MRS metavoxels $c \in C$ superposed on the corresponding T2-w MRI section. Representative spectra from (b) class 2 (probably benign) and (c) class 4 (probably CaP) are shown. Note that classifying such similar looking spectra manually is a challenging, error prone, and laborious task for radiologists.

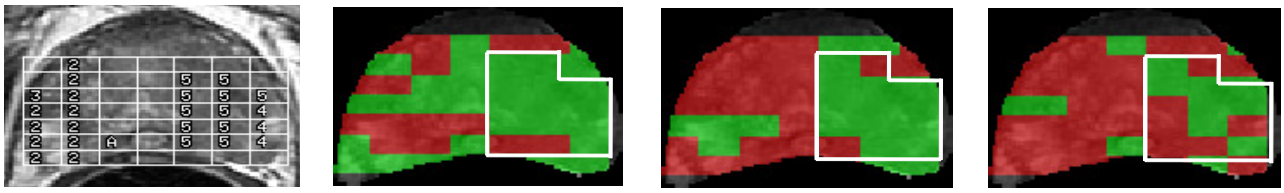
2) Texture Feature extraction from MRI: 38 texture features scenes were obtained from each MRI scene. Our feature representations were motivated by previous demonstration of their utility in discriminating between the CaP and non-CaP classes [10]. We calculated the feature scenes $\hat{\mathcal{G}}_u = (\hat{C}, \hat{g}_u)$ for each $\hat{\mathcal{C}}$ by applying the feature operators $\Phi_u, u \in \{1, \ldots, 38\}$, within a local neighborhood associated with every $\hat{c} \in \hat{C}$. Hence $\hat{g}_u(\hat{c})$ is the feature value associated with feature operator Φ_u at voxel $\hat{c}$. Below is a brief description of the texture features extracted from the T2-w scenes.

1) **Gradient:** Thirteen non-steerable gradient features were obtained via convolution of $\hat{\mathcal{C}}$ with the Sobel, Kirsch and standard derivative operators [11].
2) **First Order Statistical:** These include features corresponding to the mean, median, standard deviation, and range for the gray values of voxels within a sliding window neighborhood around each $\hat{c} \in \hat{C}$.
3) **Second Order Statistical Features:** A total of 13 Haralick features were derived at every voxel $\hat{c} \in \hat{C}$, based on calculation of a co-occurrence matrix of the intensities $\hat{g}(\hat{c})$ within $\hat{\mathcal{C}}$. The features included energy, entropy, inertia, contrast, correlation, sum average, sum variance, sum entropy, difference average, difference variance, difference entropy, local homogeneity and average deviation.

We define a T2-w MRI texture feature vector for each metavoxel $c \in C$ by taking the average of the feature values within the corresponding metavoxel as $g_u(c) = \frac{1}{|R(c)|} \sum_{\hat{c} \in R(c)} [\hat{g}_u(\hat{c})]$ where $|R(c)|$ represents the cardinality of the set of voxels contained in the space between any 2 adjacent meta-voxels. The corresponding feature vector is then given as $\boldsymbol{G}(c) = [g_u(c) | u \in \{1, \ldots, 38\}], \forall c \in C$.

C. Data Representation using Graph Embedding (GE)

The aim of GE [7] is to find an embedding vector $\boldsymbol{S}(c_i)$, $\forall c_i \in C$, $i \in \{1, \ldots, |C|\}$, such that if locations $c_i, c_j \in C, i, j \in \{1, \ldots, |C|\}$, are adjacent in the high dimensional feature space, then $||\boldsymbol{S}(c_i) - \boldsymbol{S}(c_j)||_2$ should be small, where $||.||_2$ represents the Euclidean norm. This will only be true if the distances between all $c_i, c_j \in C$ are preserved in the low dimensional mapping of the data. To



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

Fig. 2. (a) Original T2-w image, (b)-(d) demonstrate the classification results superposed on a single T2 slice, obtained by thresholding the classifiers (b) $\mathbf{h}^{MRS}$, (c) $\mathbf{h}^{T2MRS}$ and (d) $\mathbf{h}^{T2MRS}_{Ind}$ respectively at operating point (Θ=0.6).

compute the optimal embedding, we first define adjacency matrix $W_{GE} \in \Re^{|C|\times|C|}$ as

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

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

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

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

1) Data representation for MRS: At each location c, high dimensional MRS spectrum $\boldsymbol{F}(c)$ is reduced to a lower dimensional embedding $\boldsymbol{S}^{MRS}(c)$ using GE. We have shown previously that GE provides a low dimensional representation of MRS data that is more amenable for the purposes of classification [3] compared to such feature representation methods as principal component analysis and z-score.

