Multimodal wavelet embedding representation for data combination (MaWERiC): integrating magnetic resonance imaging and spectroscopy for prostate cancer detection

P. Tiwari[a], S. Viswanath[a], J. Kurhanewicz[b], A. Sridhar[a] and A. Madabhushi[a]*

Recently, both Magnetic Resonance (MR) Imaging (MRI) and Spectroscopy (MRS) have emerged as promising tools for detection of prostate cancer (CaP). However, due to the inherent dimensionality differences in MR imaging and spectral information, quantitative integration of T_2 weighted MRI (T_2w MRI) and MRS for improved CaP detection has been a major challenge. In this paper, we present a novel computerized decision support system called multimodal wavelet embedding representation for data combination (MaWERiC) that employs, (i) wavelet theory to extract 171 Haar wavelet features from MRS and 54 Gabor features from T_2w MRI, (ii) dimensionality reduction to individually project wavelet features from MRS and T_2w MRI into a common reduced Eigen vector space, and (iii), a random forest classifier for automated prostate cancer detection on a per voxel basis from combined 1.5 T *in vivo* MRI and MRS. A total of 36 1.5 T endorectal *in vivo* T_2w MRI and MRS patient studies were evaluated per voxel by MaWERiC using a three-fold cross validation approach over 25 iterations. Ground truth for evaluation of results was obtained by an expert radiologist annotations of prostate cancer on a per voxel basis who compared each MRI section with corresponding *ex vivo* wholemount histology sections with the disease extent mapped out on histology. Results suggest that MaWERiC based MRS T_2w meta-classifier (mean AUC, $\mu = 0.89 \pm 0.02$) significantly outperformed (i) a T_2w MRI (using wavelet texture features) classifier ($\mu = 0.55 \pm 0.02$), (ii) a MRS (using metabolite ratios) classifier ($\mu = 0.77 \pm 0.03$), (iii) a decision fusion classifier obtained by combining individual T_2w MRI and MRS classifier outputs ($\mu = 0.85 \pm 0.03$), and (iv) a data combination method involving a combination of metabolic MRS and MR signal intensity features ($\mu = 0.66 \pm 0.02$). Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: multimodal integration; magnetic resonance imaging; magnetic resonance spectroscopy; Haar wavelets; Gabor texture features; PCA, random forest classifier; prostate cancer

INTRODUCTION

In the biomedical field, there is a genuine need for developing data integration strategies for combining discriminatory features from multiple modalities to develop multimodal meta-classifiers for improved disease detection, diagnosis, and prognosis (1). While data integration strategies have been proposed for quantitatively combining multiple imaging modalities (2–4), these tools are not readily applicable to integration of imaging and non-imaging data because of differences in scale and feature dimensionality. For instance, consider the difficulties in quantitatively combining at the voxel level T_2-weighted (T_2w) 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 encodes a different type (structural or metabolic) and dimensionality of information. Nevertheless, 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 T_2w MRI and MRS involves visually identifying hypointense regions on T_2w MRI followed by inspection of spectrum at corresponding spatial locations to identify changes

in metabolite (choline, creatine citrate) ratios (5). However, some studies have shown that manual interpretation and visual integration of multimodal data is subjective and thus prone to inter-and intra-observer variability (6,7). It is therefore preferable to build

* Correspondence to: Anant Madabhushi, Department of Biomedical Engineering, Rutgers, The State University of New Jersey, 599 Taylor Road, Piscataway, NJ, USA.
E-mail: anantm@rci.rutgers.edu

a P. Tiwari, S. Viswanath, A. Sridhar, A. Madabhushi
Department of Biomedical Engineering, Rutgers University, 599 Taylor Road, Piscataway, NJ, USA

b J. Kurhanewicz
University of California, Department of Radiology and Biomedical Imaging, San Francisco, CA, USA

Abbreviation used: *CaP, prostate cancer; COD, combination of data; COI, combination of interpretation; CSS, computerized support system; MaWERiC, multimodal wavelet embedding representation for data combination; PBT, probabilistic boosting tree classifier; PCA, principal component analysis; RF, random forest classifier; SVM, support vector machine classifier.*

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 major challenge 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 study, 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 differences in scales and dimensionalities. Additionally, MaWERiC provides the ability to build meta-classifiers from different modalities once the data representation and integration issues are resolved. While MaWERiC could be applied to a number of different domains and applications, in this study, we demonstrate its applicability in building meta-classifiers for improved prostate cancer detection using T_2w MRI and MRS.

Prostate cancer (CaP) is the second leading cause of cancer related deaths among men, with 217 730 new cases and 32 050 estimated deaths in the USA in 2010 (9). The current gold standard for CaP detection is a sextant transrectal ultrasound examination that is known to have a low detection sensitivity (20–25%) due to poor image resolution of the ultrasound (10). In the last decade, T_2w MRI has shown great potential for characterizing disease presence and staging of CaP (11). MRS has recently emerged as a complement to traditional T_2w 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 on individual modality (5,13–16), few attempts have been made to quantitatively combine the different information channels (17,18). While some studies have proposed CSS classifiers for combining multiple MR imaging protocols (T_2-w, dynamic contrast enhanced (DCE), line-scan diffusion, diffusion weighted imaging, T_2-mapping MRI) (3,8), to our knowledge, no previous strategies for quantitative integration of T_2w MRI and spectroscopy (imaging and non-imaging modalities) for CaP detection have been suggested.

In this study, we aimed to leverage MaWERiC to build an integrated structural metabolic meta-classifier that would assign a probability of CaP presence at every spatial location on *in-vivo* prostate T_2-w-MRI/MRS studies. Figure 1 illustrates the organization of our MaWERiC strategy. Our approach was based on combining wavelet (Gabor and Haar filters) features extracted from both T_2w MRI and MRS modalities (19). The advantage of using wavelet transform (19) is that it provides 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 multiscale and multi orientation texture characterization and representation of an image. Haar wavelet decomposition is a commonly used signal filtering technique that provides a way of extracting the class discriminating frequency components that could yield higher classifier accuracy compared to the original signal (19,24). An advantage of the

Haar wavelet 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 (24). Both 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 study, multiresolution features for T_2w MRI and MRS were obtained using Gabor and Haar wavelet filters.

While the wavelet based representations of MRS and T_2w MRI channels provide a uniform feature representation of the data, the feature vectors obtained by the application of Gabor and Haar wavelet filters are of very high dimensionality and are therefore subject to the curse of dimensionality (26). As a result, a subsequent but necessary step is the application of principal component analysis (PCA) to the high dimensional feature vectors obtained (by wavelet decomposition) from each individual data modality (T_2w MRI, MRS) to obtain a reduced dimensional representation of the data and make the data representation amenable to the application of the classifier (27). The representation of MRS and T_2w MRI data in terms of eigenvectors (obtained by PCA) also allows for 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 been successfully used in various biomedical classification applications (29–31). Advantages of RF include: (i) the ability to integrate a large number of input variables, (ii) robustness to noise in the data, and (iii) relatively few tuning parameters.

