Published in final edited form as:

NMR Biomed. 2012 April ; 25(4): 607–619. doi:10.1002/nbm.1777.

Multimodal Wavelet Embedding Representation for data Combination (MaWERiC): Integrating Magnetic Resonance Imaging and Spectroscopy for Prostate Cancer Detection

Pallavi Tiwari, MS¹, John Kurhanewicz, PhD², Satish Viswanath, MS¹, Akshay Sridhar, BS¹, and Anant Madabhushi, PhD^{1,*}

¹Department of Biomedical Engineering, Rutgers University, 599 Taylor Road, Piscataway, NJ 08854

²University of California, Department of Radiology, San Francisco, CA, 94143

Abstract

Rationale and Objectives—To develop a computerized data integration framework (MaWERiC) for quantitatively combining structural and metabolic information from different Magnetic Resonance (MR) imaging modalities.

Materials and Methods—In this paper, we present a novel computerized support system that we call Multimodal Wavelet Embedding Representation for data Combination (MaWERiC) which (1) employs wavelet theory and dimensionality reduction for providing a common, uniform representation of the different imaging (T2-w) and non-imaging (spectroscopy) MRI channels, and (2) leverages a random forest classifier for automated prostate cancer detection on a per voxel basis from combined 1.5 Tesla *in vivo* MRI and MRS.

Results—A total of 36 1.5 T endorectal *in vivo* T2-w MRI, MRS patient studies were evaluated on a per-voxel via MaWERiC, using a three-fold cross validation scheme across 25 iterations. Ground truth for evaluation of the results was obtained via *ex-vivo* whole-mount histology sections which served as the gold standard for expert radiologist annotations of prostate cancer on a per-voxel basis. The results suggest that MaWERiC based MRS-T2-w meta-classifier (mean AUC, $\mu = 0.89 \pm 0.02$) significantly outperformed (i) a T2-w MRI (employing wavelet texture features) classifier ($\mu = 0.55 \pm 0.02$), (ii) a MRS (employing metabolite ratios) classifier ($\mu = 0.77 \pm 0.03$), (iii) a decision-fusion classifier, obtained by combining individual T2-w MRI and MRS classifier outputs ($\mu = 0.85 \pm 0.03$) and (iv) a data combination scheme involving combination of metabolic MRS and MR signal intensity features ($\mu = 0.66 \pm 0.02$).

Conclusion—A novel data integration framework, MaWERiC, for combining imaging and non-imaging MRI channels was presented. Application to prostate cancer detection via combination of T2-w MRI and MRS data demonstrated significantly higher AUC and accuracy values compared to the individual T2-w MRI, MRS modalities and other data integration strategies.

Keywords

Multi-modal integration; Magnetic Resonance Imaging; Magnetic Resonance Spectroscopy; Haar wavelets; Gabor texture features; PCA; Random Forest classifier; prostate cancer

INTRODUCTION

In the biomedical domain, there is a real need for developing data integration strategies for combining discriminatory features from multiple modalities to develop multi-modal meta-classifiers for improved disease detection, diagnosis, and prognosis (1). While data integration methods have been proposed for quantitatively combining multiple imaging modalities (2–4), these tools are not readily extensible to integration of imaging and non-imaging data on account of differences in scale and feature dimensionality. For instance, consider the difficulties in quantitatively combining, at the voxel-level, T2-weighted (T2-w) magnetic resonance (MR) imaging (reflecting structural attributes) acquired as scalar intensity values, with MR spectroscopy (MRS) acquired as a vector (or spectrum) of metabolite concentrations; each modality encoding a different type (structural or metabolic) and dimensionality of information. Nonetheless, both modalities reflect information regarding the same region of interest they are captured from, and consequently, examining them in conjunction is crucial. Manual diagnosis of prostate cancer (CaP) on T2-w MRI and MRS involves first visually identifying hypo-intense regions on T2-w MRI, followed by inspection of spectrum at those corresponding spatial locations for changes in metabolite (choline, creatine, citrate) ratios (5). However, some studies have shown that manual interpretation and visual integration of multi-modal data is subjective and thus prone to inter-and intra-observer variability (6,7). It is hence desirable to build a data integration (DI) based computerized support system (DI-CSS) meta-classifier that can accurately extract and combine relevant information from both imaging and non-imaging data channels for improved disease classification (3,8). Such a CSS could then be integrated into a clinical setting to assist radiologists in accurate characterization, staging, as well as directing and evaluating disease treatment and treatment response.

One of the major challenges in developing a CSS for quantitative integration of imaging and non-imaging data (hereafter referred to as data level integration) is to represent the different data channels within a common, unified representation prior to data integration. In this work, we introduce a novel computerized framework to address this problem. We refer to this framework as *Multimodal Wavelet Embedding Representation for data Combination (MaWERiC)*, where the aim is to enable seamless quantitative data-level integration of imaging and non-imaging data while overcoming the differences in scales and dimensionalities. Additionally, MaWERiC provides the ability to build meta-classifiers from the different modalities once the data representation and data integration issues have been resolved. While MaWERiC could be applied to a number of different domains and applications, in this work we demonstrate the applicability of MaWERiC in building meta-classifiers for improved prostate cancer detection using T2-w MRI and MRS.

Prostate cancer (CaP) is the second leading cause of cancer related deaths amongst men with 217,730 new cases and 32,050 estimated deaths in United States in 2010 (9). The current gold standard for CaP detection is a sextant trans-rectal ultrasound examination which is known to have a low detection sensitivity (20–25%) due to the poor image resolution of ultrasound (10). In the last decade, T2-w MRI has shown great potential for characterizing disease presence as well as staging of CaP (11). MRS has recently emerged as a complement to traditional T2-w MRI for improved CaP diagnosis (12). While clinical studies have shown that the use of structural and metabolic MR information yields greater CaP detection accuracy compared to diagnosis based off any individual modality (5,13–16), few attempts have been made to quantitatively combine the different information channels (17,18). While a few groups have proposed CSS classifiers for combining multiple MR imaging protocols (T2-w, dynamic contrast enhanced (DCE), line-scan diffusion, diffusion weighted imaging (DWI), T2-mapping MRI) (3,8), to the best of our knowledge no previous methods for

quantitative integration of T2-w MRI and spectroscopy (imaging and non-imaging modalities) for CaP detection have been presented.

In this work we aim to leverage MaWERiC to build an integrated structural, metabolic meta-classifier which will assign a probability of CaP presence at every spatial location on *in-vivo* prostate T2-w-MRI/MRS studies. Figure 1 illustrates the organization of our MaWERiC strategy. Our scheme is based on combining wavelet (Gabor and Haar filters) features (19) extracted from both the T2-w MRI and MRS modalities. The advantage of using wavelet transform (19) is that it can provide multi-resolution discriminatory information from different data modalities, including but not limited to signals and images (20–22). The 2D-Gabor wavelet filter is defined as the convolution of a 2D Gaussian function with a sinusoid (23). A Gabor filter bank is then generated by variation of the associated scale and orientation parameters. This filter bank provides a means for multi-scale, multi-orientation texture characterization and representation of an image. Haar wavelet decomposition is a commonly employed signal filtering technique (19) which provides a way of extracting the class discriminating frequency components that may yield higher classifier accuracy compared to the original signal (24). An advantage of the Haar wavelet (24) is that it preserves features that are representative of abrupt changes in signals, dominant spectral peaks such as those that correspond to the most significant metabolites on MRS, while simultaneously eliminating spectral noise. Both the Gabor and Haar wavelet filters have been previously used in conjunction with CSS classifiers to distinguish between different data classes for various biomedical applications (20–22,25). In the context of this work, multi-resolution features for T2-w MRI and MRS are obtained via use of the Gabor and Haar wavelet filters respectively.

While the wavelet based representations of the MRS, T2-w MRI channels provide a uniform, feature representation of the data, the feature vectors obtained via the application of the Gabor and Haar wavelet filters are of a very high dimensionality and hence subject to the *curse of dimensionality* (26). Consequently a subsequent step is application of Principal Component Analysis (PCA) (27) to the high dimensional feature vectors obtained (via wavelet decomposition) from each individual data modality (T2-w MRI, MRS) to obtain a reduced dimensional representation of the data and thus make the data representation amenable to the application of the classifier. The representation of the MRS and T2-w MRI data in terms of the Eigenvectors (obtained via PCA) also allows for the overcoming of the scale (resolution) and dimensionality differences between the two modalities, since the wavelet representations obtained from each individual modality are of varying dimensionality, and can be reduced to the same number of Eigenvectors. Data level fusion is then performed by concatenating the principal Eigenvectors corresponding to each modality. This fused Eigenvector representation is then used to train a Random forest (RF) classifier. RF is a commonly used ensemble classifier that combines predictions from several weak classifiers to generate a more accurate and stable classifier (28). The RF classifier has previously been successfully employed for various biomedical classification applications (29–31). Advantages of RF include, (1) ability to integrate a large number of input variables, (2) robustness to noise in the data, and (3) relatively few tuning-parameters.