2) Data representation for MRI: The high dimensional T2-w MRI texture feature vector $\boldsymbol{G}(c)$ is reduced to a low dimensional Eigen vector, $\boldsymbol{S}^{T2}(c)$ using GE, where the dimensionality of $\boldsymbol{S}^{T2}(c)$ is significantly lower compared to $\boldsymbol{G}(c)$ ($|\boldsymbol{S}^{MRI}(c)| << |\boldsymbol{G}(c)|$).

D. Data fusion

1) Data level integration: Owing to the physical and dimensional differences in the MRS and MRI features, the MRS-MRI meta-classifier is created in the joint MRI and MRS embedding space. A direct concatenation of the MRI and MRS embedding coordinates can be obtained as $\boldsymbol{S}^{T2MRS} = [\boldsymbol{S}^{T2}(c), \boldsymbol{S}^{MRS}(c)]$. The concatenated feature vector $\boldsymbol{S}^{T2MRS}(c)$ is then used for training a probabilistic boosting tree classifier.

2) Decision level integration: Decision integration refers to the combination of weak classifiers (based on individual modalities) via some pre-decided rule such as averaging or majority voting. Every $c \in C$ is assigned to one of several classes via multiple weak classifiers, $\mathbf{h}_n(c), n \in \{1,\ldots,N\}$, where $\mathbf{h}_n(c)$ can either be a probability or a class membership (such as (0, 1)). In our case, $N = 2$ (MRI, MRS). Classifiers $\mathbf{h}_1, \mathbf{h}_2$ are individually trained on $\boldsymbol{F}(c)$ and $\boldsymbol{G}(c)$, for $c \in C$. Independence assumption can then be invoked to fuse $\mathbf{h}_1$ and $\mathbf{h}_2$ at each $c \in C$ as $\mathbf{h}_{Ind}(c) = \mathbf{h}_1(c) \times \mathbf{h}_2(c)$.

E. Classification using Probabilistic boosting tree (PBT)

The PBT algorithm [12] is a combination of the Adaboost and decision trees classifiers. It iteratively generates a tree structure of length L in the training stage where each node of the tree is boosted with T weak classifiers. The hierarchical tree is obtained by dividing new samples in two subsets of $\tilde{\boldsymbol{F}}_{Right}$ and $\tilde{\boldsymbol{F}}_{Left}$ and recursively training the left and right sub-trees using Adaboost. To solve for overfitting, error parameter ϵ is introduced such that samples falling in the range $[\frac{1}{2}-\epsilon, \frac{1}{2}+\epsilon]$ are assigned to both subtrees with probabilities $(\boldsymbol{F}(c), p(1|c)) \rightarrow \tilde{\boldsymbol{F}}_{Right}(c)$ and $(\boldsymbol{F}(c), p(0|c)) \rightarrow \tilde{\boldsymbol{F}}_{Left}(c)$, where the function $p(Y|c)$ represents the posterior class conditional probability of c belonging to class $Y \in \{0,1\}$. The algorithm stops when misclassification error hits a predefined threshold θ.

During testing, the conditional probability of each $c \in C$ is calculated at each node based on the learned hierarchical tree. PBT generates a posterior conditional probability for CaP, $p(w_T|\boldsymbol{S}^{\phi})$, $\phi \in \{T2, MRS, T2MRS\}$ for each spectral location $c \in C$, based on the embedding vector $\boldsymbol{S}^{\phi}(c)$, where w_T represents CaP class (4, 5 spectra). We define $\mathbf{h}^{\phi,\rho}(c)$ as the binary prediction result at each threshold $\rho \in [0,1]$ such that $\mathbf{h}^{\phi,\rho}(c) = 1$ when $p(w_T|\boldsymbol{S}^{\phi}) \geq \rho$, 0 otherwise.