PREVIOUS WORK

Quantitative data integration strategies 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 (combination of data (COD)) involves combining the data prior to classification. The second class, known as decision level classification (combination of interpretation (COI)) involves training individual data classifiers (unimodal classification) and combining the outputs from each classifier (32). In the context of COD strategies, data can be combined either in its original acquired form or following the application of an appropriate transformation to each data modality to obtain improved discriminatory information beyond the original acquired data. MaWERiC is a COD approach 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 multimodal MRI.

In subsequent sections, we discuss both individual unimodal classifiers for T_2w MRI (22), MRS (33) and multimodal meta-classifiers for CaP detection using multiparametric MRI. We also briefly discuss previous combination strategies involving MRI-MRS for brain tumor classification (17,18).
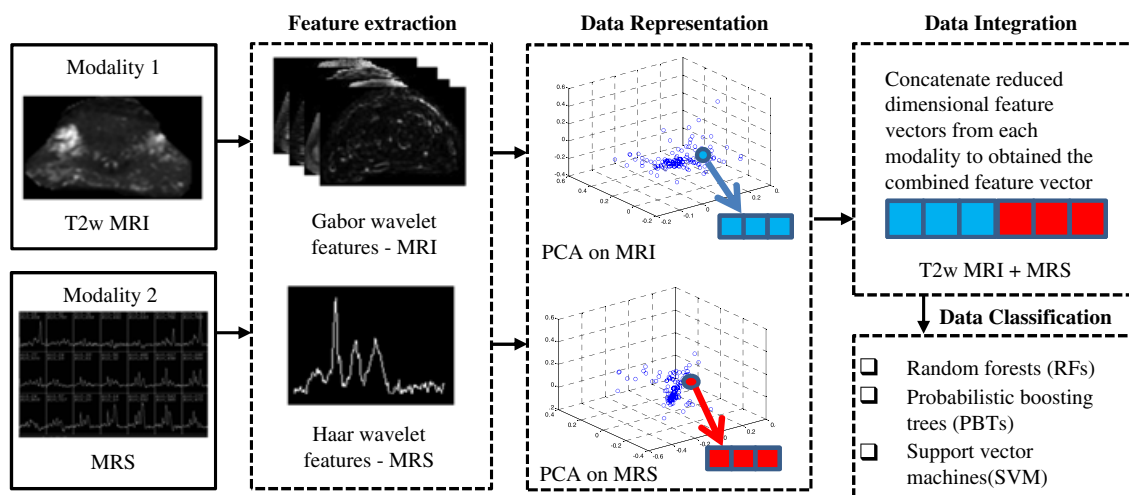
Figure 1. Flowchart showing various components and methodological overview of our data integration method. Wavelet features were first extracted individually from T$_2$w MRI and MRS, followed by dimensionality reduction using PCA. The reduced dimensional vectors, now existing in the same dimensions and scale, were concatenated during data combination followed by data classification.

Unimodal classifiers for T$_2$w MRI

Since image intensity on T$_2$w MRI is susceptible to artifacts such as bias field inhomogeneity (34) and intensity nonstandardness (35), researchers have explored alternate representations of T$_2$w image intensities (e.g., Gabor or wavelet based texture features) to build classifiers for predicting CaP presence on MRI (22). Madabhushi *et al.* presented a supervised CSS system for detection of CaP from four Tesla *ex vivo* prostate T$_2$w MRI, where 33 3D texture features (statistical, gradient, Gabor) were quantitatively extracted at each voxel (T$_2$w MRI spatial resolution) (22). These extracted features were then used to train a number of supervised classifiers (AdaBoost, Bayes, decision trees) which were employed to assign a probability of CaP presence at each image voxel.

Unimodal classifiers for MRS

Previous CSS strategies developed in the context of MRS data can be broadly divided into two main categories: signal quantification (model dependent) (36–38) and statistical pattern recognition based (model independent) strategies (30,33,39,40). Commonly used MRS quantification strategies include VARPRO (36), AMARES (37) and QUEST (38), which are software utilities, where the objective is to minimize the squared distance between acquired data and a model basis function built on prior information about the metabolic profiles of a typical MR spectrum. Conversely, pattern recognition based features try to capture the underlying variance in the data using regression analysis. Kelm *et al.* (30) compared classification techniques for prostate MRS data based on pattern recognition strategies such as PCA (27) and independent component analysis (ICA) (41) with quantification based feature extraction strategies using SVM, RF and Gaussian process classifiers. They showed that pattern recognition based classifiers provided better classification results for CaP detection compared to MRS quantification strategies. 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 method in conjunction with non-linear dimensionality reduction (NLDR)

strategies (33). NLDR strategies were used 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 for eighteen 1.5 T prostate MRS studies. Luts *et al.* presented a method that leveraged ICA and Relief-F in conjunction with SVM and linear discriminant analysis classifiers for brain tumor classification using MRS (40).

Combining imaging-imaging MRI channels

Chan *et al.* presented a statistical classifier that integrated texture features from multi-protocol 1.5 T *in vivo* MRI to generate a statistical probability map representing likelihoods of cancer in different regions of the prostate (3). Liu *et al.* examined multiparametric *in vivo* MRI maps (T$_2$-w, DCE, DWI) within a fuzzy Markov random fields framework (42). The maps were generated by curve fitting data from each protocol with ROI limited to the peripheral zone of the prostate, while evaluation of the results was performed in manually delineated CaP regions on MRI (with corresponding whole-mount histology and *ex vivo* MRI data used for reference). Ampeliotis *et al.* explored the use of image intensity features from both DCE and T$_2$w MRI data for the classification of CaP (43). A statistically significant improvement in classifier performance was reported when fusing modalities compared to using of individual modalities. Another multiprotocol (DCE and T$_2$-w) MRI based CSS has been presented that combines pharmacokinetic features from DCE with T$_2$w image intensities (44).

Decision integration strategies for MRI and MRS

To our knowledge, no data or decision integration strategies for combining imaging and spectroscopy for use in CaP detection have been reported. Jesneck *et al.* proposed a decision integration approach 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 (45). Another COI approach has been suggested that combines classifier outputs

from three heterogeneous modalities: face recognition, voice recognition, and hand geometry within a Bayesian framework for improved biometric based personnel identification (46). However, a major disadvantage of such decision integration based strategies is that all inter-source dependencies between modalities could be lost since each modality is treated independently (26).