PREVIOUS WORK

Quantitative data integration schemes with the intent of building meta-classifiers that combine discriminatory information from different data channels can be categorized into two major classes. The first class of approaches (combination of data (COD)) involves combining the data prior to classification. The second class of methods, known as decision level classification (combination of interpretation (COI)) (32), involves training individual data classifiers (uni-modal classification) and combining the outputs from each classifier. In

the context of COD approaches, data can be combined either (a) in its original acquired form or (b) following the application of an appropriate transformation to each data modality to obtain improved discriminatory information over and beyond the original acquired data. MaWERiC is a COD scheme where wavelets and PCA are used to extract discriminatory features from each modality prior to data integration. The meta-classifier trained on this wavelet, PCA based data representation is then applied to the problem of CaP detection from multi-modal MRI.

In the subsequent subsections, we discuss both individual unimodal classifiers for T2-w MRI (22) and MRS (33), and multi-modal meta-classifiers for CaP detection using multi-parametric MRI. We also briefly discuss previous combination strategies involving MRI-MRS for brain tumor classification (17,18).

Unimodal classifiers for T2-w MRI

Since image intensity on T2-w MRI, is susceptible to artifacts such as bias field inhomogeneity (34) and intensity non-standardness (35), researchers have explored alternate representations of T2-w image intensities (e.g. Gabor or wavelet based texture features (22)) to build classifiers for predicting CaP presence on MRI. In (22), Madabhushi *et al.* presented a supervised CSS system for detection of CaP from 4 Tesla (T) *ex vivo* prostate T2-w MRI where 33 3D texture features (statistical, gradient, and Gabor) were quantitatively extracted at each voxel (T2-w MRI spatial resolution). These extracted features were then used to train a number of supervised classifiers (Adaboost, Bayes, and Decision Trees) which were employed to assign a probability of CaP presence at each image voxel.

Unimodal classifiers for MRS

Previous CSS approaches that have been developed in the context of MRS data can be broadly divided into two main categories: (a) signal quantification (model dependent) (36–38), and (b) statistical pattern recognition based (model independent) approaches (30,33,39,40). Commonly used MRS quantification methods include VARPRO (36), AMARES (37), and QUEST (38), which are software utilities where the objective is to minimize the squared distance between the acquired data and a model basis function built on prior knowledge about the metabolic profiles of a typical MR spectrum. Pattern recognition based features on the other hand, try to capture the underlying variance in the data using regression analysis. Kelm *et al.* (30) presented a comparative study of classification techniques for prostate MRS data based on pattern recognition methods such as PCA (27) and Independent Component Analysis (ICA) (41) against quantification based feature extraction methods using SVM, RF and Gaussian processes classifiers. They showed that pattern recognition based classifiers provided better classification results for CaP detection compared to MRS quantification schemes. In (33) Tiwari *et al.* presented a CSS for CaP detection using 1.5 Tesla *in vivo* prostate MRS where each prostate spectrum was classified, on a per voxel basis, as either belonging to cancerous or non-cancerous classes using a hierarchical, clustering scheme in conjunction with non-linear dimensionality reduction (NLDR) methods. NLDR schemes were employed to obtain a low dimensional representation of high dimensional MR spectra, followed by hierarchical *k*-means clustering to identify CaP signatures in the prostate. A sensitivity of 89.33% and a specificity of 79.79%, on a per voxel basis, were obtained across a total of 18 1.5 T prostate MRS studies. Luts *et al.* (40) presented a method which leveraged ICA and Relief-F in conjunction with SVM and linear discriminant analysis classifiers for brain tumor classification using MRS.

Combining imaging-imaging MRI channels

In (3) Chan *et al.* presented a statistical classifier which integrated texture features from multi-protocol 1.5 T *in vivo* MRI to generate a statistical probability map representing

likelihoods of cancer for different regions within the prostate. Liu *et al.* (42) examined multi parametric *in vivo* MRI maps (T2-w, DCE, DWI) within a fuzzy Markov Random Fields framework. The maps were generated via curve fitting of data from each of the protocols with the ROI limited to the peripheral zone of the prostate, while the evaluation of the results was done against manually delineated CaP regions on MRI (with corresponding whole-mount histology and *ex vivo* MRI data used for reference). Ampeliotis *et al.* (43) explored the use of image intensity features from both DCE and T2-w MRI data for the classification of CaP. A statistically significant improvement in classifier performance when fusing modalities, over the use of individual modalities, was reported. Another multi-protocol (DCE and T2-w) MRI based CSS was presented in (44) which combined pharmacokinetic features from DCE along with T2-w image intensities.

Decision Integration strategies for integrating MRI, MRS

To the best of our knowledge, no data or decision integration methods for combining imaging and spectroscopy in the context of prostate cancer have been proposed. Jesneck *et al.* (45) proposed a decision integration scheme where probabilities for breast cancer presence obtained from classifiers built individually from features extracted from different imaging modalities (sonogram, mammogram) and patient history (non-imaging) were combined to obtain an integrated classifier for improved breast cancer diagnosis. Another COI scheme was presented in (46), which combined classifier outputs from three heterogeneous modalities: face recognition, voice recognition, and hand geometry within a Bayesian framework for improved biometric based personnel identification. However, one of the major disadvantages of such decision integration based approaches is that all inter-source dependencies between modalities may be lost, given that each modality is being treated independently (26).

Data Integration strategies for integrating MRI, MRS

A data integration method involving integration of multi-protocol MR image intensities (T1-w, T2-w, proton density-weighted, and gadolinium-DTPA) with the areas under spectral peaks of specific metabolites (*myo*-inositol, glucose, choline, creatine, glutamate/glutamine, *N*-acetyl aspartate, lactate/fatty acids and fatty acids) from MRS was presented in (18) for classifying four brain tumor types (Tumor II, III, IV, meningioma), healthy tissues and cerebrospinal fluid (CSF). All MRS and MRI features were directly concatenated into a single joint feature vector and employed in conjunction with a Mahalanobis distance based classifier. The classifier results showed that the voxel-level classification obtained via this multi-modal feature combination was significantly superior compared to the results obtained using unimodal classifiers.

In this paper we aim to demonstrate that MaWERiC provides significant improvements over previous COD and COI approaches for prostate cancer detection using multi-modal MRI (18,45). We show the utility of MaWERiC for developing the first CSS scheme which quantitatively combines T2-w MRI and MRS for CaP detection, and demonstrate significant improvements over using individual modalities and other previously presented state of the art data fusion strategies (18,45).

MATERIALS AND METHODS

Data

A total of 36 1.5 Tesla (T) T2-w MRI, MRS studies were obtained prior to radical prostatectomy. All of these studies were biopsy proven prostate cancer patient studies that were clinically referred for a prostate cancer MR staging exam for improved therapeutic selection. MR imaging was performed by using a 1.5-T whole-body MR imaging unit

(Signa; GE Medical Systems, Milwaukee, Wisconsin). The patients were imaged while in the supine position by using a body coil for signal excitation and a pelvic phased-array coil (GE Medical Systems) combined with a balloon-covered expandable endorectal coil (Medrad, Pittsburgh, PA) for signal reception. All MR images were routinely post-processed to compensate for the reception profile of the endorectal and pelvic phased-array coils. A spectroscopic MR imaging volume was then selected by an expert to maximize coverage of the prostate while minimizing the inclusion of peri-prostatic fat and rectal air. Three-dimensional proton (^1H) MR spectroscopic imaging data were acquired by using a water and lipid-suppressed double-spin-echo point-resolved spectroscopic sequence optimized for the quantitative detection of both choline and citrate. Water and lipid suppression was achieved by using the band selective inversion with gradient dephasing technique (47). To eliminate signals from adjacent tissues, especially periprostatic lipids and the rectal wall (48), outer voxel saturation pulses also were used. Data sets were acquired as $16 \times 8 \times 8$ phase-encoded spectral arrays (1024 voxels) by using a nominal spectral resolution of 0.24–0.34 cm^3 , 1000/130, and a 17-minute acquisition time.