III. Results

A. Qualitative Results

Figure 2 shows the original T2-w MRI image (2(a)) overlaid with the MR spectral grid. Figures 2(b)-(d) show the binary prediction results for CaP (identified as green) obtained via PBT from $\mathbf{h}^{MRS}$ (2(b)), $\mathbf{h}^{T2MRS}$ (2(c)), and $\mathbf{h}^{T2MRS}_{Ind}$ (2(d)) for ρ=Θ=0.6, where Θ is the operating point of the ROC curve. Note the high detection accuracy obtained via data level integration of T2-w MRI and MRS compared to the individual MRI, MRS classifiers. Also note that $\mathbf{h}^{T2MRS}$ yields higher CaP detection sensitivity, specificity compared to $\mathbf{h}^{T2MRS}_{Ind}$. The white outline in Figure 2(b)-(d) shows the spatial locations of the CaP spectra (4/5).

B. Quantitative Results

Figure 3(a) shows average ROC curves obtained for $\mathbf{h}^{T2}$, $\mathbf{h}^{MRS}$, and $\mathbf{h}^{T2MRS}$ across 25 runs of randomized cross validation for $m = 10$ using 300 training samples from each class (1, 2 and 4, 5). For each randomized run, 300 samples from each class were randomly chosen for training and the remaining for testing. PBT was trained using $\theta = 0.45$ and $\epsilon = 0.4$ as suggested in [12]. Five weak classifiers were used to train each node of PBT using AdaBoost for

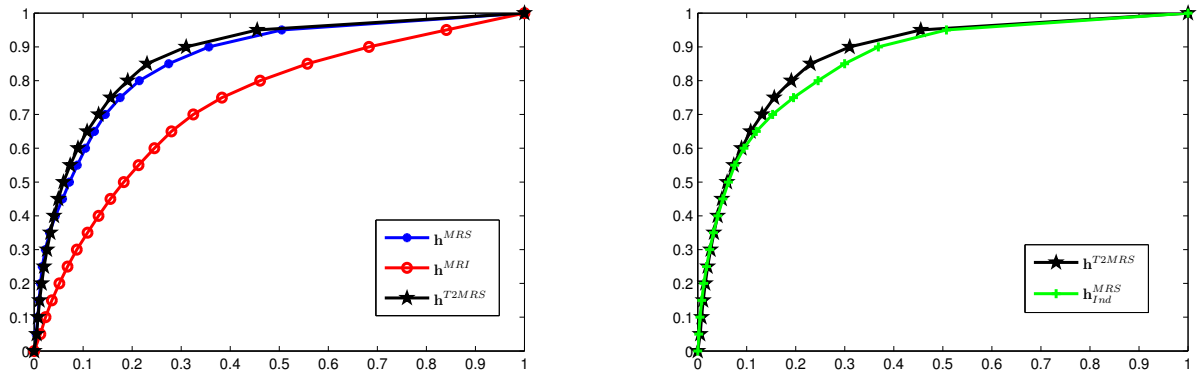


Fig. 3. Average ROC curves across 300 training samples over 25 cross validation runs for classifiers (a) $\mathbf{h}^{T2}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{T2MRS}$, and (b) $\mathbf{h}^{T2MRS}$, $\mathbf{h}_{Ind}^{T2MRS}$. Different colors correspond to different classifiers. The best performance in both (a) and (b) corresponds to the classifier (shown in black) based on data level integration of structural and metabolic data ($\mathbf{h}^{T2MRS}$).

$L = 4$. The highest AUC value corresponds to the classifier $\mathbf{h}^{T2MRS}$ (shown in black), while the lowest is for $\mathbf{h}^{T2}$ (shown in red). Figure 3(b) shows average ROC curves obtained from $\mathbf{h}^{T2MRS}$ and $\mathbf{h}_{Ind}^{T2MRS}$ across 25 runs of randomized cross validation using 300 training samples. The higher AUC values for $\mathbf{h}^{T2MRS}$ compared to $\mathbf{h}_{Ind}^{T2MRS}$ suggests that data level integration is superior to decision level classification. AUC and accuracy values for $\mathbf{h}^{T2}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{T2MRS}$, and $\mathbf{h}_{Ind}^{T2MRS}$ averaged over 25 cross validation runs are summarized in Table 1 with corresponding standard deviations. Paired student t-tests were also conducted for AUC at the operating point of the average ROC curves, with the null hypothesis being no improvement in performance of $\mathbf{h}^{T2MRS}$ when compared to the other 3 classifiers ($\mathbf{h}^{T2}$, $\mathbf{h}^{MRS}$, $\mathbf{h}_{Ind}^{T2MRS}$). Significantly superior performance ($p < 0.05$) was observed for $\mathbf{h}^{T2MRS}$ suggesting that integrating structural textural features and metabolic information at the data-level offers the most optimal results for CaP detection.