Data integration strategies for MRI and MRS

A data integration approach involving integration of multiprotocol MR image intensities (T$_1$-w, T$_2$-w, proton density-weighted, gadolinium-DTPA) with the areas under spectral peaks of specific metabolites (*myo*-inositol, glucose, choline, creatine, glutamate/glutamine, *N*-acetyl aspartate, lactate/fatty acids, fatty acids) from MRS has been presented for classifying four brain tumor types (Types II, III, IV and meningioma), healthy tissues, and cerebrospinal fluid (CSF) (18). MRS and MRI features were directly concatenated into a single joint feature vector and used in conjunction with a Mahalanobis distance based classifier. Classifier results showed that the voxel level classification obtained by this multimodal feature combination was significantly superior compared to the results obtained using unimodal classifiers.

In this study, we aimed to show that MaWERiC provides significant improvements over previous COD and COI strategies for prostate cancer detection using multimodal MRI (18,45). We showed the utility of MaWERiC for developing the first CSS method that quantitatively combines T$_2$w MRI and MRS for CaP detection and demonstrates significant improvements over using individual modalities and other previous state-of-the-art data fusion strategies (18,45).

MATERIALS AND METHODS

Data

Thirty-six 1.5 Tesla (T) T$_2$w MRI and MRS patient studies prior to radical prostatectomy were reviewed. Studies were of biopsy proven prostate cancer patients that were clinically referred for a prostate cancer MR staging exam for improved therapeutic selection. MR imaging was performed using a 1.5-T whole-body MR imaging unit (Sigma; GE Medical Systems, Milwaukee, WI, USA). Patients were imaged in supine position using a body coil for signal excitation and a pelvic phased-array coil (GE Medical Systems, Milwaukee, WI, USA) combined with a balloon-covered expandable endorectal coil (Medrad, Pittsburgh, PA, USA) for signal reception. 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 periprostatic fat and rectal air. 3D proton (1-H) MR spectroscopic imaging data were acquired using a water and lipid suppressed double spin-echo point resolved spectroscopic sequence optimized for quantitative detection of both choline and citrate. Water and lipid suppression were achieved by using the band selective inversion with a gradient dephasing technique (47). To eliminate signals from adjacent tissues, especially periprostatic lipids and the rectal wall, outer voxel saturation pulses were also used (48). Data sets were acquired as $16 \times 8 \times 8$ phase-encoded spectral arrays (1024 voxels) using a nominal spectral resolution of 0.24–0.34 cm^3, 1000/130, and a 17-min acquisition time.
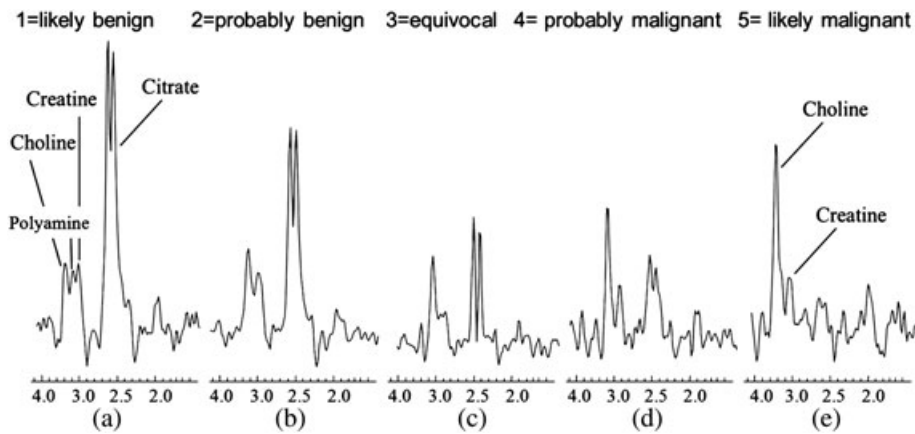
Preprocessing

3D MR spectroscopic imaging data were processed and aligned with the corresponding T$_2$w imaging data using a combination of inhouse software interactive display language (Research Systems Inc., Boulder, Colorado, USA) software tools (48). The raw spectral data were apodised 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, line width) were estimated 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 aligned the spectral peaks with the expected locations of choline, creatine, citrate, and residual water was then estimated. Subsequently, the spectra were phased using the phase of the residual water and metabolite resonances. Baseline values were corrected using a local nonlinear fit to the non-peak regions of the spectra. Subsequent feature extraction and classification steps were performed using algorithms developed with the MATLAB (MathWorks Inc., Natick, MA, USA) programming environment.

Ground truth annotations

For all experiments considered in this study, *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 by visually registering corresponding histological and radiological sections; correspondence between sections was determined manually by visually determining anatomical fiducials on the histology and 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*. that was used to visually classify each spectrum as being: (a) scale 1, definitely benign; (b) scale 2, probably benign; (c) scale 3, equivocal- (d) scale 4, probably cancer; and (e) scale 5, definitely cancer. All spectra labeled as 4 or 5 were assumed to be CaP and spectra labeled as 1 or 2 were assumed as benign (49). The voxels labeled as three and atrophic (a) were assumed to be indeterminate, and consequently excluded from our analysis. The 36 studies comprised 2120 class 1 and 2 spectra and 1026 class 4 and 5 spectra (Table 1). The class labels for the individual spectral voxels assigned by 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. The ground truth surrogate was then used for training and evaluating the MaWERiC classifiers.

Notation

We defined a metavoxel in the MRS grid as: $c \in C$, where C is a 3D grid of MRS metavoxels. For each $c \in C$, $\mathbf{F}(c) = [f\alpha(c)|\alpha \in \{1, \ldots M\}]$ represents the MR spectral vector, reflecting the frequency component of each of the M metabolites being measured (33). Note that the MRS metavoxel and T$_2$w MRI voxel are at different resolutions, where 1 MRS metavoxel corresponds to approximately 90 T$_2$w MRI voxels. Feature extraction from T$_2$w MRI was performed on a per voxel (T$_2$w resolution) basis (3,22). The responses of various texture filters (described in the next section) were averaged over all voxels within each metavoxel $c \in C$. The

1=likely benign   2=probably benign   3=equivocal   4= probably malignant   5= likely malignant

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

Table 1. Number of spectra and patients for each scale as annotated by an 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

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: $h(c), h \in \{RF, SVM, PBT\}$, where **RF** is a random forest, **SVM** is a support vector machine and **PBT** is a probabilistic boosting tree classifier (described in subsequent sections). Similarly, notation for a classifier trained in conjunction with different feature vectors was identical to the feature vectors and involved replacing the **F** with **h** (e.g., a classifier that leverages the features in $F^{T_2w}(c)$ is denoted as $h^{T_2w}(c)$. Description of each of the feature vectors evaluated in this study is provided in Table 2.