Pre-processing—Three-dimensional, MR spectroscopic imaging data were processed and aligned with the corresponding T2-w imaging data using a combination of in-house software and Interactive Display Language (Research Systems, Boulder, Colorado) software tools (48). The raw spectral data were apodized with a 1-Hz Gaussian function and Fourier transformed in the time domain and in three spatial domains. Choline, creatine, and citrate peak parameters (i.e., peak area, peak height, peak location, and line width) were estimated by using an iterative procedure that was used to first identify statistically significant peaks (those with a signal-to-noise ratio higher than 5) in the magnitude spectrum. The frequency shift that best aligns the spectral peaks with the expected locations of choline, creatine, citrate, and residual water is then estimated. Subsequently, the spectra are phased by using the phase of the residual water and the metabolite resonances. Baseline values were corrected by using a local nonlinear fit to the non-peak regions of the spectra. Subsequent feature extraction and classification steps were performed using algorithms developed within the MATLAB (The MathWorks, Inc.) programming environment.

Ground Truth annotations—For all the studies considered in this work, *ex vivo* whole mount histological sections obtained from radical prostatectomy specimens were available. The “ground truth” CaP extent on the MR imaging was manually delineated by an expert (JK) by visually registering corresponding histological and radiological sections; correspondence between sections having been determined manually by visually determining anatomical fiducials on the histology and the imaging. Having delineated the CaP extent on the MR imaging, an expert spectroscopist then labeled the spectral voxels within the CaP annotated regions on the MRI/MRS according to the 5-point scale. Figure 2 shows the standardized 5-point scale developed by Jung *et al.* (49) which was used to visually classify each spectrum as being either (a) definitely benign (scale 1), (b) probably benign (scale 2), (c) equivocal (scale 3), (d) probably cancer (scale 4), and (e) definitely cancer (scale 5). In this study, all spectra labeled (4, 5) were assumed to be CaP and all spectra labeled as (1, 2) were assumed as benign. The voxels labeled as 3 and atrophic (A) were assumed to be indeterminate and consequently excluded from our analysis. The 36 studies comprised 2120 class 1, 2 and 1026 class 4, 5 spectra (Table 1). The class labels for the individual spectral voxels, assigned via a combination of manual registration of histology and MRI and subsequent visual inspection, were used as the surrogate ground truth for CaP extent on the MRI/MRS. This ground truth surrogate is then used for training and evaluation of the MaWERiC classifiers.

Notation—We define a metavoxel in the MRS grid as $c \in C$, where C is a 3D grid of MRS metavoxels. For each $c \in C$, $F(c) = [f_\alpha(c)/\alpha \in \{1, \dots, M\}]$, represents the MR spectral vector, reflecting the frequency component of each of the M metabolites being measured (33). Note that the MRS metavoxel and T2-w MRI voxel are at different resolutions where 1 MRS metavoxel corresponds to approximately 90 T2-w MRI voxels. Feature extraction (3,22) from T2-w MRI is performed on a per voxel (T2-w resolution) basis. The responses of various texture filters (described in the next Section) are averaged over all voxels within each metavoxel $c \in C$. The T2-w MRI feature vector is obtained by calculating a mean feature vector at each $c \in C$. Hence at every metavoxel $c \in C$, the corresponding intensity feature vector is denoted as $F^{T2}(c)$ while the corresponding mean Gabor wavelet feature vector (details in the next section) is denoted as $F^{T2w}(c)$. For MRS, the feature vector comprised of ratios of concentrations of metabolites is denoted as $F^{MRS}(c)$, while the corresponding Haar wavelet feature vector for each $c \in C$ is denoted as $F^{MRSw}(c)$. A classifier is defined as $h(c)$, $h \in \{\mathbf{RF}, \mathbf{SVM}, \mathbf{PBT}\}$, where $\mathbf{RF}$ is a random forest, $\mathbf{SVM}$ is a support vector machine, and $\mathbf{PBT}$ is a probabilistic boosting tree classifier (described in the subsequent Sections). Similarly, notation for a classifier trained in conjunction with different feature vectors is identical to the feature vectors and involves replacing the F with h (e.g. a classifier that leverages the features in $F^{T2w}(c)$ is denoted as $h^{T2w}(c)$). Description of each of the feature vectors evaluated in this work is provided in Table 2.

Methodology

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

C.1 Wavelet feature extraction

a) Haar wavelet features for MRS: The spectral signal $F(c)$ is convolved simultaneously with a high pass (ξ_h) and a low pass filter (ψ_l) to obtain the corresponding high ($\mathcal{H}_h$) and low ($\mathcal{L}_l$) filter coefficients as,

$$\mathcal{H}_h = F(c) * \xi_h, \mathcal{L}_l = F(c) * \psi_l,$$

where $*$ is the convolution operator and dimensionality of coefficients $\mathcal{H}_h$ and $\mathcal{L}_l$ is $M/2$ (M is the dimensionality of each spectrum). Down-sampling or decimation by a factor 2 is performed on coefficients after each pass through filters during the multi-level decomposition. In this work we considered the wavelet packet decomposition (WPD) scheme for extracting MRS wavelet features as suggested in (50), in which both $\mathcal{L}_l$ and $\mathcal{H}_h$ coefficients are iteratively decomposed using high (ξ_h) and low (ψ_l) pass filters, into a full tree like structure of a pre-defined length K , producing a total of 2^K coefficients. The reconstruction of the signal in WPD is then performed by using the best basis algorithm (51) which combines the coefficients that minimize the entropy at each level of the tree. Hence, for each MR spectrum $F(c)$, at each $c \in C$, an M dimensional wavelet feature vector $F^{MRSw}(c)$ is extracted using a Haar wavelet basis function. M varies as a function of the number of coefficients retained by the best basis algorithm, which in turn aims to minimize the entropy for each spectrum.

b) Gab or wavelet features for T2-w MRI: Gabor operators are defined as the convolution of a 2D Gaussian function with a sinusoid (23), such that the cosine component is considered real and the sine component imaginary. Hence, at every metavoxel $c \in C$, a total of 54 Gabor features, $F_k^{T2w}(c)$, $k \in \{1, \dots, 54\}$, are obtained at 9 different scales and 6

orientations similar to that shown in (22) and are represented by a Gabor feature vector $\mathbf{F}^{T2w}(c)$. Further details on the implementation of Gabor texture features for feature extraction can be found in (22).

C.2 Lower dimensional data representation using Principal Component

Analysis—PCA is a linear dimensionality reduction technique (27) which applies a linear transformation to the data to obtain the most uncorrelated features by maximizing the variance within the data. The variance in the data is then expressed in terms of orthogonal Eigenvectors. The Eigenvectors that contain the most variance in the data represent the principal components. At each metavoxel $c \in C$, the high-dimensional MRS wavelet feature vector $\mathbf{F}^{MRSw}(c)$ is reduced to transformed Eigenvector $\mathbf{F}_{PCA}^{MRSw}(c)=[e_1, e_2, e_3, \dots, e_M]$ using PCA, where $[e_1, e_2, \dots, e_M]$ represent the Eigenvectors obtained from Eigenvalue decomposition of the data ranked in order of greatest variance. Thus the first m Eigen vectors that represent a pre-specified percentage of the variance in the data are extracted, while the remaining Eigenvectors are discarded. The high dimensional T2-w MRI Gabor feature vector $\mathbf{F}^{T2w}(c)$ is similarly reduced to a lower dimensional representation, $\mathbf{F}_{PCA}^{T2w}(c)$ using PCA.

C.3 Data combination—Owing to the physical and dimensionality differences between the MR spectra and the T2-w MRI features, the MaWERiC meta-classifier is created in the joint T2-w MRI and MRS Eigen space obtained via PCA. Following the mapping of $\mathbf{F}^{T2w}$ and $\mathbf{F}^{MRSw}$ to reduced dimensional Eigenvector representations, $\mathbf{F}_{PCA}^{T2w}$ and $\mathbf{F}_{PCA}^{MRSw}$, a new concatenated feature vector $\mathbf{F}_{PCA}^{Int}=[\mathbf{F}_{PCA}^{T2w}(c), \mathbf{F}_{PCA}^{MRSw}(c)]$ is obtained.

C.4 Data Classification using a Random Forest classifier—The RF classifier uses the majority voting rule for class assignment by combining decisions from an ensemble of bagged (bootstrapped aggregated) (52) decision trees. The C4.5 decision tree (53) is a multistage classifier which creates a tree like structure by breaking down a complex decision process into a collection of simpler decisions for predicting the best possible outcome solution by combining the simple decisions. RF further combines these decisions obtained from different decision trees to provide a more optimal and stable solution. For a given training set, N bootstrapped subsets are created with replacement of the training data. Based on each training subset, a C4.5 decision tree (53) classifier $\mathbf{h}_j, j \in \{1, \dots, N\}$, is constructed. The class label (CaP or normal) $\mathbf{h}_j(c)$ for each metavoxel $c \in C$, based on the feature vector $\mathbf{F}_{PCA}^{T2w}(c)$, is then obtained using the decision trees $\mathbf{h}_j, j \in \{1, 2, \dots, N\}$; $\mathbf{h}_j(c) = 1$ if c is classified as CaP (scale 4, 5) and $\mathbf{h}_j(c) = 0$ otherwise. The final class likelihood that c belongs to the CaP class, via the RF classifier, is obtained by aggregating the decisions of

individual weak learners as $\frac{1}{N} \sum_{j=1}^N \mathbf{h}_j(c)$. The higher the value of this class likelihood, the more likely c belongs to the CaP class. We define $\mathbf{h}^\rho(c)$ as the binary prediction result for the classifier, at each threshold $\rho \in [0,1]$ such that $\mathbf{h}^\rho(c) = 1$ when $\mathbf{h}(c) \geq \rho$, 0 otherwise.