Classifier	AUC	Accuracy
$\mathbf{h}^{T2}$	0.7327±0.0236	0.6772±0.0342
$\mathbf{h}^{MRS}$	0.8617±0.0191	0.7944±0.0333
$\mathbf{h}^{T2MRS}$	0.8743±0.0141	0.7951±0.0254
$\mathbf{h}_{Ind}^{T2MRS}$	0.8592±0.0161	0.7926±0.0402

TABLE I

AVERAGE AND STANDARD DEVIATION OF AUC AND ACCURACY VALUES T2-W MRI, MRS, DATA LEVEL, AND DECISION LEVEL CAP CLASSIFIERS OVER 25 RUNS OF RANDOMIZED CROSS VALIDATION.

IV. CONCLUDING REMARKS

In this paper, we presented a novel multi-modal classifier to quantitatively combine structural and metabolic information obtained from T2-w MRI and MRS for CaP detection. Our scheme attempts to combine the multiple, heterogeneous modalities at the data level, as opposed to the decision level. In order to overcome the inter-modal feature and dimensionality differences, Graph Embedding, a NLDR scheme was used to homogeneously represent disparate sources of data. Data fusion was then performed in the reduced dimensional Eigen space. A Probabilistic Boosting tree classifier was trained to discriminate between CaP (4/5) and benign (1/2) spectra using the integrated T2-w MRI, MRS embedding vectors. Randomized cross validation results on a total of 15 T2-w MRI, MRS studies demonstrate that (a) classification based on data-level integration of T2-w MRI, MRS performs significantly better compared to either modality individually, and (b) data level integration, obtained by concatenating low dimensional embedding coordinates of T2-w MRI, MRS outperforms a decision level classifier obtained by combining individual classifier decisions from T2-w MRI and MRS respectively. The results presented in this paper also suggest that our method of data level fusion via use of GE could be applied to combining and classification of other imaging and non-imaging data modalities.

REFERENCES

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

[2] Kurhanewicz, J., et al., "The prostate: MR Imaging and Spectroscopy. Present and future," *Radiology* vol. 38, pp. 115-138, 2000.

[3] Tiwari, P., Rosen, M., Madabhushi, A., "A Hierachical Unsupervised Spectral Clustering Scheme for Detection of Prostate Cancer from Magnetic Resonance Spectroscopy", *MICCAI,*, pp. 278-286, 2007.

[4] Viswanath, S., Tiwari, P., Rosen, M., Madabhushi, A., "A meta-classifier for detecting prostate cancer by quantitative integration of in vivo Magnetic Resonance Spectroscopy and Magnetic Resonance Imaging", *SPIE Medical Imaging*, vol. 6915, 2008.

[5] Rohlfing, T., Pfefferbaum, A., et al., "Information Fusion in Biomedical Image Analysis: Combination of Data vs. Combination of Interpretations", *IPMI*, vol. 3565, pp.150-161, 2005.

[6] Simonetti, AW., Melssen, WJ., et al., "Combination of Feature-Reduced MR Spectroscopic and MR Imaging Data for Improved Brain Tumor Classification.", *NMR in Biomedicine*, vol. 18, pp. 34-43, 2005.

[7] Jianbo Shi and Jitendra Malik, "Normalized Cuts and Image Segmentation", *IEEE TPAMI*, vol. 22(8), pp. 888-905, 2000.

[8] Jung, J., Coakley, F., Vigneron, D., et al., "Prostate depiction at endorectal MR Spectroscopic Imaging: Investigation of a standardized evaluation system", *Radiology*, vol. 233, pp. 701-708, 2004.

[9] Wetter, A., Engl, T., et al., "Combined MRI and MR Spectroscopy of the Prostate Before Radical Prostatectomy.", *Radiology*, vol. 187, pp. 724-730, 2006.

[10] Madabhushi, A., Feldman, M., et al., "Automated Detection of Prostatic Adenocarcinoma from High-Resolution Ex Vivo MRI.", *IEEE TMI*, vol. 24(12), pp. 1611-1625, 2005.

[11] Russ, J.C., "The Image Processing Handbook", *CRC Press*, 2007.

[12] Tu, Zhuowen, "Probabilistic Boosting-Tree: Learning Discriminative Models for Classification, Recognition, and Clustering", *IEEE ICCV*, vol. 2, pp. 1589-1596, 2005.