T$_2$w MRI feature vector was obtained by calculating a mean feature vector at each $c \in C$. As a result, at every metavoxel, $c \in C$, the corresponding intensity feature vector is denoted as $F^{T_2}(c)$ while the corresponding mean Gabor wavelet feature vector (details in the next section) is denoted as $F^{T_2w}(c)$. For MRS, the feature vector comprised of ratios of concentrations of

METHODOLOGY

The MaWERiC method is comprised of four modules: C.1, wavelet feature extraction; C.2, data representation using PCA; C.3, data combination; and C.4, data classification (Fig. 1). We describe each of these modules in detail in the subsequent subsections.

Table 2. Description of different feature notations and associated dimensionality of each feature vector evaluated

Notation	Description	Number of Dimensions
F^{T_2w}	Gabor feature vector for T$_2$w MRI	34
F^{T_2r1}	non-steerable Gradient T$_2$w MRI feature vector	13
F^{T_2r2}	first order statistical T$_2$w MRI feature vector	8
F^{T_2r3}	Haralick T$_2$w MRI feature vector	13
F^{T_2t}	Combined ensemble of T$_2$w MRI feature vector $F^{T_2t} = F^{T_2w}, F^{T_2r1}, F^{T_2r2}, F^{T_2r3}$	68
F^{MRS}	MRS metabolite and ratio feature vector $F^{MRS}(c) = A_{ch}, A_{cr}, A_{cit}, A_{ch+cr}/A_{cit}, A_{ch}/A_{cr}$	5
F^{MRSw}	MRS wavelet feature vector	Variable (depends on the coefficients chosen to maximize energy)
$F^{T_2w}_{PCA}$	Reduced dimensional T$_2$w MRI Eigen feature vector obtained via PCA on F^{T_2w}	15
F^{MRSw}_{PCA}	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 $F^{Int}(c) = [F^{MRS}(c), F^{T_2}(c)]$	6
F^{Int}_{PCA}	Wavelet PCA based MaWERiC feature vector	30

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 (ψ_l) filter to obtain the corresponding high (H$_h$) and low (L$_l$) filter coefficients as follows:

$$H_h = F(c) * \xi_h, L_l = F(c) * \psi_l,$$

where * is the convolution operator and dimensionality of coefficients H$_h$ and L$_l$ is $M/2$ (M is the dimensionality of each spectrum). Downsampling or decimation by a factor of 2 was performed on coefficients after each pass through filters during multilevel decomposition. We considered the wavelet packet decomposition (WPD) method for extracting MRS wavelet features as suggested in (50), in which both L$_l$ and H$_h$ coefficients are iteratively decomposed using high (ξ_h) and low (ψ_l) pass filters into a full tree like structure of a predefined length K, producing a total of 2^K coefficients. The reconstruction of the signal in WPD was then performed using the best basis algorithm that combined the coefficients that minimized the entropy at each level of the tree (51). As a result, for each MR spectrum $F(c)$ at each $c \in C$, an M dimensional wavelet feature vector $F^{MRSw}(c)$ was 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) *Gabor wavelet features for T$_2$w MRI*

Gabor operators are defined as the convolution of a 2D Gaussian function with a sinusoid, such that the cosine component is considered real and the sine component imaginary (23). As a result, at every metavoxel $c \in C$, a total of 54 Gabor features $F_k^{T_2w}(c), k \in \{1, \ldots, 54\}$ were obtained at nine different scales and six orientations similar to that shown in (22) and are represented by a Gabor feature vector $F^{T_2w}(c)$. 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 that applies a linear transformation to data to obtain the most uncorrelated features by maximizing variance within the data (27). 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 $F^{MRSw}(c)$ is reduced to transformed eigenvector $F_{PCA}^{MRSw}(c) = [e_1, e_2, e_3, \ldots, e_M]$ using PCA, where $[e_1, e_2, \ldots, e_M]$ represents the eigenvectors obtained from eigenvalue decomposition of the data ranked in order of greatest variance. Therefore, the first m eigen vectors that represent a prespecified percentage of the variance in the data are extracted while the remaining eigenvectors are discarded. The high dimensional T$_2$w MRI Gabor feature vector $F^{T_2w}(c)$ is similarly reduced to a lower dimensional representation, $F_{PCA}^{T_2w}(c)$, using PCA.

C.3 Data combination

Because of the physical and dimensionality differences between MR spectra and T$_2$w MRI features, the MaWERiC meta-classifier is created in the joint T$_2$w MRI and MRS eigen space obtained by PCA. Following the mapping of F^{T_2w} and F^{MRSw} to the reduced dimensional eigenvector representations, $F_{PCA}^{T_2w}$ and F_{PCA}^{MRSw}, a new concatenated feature vector $F_{PCA}^{Int} = [F_{PCA}^{T_2w}(c), F_{PCA}^{MRSw}(c)]$, 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) decision trees (52). The C4.5 decision tree is a multistage classifier that 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 (53). RF further combines these decisions 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 classifier $\mathbf{h}_j, j \in \{1, \ldots, N\}$, is constructed (53). The class label (CaP or normal) $\mathbf{h}j(c)$ for each metavoxel $c \in C$ based on the feature vector $F_{PCA}^{T_2w}(c)$ is then obtained using the decision trees $\mathbf{h}_j, j \in \{1, 2, \ldots, N\}$; $\mathbf{h}_j(c) = 1$ if c is classified as CaP (scale 4, 5), and $\mathbf{h}_j(c) = 0$ otherwise. The final class likelihood that c belongs to the CaP class by 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 defined $\mathbf{h}^\rho(c)$ as the binary prediction result for the classifier at each threshold $\rho \in [0, 1]$ such that $\mathbf{h}^\rho(c) = 1$ when $\mathbf{h}(c) \geq \rho$ or 0 otherwise.

EXPERIMENTAL DESIGN AND EVALUATION

Comparative strategies

In the following subsections, we evaluate and compare the individual modules (feature extraction, classification, data integration) that comprise MaWERiC with: (i) other feature extraction strategies used in the context of automated CaP detection for individual T$_2$w MRI and MRS modalities (22,54); (ii) a data integration method similar to a COD method as presented in (18) that combines MRS metabolite features with T$_2$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 for T$_2$w MRI and MRS

We now discuss some of the feature extraction and quantification strategies previously proposed in the context of MRS (30,40) and T$_2$w MRI (3,22,44) that we implemented to quantitatively compare with 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 the presence of CaP on MRS based on choline (A_{ch}), creatine (A_{cr}), citrate peaks (A_{cit}), and the A_{ch+cr}/A_{cit} ratios. Variations in these values from predefined normal ranges ($A_{ch+cr}/A_{cit} < 1$) are highly indicative of the presence of CaP (57,58). To compare MaWERiC with metabolic features used clinically, we created a metabolic feature vector for MRS by calculating the area under the choline (A^{ch}), creatine (A^{cr}), and 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 $c \in C$ was then defined by a metabolite feature vector $\boldsymbol{F}^{MRS}(c) = [A_{ch}, A_{cr}, A_{cit}, A_{ch+cr}/A_{cit}, A_{ch}/A_{cr}]$.

b) Texture features for T$_2$w MRI

Following is a brief description of other individual texture features that have previously been explored in conjunction with classifiers for discriminating between CaP and normal areas on T$_2$w MRI (3,22).