EXPERIMENTAL DESIGN AND EVALUATION

Comparative Strategies

In the following sub-sections, we evaluate and compare the individual modules (feature extraction, classification, data integration) comprising MaWERiC with (i) other feature extraction schemes (22,54) used in the context of automated CaP detection for individual T2-w MRI, MRS modalities, (ii) a data integration scheme similar to a COD scheme presented in (18) that combines MRS metabolite features with T2-w MRI intensities, (iii) a

decision integration strategy and (iv) two other ensemble classifiers, SVMs (55) and probabilistic boosting trees (PBT) (56).

Comparative feature extraction strategies (T2-w MRI, MRS)

Below, we discuss some of the feature extraction and quantification methods previously proposed in the context of MRS (30,40) and T2-w MRI (3,22,44) and that we have implemented in order to quantitatively compare against MaWERiC. The results of these comparative studies will be described later in the Results Section.

a) Metabolic Peak features for MRS—In the clinic, radiologists typically assess presence of CaP on MRS based on the choline (A_{ch}), creatine (A_{cr}), citrate peaks (A_{cit}) and the A_{ch+cr}/A_{cit} ratio. Variations in these values from predefined normal ranges ($A_{ch+cr}/A_{cit} < 1$) is highly indicative of the presence of the disease (57,58). To compare MaWERiC with metabolic features used clinically, we created a metabolic feature vector for MRS, by calculating area under the choline (A^{ch}), creatine (A^{cr}), citrate (A^{cit}) peaks using the composite trapezoidal rule and recording the corresponding ratios (A_{ch}/A_{cr} , A_{ch+cr}/A_{cit}) (54). Each $c \in C$ is then defined by a metabolite feature vector $\mathbf{F}^{MRS}(c) = [A_{ch}, A_{cr}, A_{cit}, A_{ch+cr}/A_{cit}, A_{ch}/A_{cr}]$.

b) Texture features for T2-w MRI—Below is a brief description of other individual texture features which have previously been explored in conjunction with classifiers for discriminating between CaP and normal areas on T2-w MRI (3,22).

- *Non-steerable Gradient*: Thirteen non-steerable gradient features for each voxel on the T2-w MRI scene were obtained via convolution of the T2-w MRI scene with the Sobel, Kirsch and standard derivative operators at every spatial location (59).
- *First Order Statistical*: A total of 8 first-order statistical features including mean, median, standard deviation, and range of gray scale image intensities within a sliding window neighborhood of 3×3 pixels centered around each spatial location in the T2-w MRI scene were extracted (59).
- *Second Order Statistical*: A total of 13 Haralick features including energy, entropy, inertia, contrast, correlation, sum average, sum variance, sum entropy, difference average, difference variance, difference entropy, local homogeneity and average deviation were extracted within a sliding window neighborhood of 3×3 pixels centered around each voxel in the T2-w MRI scene (60).

For each class of texture features (first order statistical, second order statistical, non-steerable gradient), corresponding T2-w texture feature vectors $\mathbf{F}^{T2\tau_i}$, $i \in \{1,2,3\}$, are constructed at every $c \in C$. A combined ensemble of texture features is defined as $\mathbf{F}^{T2t} = [\mathbf{F}^{T2w}, \mathbf{F}^{T2\tau_1}, \mathbf{F}^{T2\tau_2}, \mathbf{F}^{T2\tau_3}]$ obtained by concatenating all textural attributes obtained from T2-w MRI. PCA was used to reduce each individual texture feature, $\mathbf{F}^{T2\tau_i}$, $i \in \{1, 2, 3\}$ to the corresponding low dimensional representation, $\mathbf{F}_{PCA}^{T2\tau_i}(c)$, and this was then used for classification.

Comparative Data integration strategies

a) Classifier combination (COI)—Classifiers $\mathbf{h}^{T2w}(c)$, $\mathbf{h}^{MRS}(c)$ are individually trained on $\mathbf{F}_{PCA}^{T2w}(c)$ and $\mathbf{F}^{MRS}(c)$, for all $c \in C$. The independence assumption (26) can then be invoked to fuse $\mathbf{h}_{PCA}^{T2w(\rho)}(c)$ and $\mathbf{h}^{MRS(\rho)}(c)$ at each $c \in C$, and at every threshold ρ as $\mathbf{h}_d^{Int}(c) = \mathbf{h}_{PCA}^{T2w(\rho)}(c) \times \mathbf{h}^{MRS(\rho)}(c)$, $\mathbf{h} \in \{\mathbf{RF}, \mathbf{PBT}, \mathbf{SVM}\}$.

b) Data combination (COD) via MRS metabolic area and ratio features and T2-w image intensity—A combined feature vector $F^{Int}(c) = [F^{MRS}(c), F^{T2}(c)]$ is obtained by concatenating the MRS metabolite area and ratio features ($F^{MRS}(c)$) with the mean intensity feature ($F^{T2}(c)$) for each metaboxel $c \in C$. RF classifier along with PBT and SVM classifiers (described in the next section) are then trained using $F^{Int}(c)$ to obtain the meta-classifiers $h^{Int}(c)$, $h \in \{RF, PBT, SVM\}$.

Comparative classifier strategies

a) Probabilistic Boosting Tree (PBT) classifier—The PBT algorithm (56) is a combination of the decision tree (53) and Adaboost (61) classifiers. Adaboost is an ensemble classifier obtained by combining classifier predictions from several weak classifiers. PBT combines decision tree and Adaboost by iteratively generating a tree structure of a predefined length in the training stage where each node of the tree is boosted with τ weak classifiers. The hierarchical tree is obtained by dividing the training samples in two left and right subsets and recursively training the left and right sub-trees using Adaboost (61). During testing, the conditional probability that any $c \in C$ belongs to the CaP class, given the combined MRI-MRS feature vector, $F_{PCA}^{Int}(c)$, is calculated at each node based on the learned hierarchical tree.

b) Support vector machine (SVM) classifier—SVM aims at identifying the “best possible” hyper plane that can accurately separate the data into two classes. SVM classifier (55) is constructed by using a kernel function which projects the training data into a higher-dimensional space via an implicit feature mapping in the dot product space. In our implementation, the radial basis function (RBF) kernel was employed to project the training data into a higher dimensional space. In contrast to PBTs and RFs where a probability (or likelihood) is generated for each voxel belonging to a class, SVM classifiers are typically used to generate a hard decision; $h(c) = 1$ if metaboxel c is identified as CaP and $h(c) = 0$, otherwise. However, a pseudo-likelihood that any meta-voxel c belongs to a class can be generated by calculating how far or close each c is from the SVM decision hyperplane during classification and converting this distance in terms of likelihood of each c belonging to a class. Thus the greater the distance of c from the hyperplane, the higher the likelihood that it belongs to a particular class; the proximity of an object to the hyperplane reflects greater ambiguity with respect to class membership.

Performance Measures

The classification performance of MaWERiC strategy was compared against related state-of-the-art feature extraction, classifier, and data fusion strategies via (a) area under the Receiver Operating Characteristic (ROC) (62) curve (μ^{AUC}), and (b) classification accuracy (μ^{Acc}) at the operating point on the ROC curve. Both performance measures were reported for voxel-level classification.

Classifier Accuracy

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

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

Experimental Setup

Experiment 1: Comparison of MaWERiC against uni-modal classifiers (T2-w MRI, MRS)—MaWERiC was compared against individual feature extraction strategies for T2-w MRI and MRS. Individual features obtained from T2-w MRI and MRS were also quantitatively evaluated against each other to determine the best performing T2-w MRI and MRS features in terms of μ^{AUC} and μ^{Acc} .

Experiment 2: Comparison of MaWERiC against other COD and COI strategies—MaWERiC was compared against current state of the art COD, and COI strategies, involving direct combination of metabolic features with T2-w image intensities and combination of individual classifier predictions respectively for MRI-MRS integration, where binary predictions from the two uni-modal classifiers were combined using a dot product operation to obtain the final classification.

Experiment 3: Comparison of classifiers (RFs against PBTs and SVMs)—Performance of SVMs and PBT classifiers was compared against the RF classifier (employed for MaWERiC), and across other comparative studies (uni-modal T2-w MRI, MRS strategies in Experiment 1 and COD, COI strategies in experiment 2) using μ^{AUC} and μ^{Acc} measures.

RESULTS

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

Qualitative results of classifications obtained from Gabor T2-w MRI ($\mathbf{h}^{T2w}$), Metabolic MRS features ($\mathbf{h}^{MRS}$), COD ($\mathbf{h}^{Int}$), COI ($\mathbf{h}_d^{Int}$), and MaWERiC ($\mathbf{h}_{PCA}^{Int}$) using a RF classifier are shown in Figure 3. Probability heat maps for each strategy were obtained, where the spatial locations shown in red (Figures 3(b)–(f)) were identified as having a higher probability of CaP as determined by classifiers $\mathbf{h}^{T2w}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}_d^{Int}$, and $\mathbf{h}_{PCA}^{Int}$ on a single T2-w slice. Locations shown in blue were identified as having a higher probability of being benign by the classifiers. The white outline in Figure 3(a) shows the ground truth (outlined with a white rectangle) for CaP as annotated by an expert. Note the high CaP detection sensitivity and specificity of MaWERiC (Figure 3(f)) compared to individual uni-modal T2-w MRI (Figure 3(b)) and MRS (Figure 3(c)).

Figure 4(a) shows the AUC results, while Figure 4(b) shows the accuracy results for different feature extraction strategies ($\mathbf{h}^{T2w}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}_d^{Int}$, and $\mathbf{h}_{PCA}^{Int}$) obtained via a RF classifier over 25 runs of cross validation using box-and-whiskers plots respectively. Note that $m = 15$ was used to reduce the dimensionality of T2w MRI, MRS features since it captured ~93% of the variance across MRS and T2w MRI features. Hence, the dimensionality of MaWERiC used for evaluation was $m = 30$. Table 3 shows the quantitative results in terms of AUC and accuracy across various feature extraction and classifier strategies ($\mathbf{h}^{T2w}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}_d^{Int}$, $\mathbf{h}_{PCA}^{Int}$) under evaluation. The μ^{AUC} and μ^{Acc} results shown in Table 3 across 25 iterations of 3-fold cross validation suggest higher CaP detection

accuracy using MaWERiC ($\mu^{AUC} = 0.89 \pm 0.02$, $\mu^{Acc} = 0.83 \pm 0.03$) against both T2-w MRI ($\mu^{AUC} = 0.55 \pm 0.02$, $\mu^{Acc} = 0.54 \pm 0.01$) and MRS ($\mu^{AUC} = 0.77 \pm 0.03$, $\mu^{Acc} = 0.72 \pm 0.02$) for a RF classifier. Note that a higher accuracy for MaWERiC was observed across the other two classifiers (SVM and PBT) as well. Table 4 shows the p -values of paired student t-tests conducted over μ^{AUC} values for comparing statistical significant difference of MaWERiC against all the other comparative feature extraction strategies ($\mathbf{h}^{T2w}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}_d^{Int}$), with the null hypothesis being equal classification performance from MaWERiC when compared to the other feature extraction strategies. Significantly superior performance for MaWERiC ($p < 0.05$) was observed for all pairwise comparisons

($\mathbf{h}_{PCA}^{Int} - \mathbf{h}^{T2w}$, $\mathbf{h}_{PCA}^{Int} - \mathbf{h}^{MRS}$, $\mathbf{h}_{PCA}^{Int} - \mathbf{h}^{Int}$, $\mathbf{h}_{PCA}^{Int} - \mathbf{h}_d^{Int}$). Table 5 shows the individual μ^{AUC} and μ^{Acc} values (obtained across 25 runs of 3-fold cross validation) using each set of texture features (1st order statistical ($\mathbf{h}^{T2r1}$), 2nd order statistical ($\mathbf{h}^{T2r2}$), Gradient ($\mathbf{h}^{T2r3}$) and Gabor ($\mathbf{h}^{T2w}$)), extracted from T2-w MRI across the three sets of classifiers (SVM, RF and PBT). Except in the case of the PBT classifier, Gabor ($\mathbf{h}^{T2w}$) was found to outperform the other first, second-order statistical and gradient texture features ($\mathbf{h}^{T2r1}$, $\mathbf{h}^{T2r2}$, and $\mathbf{h}^{T2r3}$) for both the RF and SVM classifiers.

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

The qualitative results in Figure 3 and box-plots in Figure 4 suggest that MaWERiC ($\mathbf{h}_{PCA}^{Int}$) (Figure 3(f)) yields a higher detection accuracy compared to state-of-the-art COD ($\mathbf{h}^{Int}$) (Figure 3(d)) and COI ($\mathbf{h}_d^{Int}$) (Figure 3(e)) strategies.

Table 3 demonstrates the quantitative results which suggest a significantly higher CaP detection accuracy of MaWERiC, $\mathbf{h}_{PCA}^{Int}$ ($\mu^{AUC} = 0.89 \pm 0.02$, $\mu^{Acc} = 0.83 \pm 0.03$) compared to both COD, $\mathbf{h}^{Int}$ ($\mu^{AUC} = 0.66 \pm 0.02$, $\mu^{Acc} = 0.62 \pm 0.02$), and COI, $\mathbf{h}_d^{Int}$ ($\mu^{AUC} = 0.85 \pm 0.03$, $\mu^{Acc} = 0.78 \pm 0.03$) integration strategies using a RF classifier. MaWERiC results were found to be significantly better than the other comparative feature extraction strategies ($\mathbf{h}^{T2w}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}_d^{Int}$) across the two classifiers (SVM and PBT) as well.

Experiment 3: Comparing PBT against SVM and RF classifiers

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

DISCUSSION

To the best of our knowledge, MaWERiC is the first computerized decision support system (CSS) that provides a systematic framework for quantitative combination of structural information from T2-w MRI (imaging) with metabolic information from MRS (non-imaging) for improved CaP detection. The few COI and COD based data integration techniques previously explored in the literature (18,45) are limited in applicability due to the

ad-hoc approaches employed for overcoming dimensionality differences across modalities. For instance, in (18), Simonetti *et al.* quantitatively combined MRI and MRS by directly concatenating features obtained from the two heterogeneous data sources. However the differing dimensionalities of MRI, MRS features were not accounted for in this study, suggesting that the classifier may have been biased towards the MRS features (8 MRS versus 4 MRI features). Another approach for combining binary decisions, COI, makes an unrealistic assumption of independence across the two data modalities, although complimentary information is acquired simultaneously from the two or more sources about the same disease.

MaWERiC was evaluated on 36 1.5 T *in-vivo* MRS, T2-w MRI patient studies on a per-voxel basis and results thus obtained were compared against 4 other feature extraction strategies, using (i) MRS metabolic features, (ii) T2-w Gabor wavelet features, (iii) a COD scheme involving integration of MRS metabolic features with mean image intensity from T2-w MRI, and (iv) a COI scheme which combined the independent classification results obtained from T2-w MRI and MRS. We also evaluated three classifiers, SVM, PBT and RF, across all the aforementioned 4 strategies (uni-modal MRI, MRS, COD and COI scheme) to identify the best classifier. MaWERiC was found to significantly outperform all the other 4 feature extraction (individual MRS, T2-w MRI) COD and COI strategies, for all 3 classifiers.

To overcome concerns of bias and over-fitting of the data, we iteratively divided 36 patient studies into training and testing sets via a three-fold cross validation scheme. μ^{AUC} and μ^{Acc} values over 25 iteration runs were then reported for all 15 combinations of feature extraction, classifier, and data fusion strategies (see Table 3). In the following subsections, we discuss the detection results of MaWERiC with respect to the other feature extraction, data fusion, and classification strategies considered.

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

MaWERiC was found to significantly outperform a uni-modal classifier trained on Gabor features for T2-w MRI. MaWERiC also outperformed a MRS uni-modal classifier trained on clinically used metabolic MRS features. Our results were consistent with several multi-modal integration studies (45,63–67) which have suggested that combining orthogonal, complementary pieces of information from different modalities can improve classification accuracy compared to uni-modal data channels (5,13–16,68).

Our results demonstrate that MRS metabolite peak area and ratio features yield better classifiers (at a meta-voxel level) compared to a Gabor texture based T2-w MRI classifier. Our findings are consistent with (35) where μ^{AUC} of 0.68 was obtained using T2-w MRI compared to μ^{AUC} of 0.80 obtained using MRS, the metabolic peaks having been identified by visual inspection of 2 expert readers. In a related study (69), MRS ratios of metabolite concentrations ($\mu^{AUC}=0.89$) were shown to outperform visually identified, hypo-intense T2-w MRI features ($\mu^{AUC}=0.85$) for CaP detection on a total of 65 patient studies. Note that in these studies the AUC evaluation was done on a per patient basis, as opposed to a voxel-based evaluation, as in our work. Our findings (Figure 4(a)–(b)) suggest that T2-w MRI texture features alone may not be enough to identify CaP signatures on the prostate. Our findings are also consistent with recent 1.5 T and 3T multi-parametric clinical studies (70–72) which reported sensitivity (at the patient level) in the range of 0.45–0.55 and specificity in the range of 0.80–0.90 from T2-w MRI.