- Non-steerable gradient: Thirteen non-steerable gradient features for each voxel on the T$_2$w MRI scene were obtained by convolution of the T$_2$w MRI scene with the Sobel, Kirsch, and standard derivative operators at every spatial location (59).
- First order statistical: Eight first order statistical features including mean, median, standard deviation, and range of gray scale image intensities within a sliding window neighborhood of 3 x 3 pixels centered around each spatial location in the T$_2$w MRI scene were extracted (59).
- Second order statistical: 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 x 3 pixels centered around each voxel in the T$_2$w MRI scene (60).

For each class of texture features, corresponding T2w texture feature vectors $\boldsymbol{F}^{T_2\tau i}, i \in (1, 2, 3)$ were constructed at every $c \in C$. A combined ensemble of texture features was defined as $\boldsymbol{F}^{T_2 t} = \left[\boldsymbol{F}^{T_2 W}, \boldsymbol{F}^{T_2 \tau 1}, \boldsymbol{F}^{T_2 \tau 2}, \boldsymbol{F}^{T_2 \tau 3} \right]$ obtained by concatenating all textural attributes obtained from T$_2$w MRI. PCA was used to reduce each individual texture feature $\boldsymbol{F}^{T_2\tau i}, i \in (1, 2, 3)$ to the corresponding low dimensional representation $\boldsymbol{F}_{PCA}^{T_2\tau i}(c)$ and this was then used for classification.

Comparative data integration strategies

a) Classifier combination (COI)

Classifiers $\mathbf{h}^{T_2w}(c)$, $\mathbf{h}^{MRS}(c)$ are individually trained on $\boldsymbol{F}_{PCA}^{T_2w}(c)$ and $\boldsymbol{F}^{MRS}(c)$ for all $c \in C$. The independence assumption can then be invoked to fuse $\mathbf{h}_{PCA}^{T_2w(\rho)}(c)$ and $\mathbf{h}^{MRS(\rho)}(c)$ at each $c \in C$ and at every threshold as $\mathbf{h}_d^{Int}(c) = \mathbf{h}_{PCA}^{T_2w(\rho)}(c) \times \mathbf{h}^{MRS(\rho)}(c)$, $\mathbf{h} \in \{\mathbf{RF}, \mathbf{PBT}, \mathbf{SVM}\}$.(26)

b) Data combination (COD) by MRS metabolic area and ratio features and T$_2$w image intensity

A combined feature vector $\boldsymbol{F}^{Int}(c) = [\boldsymbol{F}^{MRS}(c), \boldsymbol{F}^{T_2}(c)]$ was obtained by concatenating the MRS metabolite area and ratio features ($\boldsymbol{F}^{MRS}(c)$) with the mean intensity feature ($\boldsymbol{F}^{T_2}(c)$) for each metavoxel $c \in C$. RF classifier along with PBT and SVM classifiers (described in the next section) were then trained using $\boldsymbol{F}^{Int}(c)$ to obtain the meta-classifiers $\mathbf{h}^{Int}(c), \mathbf{h} \in \{\mathbf{RF}, \mathbf{PBT}, \mathbf{SVM}\}$.

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 L 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 subtrees 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, $\boldsymbol{F}_{PCA}^{Int}(c)$, was calculated at each node based on the learned hierarchical tree.

b) Support vector machine (SVM) classifier

SVM aims at identifying the best possible hyperplane that can accurately separate data into two classes. The SVM classifier is constructed using a kernel function that projects the training data into a higher dimensional space by an implicit feature mapping in the dot product space (55). In our study, the radial basis function (RBF) kernel was used to project the training data into a higher dimensional space. In contrast to PBTs and RFs, where a probability or likelihood was generated for each voxel belonging to a class, SVM classifiers were typically used to generate a hard decision; $\mathbf{h}(c) = 1$ if metavoxel c is identified as CaP and $\mathbf{h}(c) = 0$ otherwise. However, a pseudo likelihood that any metavoxel c belonging to a class could be generated by calculating how far or close each c was from the SVM decision hyperplane during classification and converting this distance in terms of likelihood of each c belonging to a class. As a result, the greater the distance of c from the hyperplane, the higher the likelihood that it belonged 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 was compared with related state-of-the-art feature extraction classifier and data fusion strategies by (i) the area under the receiver operating characteristic (ROC) (62) curve (μ^{AUC}) and (ii) the 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 tradeoff between CaP detection sensitivity and specificity can be generated. Each point on the curve corresponds to the voxel level CaP detection sensitivity and specificity of the classifier ($\mathbf{h}^\rho(c)$) for some $\rho \in [0, 1]$. The operating point Θ on the ROC curve is defined as the value of ρ that yields detection sensitivity and specificity that is closest to 100%. A three-fold randomized cross-validation procedure was used to evaluate the performance of MaWERiC with other strategies. Hence, for the 36 patient studies examined in this study, three sets of spectra obtained from 12 different studies were constituted. During a single run of cross-validation, two of the three sets (corresponding to 24 studies) were chosen for training the classifier while the remaining set of 12 studies were used for independent testing. Classifier results were generated on a per voxel basis. This process was repeated until voxels from all 36 studies were classified once within a single run of

cross-validation. This randomized cross-validation process was then repeated 25 times for different training and testing sets. The mean and standard deviation of classifier AUC values (μ^{AUC}) were recorded over the 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 with unimodal classifiers T_2w MRI and MRS.

MaWERiC was compared with individual feature extraction strategies for T_2w MRI and MRS. Individual features obtained were also quantitatively evaluated with each other to determine the best performing T_2w MRI and MRS features in terms of μ^{AUC} and μ^{Acc}.

Experiment 2. Comparison of MaWERiC with other COD and COI strategies.

MaWERiC was compared with current state-of-the-art COD and COI strategies by the direct combination of metabolic features with T_2w image intensities and a combination of individual classifier predictions for MRI-MRS integration, where binary predictions from the two unimodal classifiers were combined using a dot product operation to obtain the final classification.

Experiment 3. Comparison of classifiers (RFs compared to PBTs and SVMs).

Performance of SVM and PBT classifiers was compared with the RF classifier used for MaWERiC and with other comparative

studies (unimodal T_2w MRI and MRS strategies from Experiment (1) and COD and COI strategies from Experiment (2)) using μ^{AUC} and μ^{Acc} measures.