Experiment 2: Comparing MaWERiC against other Data fusion Strategies

MaWERiC versus decision combination (COI)—MaWERiC outperformed a decision level combination scheme (32,45) in terms of μ^{AUC} and μ^{Acc} . The decision level classifier was obtained by combining the binary class decisions (AND operation) from the individual uni-modal classifiers. Decision level integration while helping to overcome the curse of dimensionality (since all the input information is reduced to a scalar valued decision), tends to implicitly treat the data channels as independent. More specifically, in case of T2-w MRI, MRS, data is acquired simultaneously providing complementary (structural and metabolic) information from each spatial location about the same disease. Decision-level fusion strategies may thus be unable to exploit the synergy between these complimentary data streams. By contrast, data level fusion strategies not only exploit the complementary information spread across the different modalities, but are also able to leverage the cross-talk between the data channels (26).

MaWERiC versus data integration using metabolic MRS and MRI intensity features (COD)—The only other work that we are aware of where MRI and MRS features were quantitatively combined at data-level has been for brain tumor detection (18). However, in this approach (18) MRS features (obtained via PCA, ICA and quantification) were directly combined with 4 intensity features from multiprotocol MRI, possibly causing the classifier to be biased towards MRS features. Although MaWERiC was compared only against the best performing COD strategy (quantification + MR intensities), one of 4 presented in (18), our superior results suggested that directly aggregating multi-modal, heterogeneous data from very different sources without accounting for differences in feature dimensionality and relative scaling, can adversely impact classifier performance (26). This is especially true if the constituent classifier features are high dimensional or are unevenly scaled. The superior classifier accuracy of MaWERiC compared to a COD meta-classifier trained using just T2-w MR image intensities and metabolic peak area features (Figures 3(a)–(b)), may be attributable to the uniform scaling and data representation provided by the MaWERiC framework.

Since the high dimensional data could be embedded into a reduced space of arbitrary dimensions, we evaluated MaWERiC across different numbers of Eigenvectors, $m \in \{5, 10, 15, 20\}$; The MaWERiC classifier was found to consistently outperform the COD classifier (18) across different values of m . $m = 15$ was chosen as the number of low dimensional embedding vectors on which to project the high dimensional T2w MRI and MRS features, since it accounts for up to 93% of the variance in the data. Note that no significant differences in μ^{AUC} and μ^{Acc} for the MaWERiC classifier were observable for $m \in \{15, 20, 25, 30\}$, these values accounting for more than 93% of variance in the data. Figure 5 shows the variation in μ^{AUC} (y-axis) and μ^{Acc} (x-axis) of MaWERiC using a random forest classifier across different values of MRS dimensions, from $m = 5$ to 40 ($m = 40$ captures 99.8% variance for MRS), with dimension for T2-w MRI fixed at $m = 15$ (captures 99.8% T2w MRI variance). As can be seen from Figure 5, the highest AUC and accuracy was obtained when dimensions ($m = 15$) were same for both T2-w MRI and MRS. It is important to note here that our choice of number of Eigenvectors was based of maximizing classifier accuracy while using a minimal number of attributes, based on the guiding principle of Occam's razor (73).

Experiment 3: Comparing RF against SVM and PBT classifiers

The three classifiers considered in this study, PBTs, RFs and SVMs are all relatively new, state of the art classifier ensembles that have been shown to be useful in different medical imaging applications (3,29–31,40,74,75). The advantage of these classifier ensembles is that they are able to incorporate information from multiple channels and data sources easily. The

RF classifier was employed as the ensemble of choice within MaWERiC due to its improved and stable performance over SVM and PBT classifiers (Table 2). The RF classifier is known to be able to reduce data variance and hence is able to provide substantial performance improvement over other ensemble classifier strategies (28). RF classifiers have also shown to be relatively more stable across different levels of noise compared to other classifier ensembles (28).

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

Although the results obtained via our MaWERiC data integration scheme significantly outperformed a number of state of the art feature extraction and fusion strategies for MRS and T2-w MRI, we also acknowledge a few limitations of our study: (1) the spectra belonging to scale 3 (identified by the expert as being indeterminate) and voxels identified as atrophic (A) were not considered for classification. We believe that spectra classified as intermediate might provide some clinical insights about the disease specific features, a topic which will be explored in future work. (2) Alternative wavelet-based (apart from Haar and Gabor) and other feature extraction strategies (e.g. independent component analysis (ICA) (30,40)) were not considered. However, our choice of Haar wavelets for MRS and Gabor wavelets for T2-w MRI was motivated by previous demonstrations of their successful employment in building accurate classifiers for CaP detection (22,24). (3) While PCA was employed to obtain a uniform, homogeneous space for representation of the different modalities, newer NLDR methods (33,76,77) have been shown to yield better low dimensional data representations compared to PCA (33). However, these NLDR methods are highly sensitive to the parameter selection and selecting the optimal parameters for two modalities would have been a challenge. (4) Ground truth for evaluation was delineated on a per-MRS voxel by an expert, after considering the disease extent mapped on the radiological imaging from corresponding histopathology. Another way of more robustly and accurately estimating spatial extent of disease on the MRI is by spatially co-registering *ex vivo* whole mount radical prostatectomy sections with corresponding *in vivo* pre-operative MRI. Our group has previously developed elastic registration algorithms for handling deformations between *ex vivo* histology and pre-operative MRI (78). However in this study, this strategy could not be leveraged due to the non-availability of digital pathology resources for digitization of whole mount histology glass slides.

CONCLUDING REMARKS

While data integration schemes for combining image based modalities have been previously presented (2,3,42), analogous methods for combining imaging and non-imaging are not extant in the literature. In this paper, we presented a novel data combination scheme, Multimodal Wavelet Embedding Representation for data Combination, MaWERiC, specifically geared towards quantitative integration of imaging and non-imaging data. MaWERiC comprises of two transformation modules, (i) wavelet transformation and (ii) principal component analysis which together provide a platform for uniform and homogeneous data integration across modalities. The homogeneous, low-dimensional representation of disparate data sources obtained via MaWERiC is then combined in the Eigen space. In this work a random forest classifier ensemble was employed in conjunction with the combined Eigenvector representation of the T2-w MRI and MRS channels to identify prostate cancer *in vivo*. Three-fold cross-validation performed over 25 iterations and

the corresponding pairwise t-test on a total of 36 1.5 Tesla *in-vivo* T2-w MRI, MRS studies demonstrate that the MaWERiC classifier significantly outperforms (a) either modality individually, (b) decision combination obtained by combining individual classifier decisions from both modalities, and (c) a classifier combining metabolite peak area and ratio features from MRS and T2-w MR image intensities.

In conclusion, the MaWERiC data integration framework provides a general framework for potentially integrating any combination of heterogeneous data modalities, independent of scales and dimensions. Future work will look at applying MaWERiC in the context of other biomedical applications such as integration of “-omics” with “imaging” data for improved disease characterization.

Acknowledgments

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

Grant Support: This work was made possible via grants from the Wallace H. Coulter Foundation, New Jersey Commission on Cancer Research, National Cancer Institute (Grant Nos. R01CA136535-01, R01CA140772 01, R21CA127186 01, and R03CA143991-01), The Cancer Institute of New Jersey, and Department of Defense (W81XWH-09).

Abbreviations

MaWERiC	Multimodal wavelet embedding representation for data combination
CSS	Computerized support system
CaP	Prostate cancer
PCA	Principal component analysis
RF	Random forest classifier
COD	combination of data
COI	combination of interpretation
PBT	Probabilistic boosting tree classifier
SVM	Support vector machine classifier

References

1. Madabhushi A, Doyle S, Lee G, Basavanahally A, Monaco J, Masters S, Tomaszewski J, Feldman M. Integrated diagnostics: a conceptual framework with examples. *Clin Chem Lab Med*. 48(7):989–998. [PubMed: 20491597]
2. Hong G, Zhang Y. Comparison and improvement of wavelet-based image fusion. *International Journal of Remote Sensing*. 2008; 29(3):673–691.
3. 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(9):2390–2398. [PubMed: 14528961]
4. Braun V, Dempf S, Tomczak R, Wunderlich A, Weller R, Richter HP. Multimodal cranial neuronavigation: direct integration of functional magnetic resonance imaging and positron emission tomography data: technical note. *Neurosurgery*. 2001; 48(5):1178–1181. discussion 1181–1172. [PubMed: 11334290]

5. Villeirs GM, Oosterlinck W, Vanherreweghe E, De Meerleer GO. A qualitative approach to combined magnetic resonance imaging and spectroscopy in the diagnosis of prostate cancer. *Eur J Radiol.* 73(2):352–356. [PubMed: 19084363]
6. 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]
7. Bonilla J, Stoner E, Grino P, Binkowitz B, Taylor A. Intra- and interobserver variability of MRI prostate volume measurements. *Prostate.* 1997; 31(2):98–102. [PubMed: 9140122]
8. Vos P, 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.* 55(6):1719. [PubMed: 20197602]
9. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics. *CA Cancer J Clin.* 2010; 60(5):277–300. [PubMed: 20610543]
10. Borboroglu PG, Comer SW, Riffenburgh RH, Amling CL. Extensive Repeat Transrectal Ultrasound Guided Prostate biopsy in Patients with Previous Benign Sextant Biopsies. *The Journal of Urology.* 2000; 163(1):158–162. [PubMed: 10604336]
11. Schiebler ML, Schnall MD, Pollack HM, Lenkinski RE, Tomaszewski JE, Wein AJ, Whittington R, Rauschnig W, Kressel HY. Current role of MR imaging in the staging of adenocarcinoma of the prostate. *Radiology.* 1993; 189(2):339–352. [PubMed: 8210358]
12. Scheidler J, Hricak H, Vigneron DB, Yu KK, Sokolov DL, Huang LR, Zaloudek CJ, Nelson SJ, Carroll PR, Kurhanewicz J. Prostate cancer: localization with three-dimensional proton MR spectroscopic imaging--clinicopathologic study. *Radiology.* 1999; 213(2):473–480. [PubMed: 10551229]
13. Swanson MG, Vigneron DB, Tran TK, Kurhanewicz J. Magnetic resonance imaging and spectroscopic imaging of prostate cancer. *Cancer Invest.* 2001; 19(5):510–523. [PubMed: 11458818]
14. 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 Int.* 2007; 99(4):786–793. [PubMed: 17223922]
15. Yu KK, Scheidler J, Hricak H, Vigneron DB, Zaloudek CJ, Males RG, Nelson SJ, Carroll PR, Kurhanewicz J. Prostate cancer: prediction of extracapsular extension with endorectal MR imaging and three-dimensional proton MR spectroscopic imaging. *Radiology.* 1999; 213(2):481–488. [PubMed: 10551230]
16. Wang L, Hricak H, Kattan MW, Chen HN, Scardino PT, Kuroiwa K. Prediction of organ-confined prostate cancer: incremental value of MR imaging and MR spectroscopic imaging to staging nomograms. *Radiology.* 2006; 238(2):597–603. [PubMed: 16344335]
17. Luts J, Laudadio T, Idema AJ, Simonetti AW, Heerschap A, Vandermeulen D, Suykens JA, Van Huffel S. Nosologic imaging of the brain: segmentation and classification using MRI and MRSI. *NMR Biomed.* 2009; 22(4):374–390. [PubMed: 19105242]
18. Simonetti A, Melssen W, Edelenyi F, van Asten J, Heerschap A, Buydens LMC. Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification. *NMR in Biomedicine.* 2005; 18(1):34–43. [PubMed: 15657908]
19. Mallat SG. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 1989; 11:674–693.
20. Jun, Z.; Regentova, E. Wavelet based feature reduction method for effective classification of hyperspectral data. *Information Technology: Coding and Computing [Computers and Communications], 2003 Proceedings ITCC 2003 International Conference on 2003;* p. 483-487.
21. Sebe, N.; Lew, MS. Wavelet based texture classification. *Pattern Recognition, 2000 Proceedings 15th International Conference on 2000;* p. 947-950.p. 943
22. Madabhushi A, Feldman MD, Metaxas DN, Tomaszewski J, Chute D. Automated detection of prostatic adenocarcinoma from high-resolution ex vivo MRI. *Medical Imaging, IEEE Transactions on.* 2005; 24(12):1611–1625.

23. Gabor D. Theory of communication. Part 1: The analysis of information Electrical Engineers -Part III: Radio and Communication Engineering. Journal of the Institution of. 1946; 93(26):429–441.
24. Mira, J.; Álvarez, J.; Arévalo, Acosta O.; Santos Peñas, M. Nature Inspired Problem-Solving Methods in Knowledge Engineering. Vol. 4528. Springer; Berlin / Heidelberg; 2007. Classification of Biomedical Signals Using a Haar 4 Wavelet Transform and a Hamming Neural Network; p. 637-646. Lecture Notes in Computer Science
25. Subramani P, Sahu R, Verma S. Feature selection using Haar wavelet power spectrum. BMC Bioinformatics. 2006; 7(1):432. [PubMed: 17022808]
26. Duda, R.; Hart, P.; Stork, D. Pattern Classification. 2. Wiley-Interscience; 2000. Report nr 0471056693
27. Hotelling H. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology. 1933; 24:417–441.
28. Breiman L. Random Forests. Machine Learning. 2001; 45(1):5–32.
29. Diaz-Uriarte R, Alvarez de Andres S. Gene selection and classification of microarray data using random forest. BMC Bioinformatics. 2006; 7:3. [PubMed: 16398926]
30. 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]
31. Li S, Fedorowicz A, Singh H, Soderholm SC. Application of the random forest method in studies of local lymph node assay based skin sensitization data. J Chem Inf Model. 2005; 45(4):952–964. [PubMed: 16045289]
32. Rohlfing, T.; Pfefferbaum, A.; Sullivan, EV.; Maurer, CR. Information Processing in Medical Imaging. 2005. Information Fusion in Biomedical Image Analysis: Combination of Data vs. Combination of Interpretations; p. 150-161.
33. 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]
34. Cohen MS, DuBois RM, Zeineh MM. Rapid and effective correction of RF inhomogeneity for high field magnetic resonance imaging. Hum Brain Mapp. 2000; 10(4):204–211. [PubMed: 10949058]
35. Futterer JJ, Heijmink SW, Scheenen TW, Veltman J, Huisman HJ, Vos P, Hulsbergen-Van de Kaa CA, Witjes JA, Krabbe PF, Heerschap A, Barentsz JO. Prostate cancer localization with dynamic contrast-enhanced MR imaging and proton MR spectroscopic imaging. Radiology. 2006; 241(2): 449–458. [PubMed: 16966484]
36. van der Veen JW, de Beer R, Luyten PR, van Ormondt D. Accurate quantification of in vivo ³¹P NMR signals using the variable projection method and prior knowledge. Magn Reson Med. 1988; 6(1):92–98. [PubMed: 3352510]
37. Vanhamme L, van den Boogaart A, Van Huffel S. Improved method for accurate and efficient quantification of MRS data with use of prior knowledge. J Magn Reson. 1997; 129(1):35–43. [PubMed: 9405214]
38. Ratiney H, Sdika M, Coenradie Y, Cavassila S, van Ormondt D, Graveron-Demilly D. Time-domain semi-parametric estimation based on a metabolite basis set. NMR Biomed. 2005; 18(1):1–13. [PubMed: 15660450]
39. Ma, J.; Sun, Z.; Dong, G.; Xie, G. Advances in Neural Networks – ISNN 2005. 2005. Wavelet Denoise on MRS Data Based on ICA and PCA; p. 748-753.
40. Luts J, Pouillet JB, Garcia-Gomez JM, Heerschap A, Robles M, Suykens JA, Van Huffel S. Effect of feature extraction for brain tumor classification based on short echo time 1H MR spectra. Magnetic Resonance in Medicine. 2008; 60(2):288–298. [PubMed: 18666120]
41. Comon P. Independent Component Analysis, a New Concept. Signal Processing. 1994; 36(3):287–314.
42. 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 Trans Med Imaging. 2009; 28(6):906–915. [PubMed: 19164079]