RESULTS

**Experiment 1.** Comparing MaWERiC with unimodal classifiers T_2w MRI and MRS.

Qualitative results of classifications obtained from Gabor T_2w MRI ($\mathbf{h}^{T_2W}$), 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 (Figs. 3b and 3f) were identified as having a higher probability of CaP as determined by classifiers $\mathbf{h}^{T_2W}$, $\mathbf{h}^{MRS}$, $\mathbf{h}_{,}^{Int} \mathbf{h}_d^{Int}$ and $\mathbf{h}_{PCA}^{Int}$ on a single T_2w slice. Locations shown in blue are identified as having a higher probability of being benign by the classifiers. The white outline in Figure 3a 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 (Fig. 3f) compared to individual unimodal T_2w MRI (Fig. 3b) and MRS (Fig. 3c).

Figure 4a shows the AUC results while Figure 4b shows the accuracy results for different feature extraction strategies ($\mathbf{h}^{T_2W}$, $\mathbf{h}^{MRS}$, $\mathbf{h}_{,}^{Int} \mathbf{h}_d^{Int}$ and $\mathbf{h}_{PCA}^{Int}$) obtained by a RF classifier over 25 runs of cross-validation using box-and-whiskers plots. Note that $m=15$ was used to reduce the dimensionality of T_2w MRI and MRS features since it captured ~93% of the variance in MRS
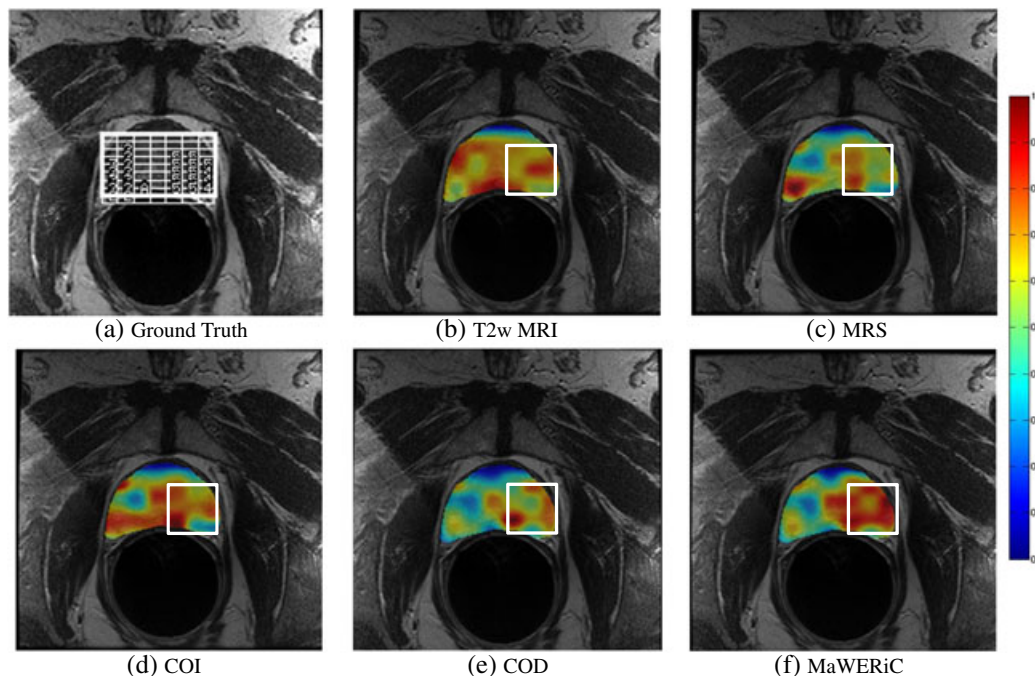
Figure 3. (a) Original T_2w image with MRS grid superposed and labeled according to the 5- point scale (2 = probably benign; 3 = indeterminate; 4 = probably cancer; 5 = definitely cancer; A = atrophy). Figures 3b to 3f: probability heat map results superposed on a single T_2 slice by interpolating the CaP probabilities at MRS resolution to a pixel level T_2w MRI resolution using Gaussian smoothing; (Fig. 3b), T_2w MRI wavelet classifier ($\mathbf{h}^{T_2W}$); (Fig. 3c), MRS classifier ($\mathbf{h}^{MRS}$); (Fig. 3d), COD method involving integration of MRI mean intensity + MRS metabolic features ($\mathbf{h}^{Int}$); (Fig. 3e), decision level integration ($\mathbf{h}_d^{Int}$); and (Fig. 3f) MaWERiC ($\mathbf{h}_{PCA}^{Int}$). 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 3b and 3f denotes the spatial extent of CaP shown on the T_2w slice. Also, note the high detection sensitivity and specificity of CaP probability (Fig. 3f) compared to other classifiers ($\mathbf{h}^{T_2W}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}_d^{Int}$) under evaluation.
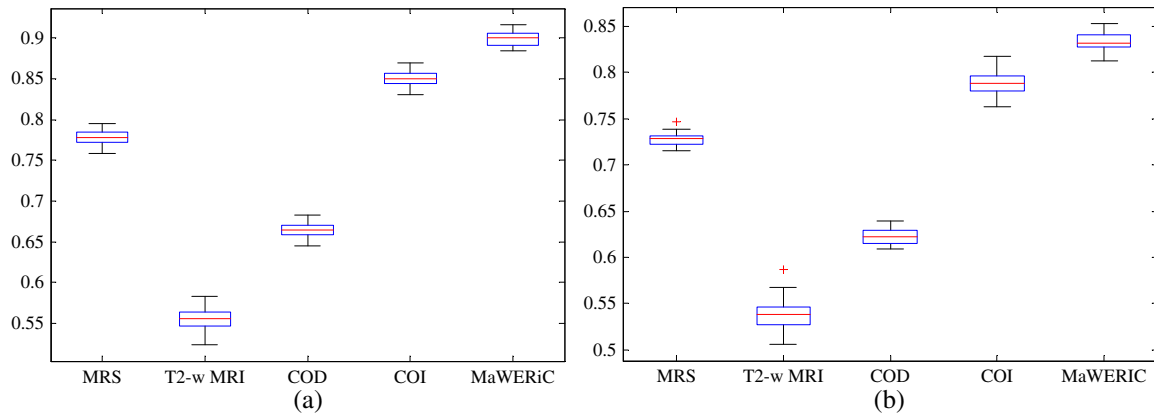
Figure 4. Box-and-whisker plot results of (Fig. 4a) AUC and (Fig. 4b) accuracy obtained over 25 runs of three-fold cross validation in 36 studies for the different feature extraction strategies using a RF classifier. Note that the red line in the middle of each box reflects the median value while the box is bounded by 25 and 75 percentile of AUC (Fig. 4a) and accuracy (Fig. 4b) values. The whisker plot extends to the minimum and maximum values (obtained in all 25 runs) outside the box and the outliers are denoted as the red plus symbol for different feature extraction strategies.