43. 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. *March 12–14.2008* 2008;1372–1377.
44. Pieter C, Vos TH, Barenstz Jelle O, Huisman Henkjan J. Computer-assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI. *Physics in Medicine and Biology*. 55(6):1719. [PubMed: 20197602]
45. Jesneck JL, Nolte LW, Baker JA, Floyd CE, Lo JY. Optimized approach to decision fusion of heterogeneous data for breast cancer diagnosis. *Med Phys*. 2006; 33(8):2945–2954. [PubMed: 16964873]
46. Veeramachaneni K, Osadciw LA, Varshney PK. An adaptive multimodal biometric management algorithm. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*. 2005; 35(3):344–356.
47. Star-Lack J, Nelson SJ, Kurhanewicz J, Huang LR, Vigneron DB. Improved water and lipid suppression for 3D PRESS CSI using RF band selective inversion with gradient dephasing (BASING). *Magn Reson Med*. 1997; 38(2):311–321. [PubMed: 9256113]
48. Tran TK, Vigneron DB, Sailasuta N, Tropp J, Le Roux P, Kurhanewicz J, Nelson S, Hurd R. Very selective suppression pulses for clinical MRSI studies of brain and prostate cancer. *Magn Reson Med*. 2000; 43(1):23–33. [PubMed: 10642728]
49. 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(3):701–708. [PubMed: 15564406]
50. Mainardi LT, Origgi D, Lucia P, Scotti G, Cerutti S. A wavelet packets decomposition algorithm for quantification of in vivo (1)H-MRS parameters. *Med Eng Phys*. 2002; 24(3):201–208. [PubMed: 12062178]
51. Coifman RR, Wickerhauser MV. Entropy-based algorithms for best basis selection. *Information Theory, IEEE Transactions on*. 1992; 38(2):713–718.
52. Prinzie A, Van den Poel D. Random forests for multiclass classification: Random multinomial logit. *Expert systems with Applications*. 2008; 34(3):1721–1732.
53. Quinlan, J. C4. 5: programs for machine learning. Morgan Kaufmann; 1993.
54. Devos A, Lukas L, Suykens JA, Vanhamme L, Tate AR, Howe FA, Majos C, Moreno-Torres A, van der Graaf M, Arus C, Van Huffel S. Classification of brain tumours using short echo time 1H MR spectra. *J Magn Reson*. 2004; 170(1):164–175. [PubMed: 15324770]
55. Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995; 20(3):273–297.
56. Zhuowen, T. Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. *Computer Vision, 2005 ICCV 2005 Tenth IEEE International Conference on 2005*; p. 1589-1596.
57. Heerschap A, Jager GJ, van der Graaf M, Barentsz JO, de la Rosette JJ, Oosterhof GO, Ruijter ET, Ruijs SH. In vivo proton MR spectroscopy reveals altered metabolite content in malignant prostate tissue. *Anticancer Res*. 1997; 17(3A):1455–1460. [PubMed: 9179183]
58. Kurhanewicz J, Vigneron DB, Hricak H, Narayan P, Carroll P, Nelson SJ. Three-dimensional H-1 MR spectroscopic imaging of the in situ human prostate with high (0.24–0. 7-cm3) spatial resolution. *Radiology*. 1996; 198(3):795–805. [PubMed: 8628874]
59. Russ, JC. *Image Processing Handbook*. 5. CRC Press, Inc; 2006. The Image Processing Handbook.
60. Haralick RM, Dinstein, Shanmugam K. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*. 1973; SMC-3:610–621.
61. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*. 1997; 55(1):119–139.
62. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem*. 1993; 39(4):561–577. [PubMed: 8472349]
63. Polikar R. Ensemble Based Systems in Decision Making. *IEEE Circuits and Systems Magazine*. 2006; 6(3):21–45.
64. Zonari P, Baraldi P, Crisi G. Multimodal MRI in the characterization of glial neoplasms: the combined role of single-voxel MR spectroscopy, diffusion imaging and echo-planar perfusion imaging. *Neuroradiology*. 2007; 49(10):795–803. [PubMed: 17619871]

65. Lanckriet GR, De Bie T, Cristianini N, Jordan MI, Noble WS. A statistical framework for genomic data fusion. *Bioinformatics*. 2004; 20(16):2626–2635. [PubMed: 15130933]
66. Breiman L. Bagging predictors. *Mach Learn*. 1996; 24(2):123–140.
67. Dietterich T. Ensemble Methods in Machine Learning. *Multiple Classifier Systems*. 2000:1–15.
68. John K, Mark GS, Sarah JN, Daniel BV. Combined magnetic resonance imaging and spectroscopic imaging approach to molecular imaging of prostate cancer. *J Magn Reson Imaging*. 2002; 16(4): 451–463. [PubMed: 12353259]
69. Squillaci E, Manenti G, Mancino S, Carlini M, Di Roma M, Colangelo V, Simonetti G. MR spectroscopy of prostate cancer. Initial clinical experience. *J Exp Clin Cancer Res*. 2005; 24(4): 523–530. [PubMed: 16471314]
70. Kurhanewicz J, Vigneron D, Carroll P, Coakley F. Multiparametric magnetic resonance imaging in prostate cancer: present and future. *Curr Opin Urol*. 2008; 18(1):71–77. [PubMed: 18090494]
71. Turkbey B, Pinto PA, Mani H, Bernardo M, Pang Y, McKinney YL, Khurana K, Ravizzini GC, Albert PS, Merino MJ, Choyke PL. Prostate Cancer: Value of Multiparametric MR Imaging at 3 T for Detection—Histopathologic Correlation. *Radiology*. 2015; 255(1):89–99. [PubMed: 20308447]
72. Kim CK, Park BK, Kim B. Localization of prostate cancer using 3T MRI: comparison of T2-weighted and dynamic contrast-enhanced imaging. *J Comput Assist Tomogr*. 2006; 30(1):7–11. [PubMed: 16365565]
73. Mohak S. Feature Selection with Conjunctions of Decision Stumps and Learning from Microarray Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. :99. (PrePrints).
74. Lee SL, Kouzani AZ, Hu EJ. Random forest based lung nodule classification aided by clustering. *Comput Med Imaging Graph*. 34(7):535–542. [PubMed: 20430583]
75. Carneiro G, Georgescu B, Good S, Comaniciu D. Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. *IEEE Trans Med Imaging*. 2008; 27(9):1342–1355. [PubMed: 18753047]
76. Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science*. 2000; 290(5500):2319–2323. [PubMed: 11125149]
77. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science*. 2000; 290(5500):2323–2326. [PubMed: 11125150]
78. Chappelow J, Bloch BN, 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*. 38(4):2005–2018. [PubMed: 21626933]

Figure 1.

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

Figure 2.

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

Figure 3.

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

Figure 4.

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

Figure 5.

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

Table 1

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

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

Table 2

Description of different feature notations and the associated dimensionality of each feature vector evaluated in this work.

Notation	Description	Number of Dimensions
F^{T2w}	Gabor feature vector for T2w MRI	34
F^{T2t1}	non-steerable Gradient T2w MRI feature vector	13
F^{T2t2}	first order statistical T2w MRI feature vector	8
F^{T2t3}	Haralick T2w MRI feature vector	13
F^{T2t}	Combined ensemble of T2w MRI feature vector $F^{T2t} = [F^{T2w}, F^{T2t1}, F^{T2t2}, F^{T2t3}]$	68
F^{MRS}	MRS metabolite and ratio feature vector $F^{MRS}(c) = [A_{chl}, A_{cr}, A_{clb}, A_{chl+cr}/A_{clb}, A_{chl}/A_{cr}]$	5
F^{MRSw}	MRS wavelet feature vector	Variable (depends on the coefficients chosen to maximize energy)
F_{PCA}^{T2w}	Reduced dimensional T2w MRI Eigen feature vector obtained via PCA on F^{T2w}	15
F_{PCA}^{MRSw}	Reduced dimensional MRS Eigen feature vector obtained via PCA on F^{MRSw}	15
F^{Int}	Combined feature vector obtained by concatenating MRS metabolite area and ratio features with mean MRI intensity $F^{Int}(c) = [F^{MRS}(c), F^{T2}(c)]$	6
F_{PCA}^{Int}	Wavelet PCA based MaWERiC feature vector	30

Table 3

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

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

Table 4

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

Method	MRS	T2-w MRI	COI	COD
MaWERiC	1.60×10^{-12}	4.06×10^{-24}	0.0017	5.62×10^{-12}

Table 5

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

Classifier	Feature	Mean AUC	Mean accuracy
PBT	Gabor	0.549 ± 0.019	0.545 ± 0.017
	Gradient	0.560 ± 0.018	0.556 ± 0.012
	1 st order statistical	0.566 ± 0.013	0.554 ± 0.016
	2 nd order statistical	0.534 ± 0.016	0.538 ± 0.021
RF	Gabor	0.554 ± 0.112	0.546 ± 0.014
	Gradient	0.547 ± 0.115	0.545 ± 0.014
	1 st order statistical	0.534 ± 0.016	0.530 ± 0.017
	2 nd order statistical	0.544 ± 0.011	0.537 ± 0.013
SVM	Gabor	0.513 ± 0.030	0.659 ± 0.013
	Gradient	0.495 ± 0.038	0.499 ± 0.029
	1 st order statistical	0.493 ± 0.026	0.498 ± 0.027
	2 nd order statistical	0.504 ± 0.041	0.511 ± 0.033