and T$_2$w MRI features. As a result, 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 ($\mathbf{h}^{T_2w}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}_{,}$, $\mathbf{h}^{Int}_d$ and $\mathbf{h}^{Int}_{PCA}$) 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 ($\mu^{AUC} = 0.89 \pm 0.02$, $\mu^{Acc} = 0.83 \pm 0.03$) compared to both T$_2$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 for 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 with all other comparative feature extraction strategies ($\mathbf{h}^{T_2w}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}^{Int}_d$), with the null hypothesis being of 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 ($\mathbf{h}^{Int}_{PCA} - \mathbf{h}^{T_2w}$, $\mathbf{h}^{Int}_{PCA} - \mathbf{h}^{MRS}$, $\mathbf{h}^{Int}_{PCA} - \mathbf{h}^{Int}$, $\mathbf{h}^{Int}_{PCA} - \mathbf{h}^{Int}_d$). Table 5 shows the individual μ^{AUC} and μ^{Acc} values (obtained in 25 runs of 3-fold cross validation) using each set of texture features: first order statistical ($\mathbf{h}^{T_{2\tau1}}$), second order statistical ($\mathbf{h}^{T_{2\tau2}}$), gradient ($\mathbf{h}^{T_{2\tau3}}$), and Gabor ($\mathbf{h}^{T_2w}$) extracted from T$_2$w MRI for three sets of classifiers (SVM, RF, PBT). Except in the case of PBT classifier, Gabor ($\mathbf{h}^{T_2w}$) was found to outperform the other first and second-order statistical and gradient texture features ($\mathbf{h}^{T_{2\tau1}}$, $\mathbf{h}^{T_{2\tau2}}$, and $\mathbf{h}^{T_{2\tau3}}$) for both the RF and SVM classifiers.

Experiment 2. Comparing MaWERiC with peak integration/average MR intensities based COD.

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

Table 3 shows the quantitative results that suggest a significantly higher CaP detection accuracy of MaWERiC; $\mathbf{h}^{Int}_{PCA}$ ($\mu^{AUC} = 0.89 \pm 0.02$, $\mu^{Acc} = 0.83 \pm 0.03$) compared to both COD $\mathbf{h}^{Int}$ ($\mu^{AUC} = 0.66 \pm 0.02$, $\mu^{Acc} = 0.62 \pm 0.02$) and COI, $\mathbf{h}^{Int}_d$ ($\mu^{AUC} = 0.85 \pm 0.03$, $\mu^{AUC} = 0.78 \pm 0.03$). MaWERiC results were found to be significantly better than the other comparative feature extraction strategies ($\mathbf{h}^{T_2w}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}$, $\mathbf{h}^{Int}_d$) in the two classifiers (SVM and PBT).

Experiment 3. Comparing PBT with SVM and RF classifiers.

PBT, SVM, and RF showed similar AUC and accuracy results for all feature combination strategies ($\mathbf{h}^{T_2w}$, $\mathbf{h}^{MRS}$, $\mathbf{h}^{Int}_d$, $\mathbf{h}^{Int}_{PCA}$). Both RF and SVM showed higher μ^{AUC} and μ^{Acc} for MaWERiC ($\mathbf{h}^{Int}_{PCA}$)

Table 3. Mean AUC and accuracy results of different feature extraction and classification techniques used for comparing different strategies with MaWERiC using 25 iterations of 3-fold cross validation in 3 classifier strategies (PBT, RF, SVM)

Index	Feature Extraction Method	Classifier	AUC	Accuracy
1	MRS ($\boldsymbol{F}^{MRS}$)	PBT	0.78 ± 0.03	0.72 ± 0.02
2	T$_2$w MRI ($\boldsymbol{F}^{T_2w}_{PCA}$)	PBT	0.54 ± 0.01	0.54 ± 0.01
3	COD ($\boldsymbol{F}^{Int}$)	PBT	0.72 ± 0.03	0.67 ± 0.03
4	COI ($\mathbf{h}^{Int}_d$)	PBT	0.82 ± 0.01	0.81 ± 0.03
5	MaWERiC ($\boldsymbol{F}^{Int}_{PCA}$)	PBT	0.88 ± 0.03	0.81 ± 0.03
6	MRS ($\boldsymbol{F}^{MRS}$)	RF	0.77 ± 0.03	0.72 ± 0.02
7	T$_2$w MRI ($\boldsymbol{F}^{T_2w}_{PCA}$)	RF	0.55 ± 0.02	0.54 ± 0.01
8	COD ($\boldsymbol{F}^{Int}$)	RF	0.66 ± 0.02	0.62 ± 0.02
9	COI ($\mathbf{h}^{Int}_d$)	RF	0.85 ± 0.03	0.78 ± 0.03
10	MaWERiC ($\boldsymbol{F}^{Int}_{PCA}$)	RF	0.89 ± 0.02	0.83 ± 0.03
11	MRS ($\boldsymbol{F}^{MRS}$)	SVM	0.73 ± 0.01	0.76 ± 0.14
12	T$_2$w MRI ($\boldsymbol{F}^{T_2w}_{PCA}$)	SVM	0.51 ± 0.03	0.65 ± 0.01
13	COD ($\boldsymbol{F}^{Int}$)	SVM	0.68 ± 0.14	0.71 ± 0.08
14	COI ($\mathbf{h}^{Int}_d$)	SVM	0.81 ± 0.22	0.77 ± 0.18
15	MaWERiC ($\boldsymbol{F}^{Int}_{PCA}$)	SVM	0.90 ± 0.24	0.84 ± 0.11

Table 4. *P*-values obtained by pairwise t-test for evaluating presence of statistically significant differences in AUC for MaWERiC compared to the 4 strategies (Gabor MRI, metabolic MRS, COI and COD schemes) under evaluation using an RF classifier

Method	MRS	T$_2$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 classifiers over 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
	First order statistical	0.566 ± 0.013	0.554 ± 0.016
	Second 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
	First order statistical	0.534 ± 0.016	0.530 ± 0.017
	Second 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
	First order statistical	0.493 ± 0.026	0.498 ± 0.027
	Second order statistical	0.504 ± 0.041	0.511 ± 0.033

compared to PBT but the results were not found to be statistically significantly different from each other. Although slightly higher μ^{AUC} and μ^{Acc} were obtained using SVM, RF was used for MaWERiC due to its stable performance over different classifier iterations. Results from SVM classifier were found to have a high standard deviation for both accuracy (0.24 for SVM compared to 0.02 for RF) and AUC values (0.11 for SVM compared to 0.03 for RF).

DISCUSSION

To our knowledge, MaWERiC is the first CSS that provides a systematic framework for the quantitative combination of structural information from T_2w MRI imaging with metabolic information from MRS for improved CaP detection. The few COI and COD based data integration strategies previously explored in the literature were limited in applicability due to the ad-hoc strategies used for overcoming dimensionality differences across modalities (18,45). For instance, Simonetti *et al*. quantitatively combined MRI and MRS by directly concatenating features obtained from the two heterogeneous data sources (18). However, the differing dimensionalities of MRI and MRS features were not accounted for in that study, suggesting that the classifier could have been biased towards MRS features (eight MRS versus four 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 two or more sources about the same disease.

MaWERiC was evaluated on 36 1.5 T *in vivo* MRS and T_2w MRI studies on a per voxel basis and results obtained were compared with four other feature extraction strategies using: (i) MRS metabolic features; (ii) T_2w Gabor wavelet features; (iii) a COD method involving integration of MRS metabolic features with mean image intensity from T_2w MRI; and (iv) a COI method that

combined the independent classification results obtained from T_2w MRI and MRS. We also evaluated three classifiers (SVM, PBT, RF) across the four strategies to identify the best classifier. MaWERiC was found to significantly outperform the other four feature extraction COD and COI strategies for all three classifiers.

To overcome concerns about bias and overfitting of the data, we iteratively divided 36 studies into training and testing sets by a three-fold cross validation method. 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 (Table 3). In the following subsections, we discuss the detection results of MaWERiC with respect to feature extraction, data fusion, and classification strategies considered in this work.

Experiment 1: Comparing MaWERiC with unimodal classifiers T_2w MRI and MRS.

MaWERiC was found to significantly outperform a unimodal classifier trained on Gabor features for T_2w MRI. MaWERiC also outperformed an MRS unimodal classifier trained on clinically used metabolic MRS features. Our results were consistent with several multimodal integration studies (45,63–67) that have suggested that combining orthogonal complementary pieces of information from different modalities can improve classification accuracy compared to unimodal data channels (5,13–16,68).

Our results show that MRS metabolite peak area and ratio features yield better classifiers at a meta voxel level compared to a Gabor texture based T_2w MRI classifier. This is consistent with (35), where μ^{AUC} of 0.68 was obtained using T_2w MRI compared to μ^{AUC} of 0.80 obtained using MRS; the metabolic peaks having been identified by visual inspection of two expert readers. In a related study, MRS ratios of metabolite concentrations (μ^{AUC}=0.89) were shown to outperform visually identified hypointense T_2w MRI features (μ^{AUC}=0.85) for CaP detection in 65 studies (69). 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 study. Our findings (Figs. 4a and 4b) suggest that T_2w MRI texture features alone might not be sufficient to identify CaP signatures in the prostate. Our findings are also consistent with recent 1.5 T and 3 T multi parametric clinical studies that 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 T_2w MRI (70–72).

Experiment 2. Comparing MaWERiC with other data fusion strategies.

MaWERiC versus decision combination (COI)

MaWERiC outperformed a decision level combination approach in terms of μ^{AUC} and μ^{Acc} (32,45). The decision level classifier was obtained by combining the binary class decisions (AND operation) from the individual unimodal classifiers. Decision level integration, while helping to overcome the curse of dimensionality (since all the input information was reduced to a scalar valued decision), tends to implicitly treat the data channels as independent. More specifically, in T_2w MRI and MRS, data is acquired simultaneously, providing complementary (structural and metabolic) information from each spatial location about the same disease. Decision level fusion strategies might therefore be unable to exploit the synergy between these complimentary data streams. By contrast, data level fusion strategies not only exploit the complementary information spread in 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)

To our knowledge, the only other study where MRI and MRS features were quantitatively combined at data levels was for brain tumor detection (18). However, in that study, MRS features obtained by PCA, ICA, and quantification were directly combined with four intensity features from multiprotocol MRI, possibly causing the classifier to be biased towards MRS features. Although MaWERiC was only compared to the best performing COD strategy (quantification + MR intensities), one of four presented in (18), our results suggest that directly aggregating multimodal heterogeneous data from 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 unevenly scaled. The superior classifier accuracy of MaWERiC compared to a COD meta classifier trained using only T$_2$w MR image intensities and metabolic peak area features (Figs. 3a and 3b) could be attributable to the uniform scaling and data representation provided by the MaWERiC framework.

Since high dimensional data could be embedded into a reduced space of arbitrary dimensions, we evaluated MaWERiC across different numbers of eigenvectors, $m \in \{15, 20, 25, 30\}$. The MaWERiC classifier was found to consistently outperform the COD classifier across different values of m (18). $m = 15$ was chosen as the number of low dimensional embedding vectors on which to project the high dimensional T$_2$w 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 MaWERiC classifier were observed for $m \in \{15, 20, 25, 30\}$; these values accounting for $> 93\%$ of variance in the data. Figure 5 shows the variations in μ^{AUC} (y-axis) and μ^{Acc} (x-axis) of MaWERiC using a RF classifier across different values of MRS dimensions from $m = 5$ to 40 ($m = 40$ captures 99.8% variance for MRS) with dimension for T$_2$w MRI fixed at $m = 15$ (captures 99.8% T$_2$w MRI variance). As seen in Figure 5, the highest AUC and accuracy were obtained when dimensions ($m = 15$) were the same for both T$_2$w MRI and MRS. It is important to note that our choice of the 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).
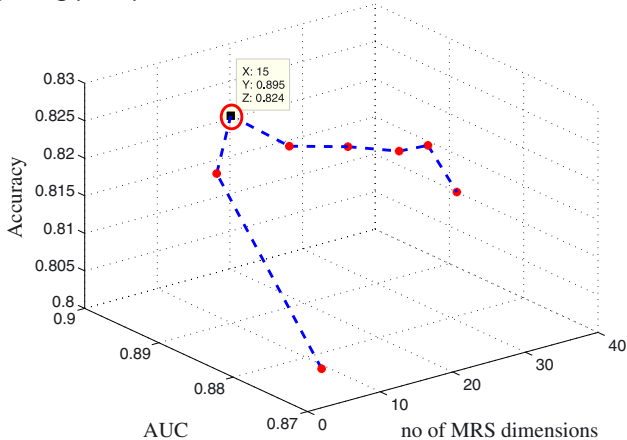
Figure 5. 3-D plot showing variations of AUC (y-axis) and accuracy (z-axis) values of MaWERiC across different PCs (x-axis) for MRS (m is fixed as 15 for T$_2$w MRI as it captures 99.8% MRI variance). Note that the highest AUC and accuracy values were obtained when the same number of dimensions were used for both T$_2$- MRI and MRS at $m = 15$.

Experiment 3. Comparing RF with SVM and PBT classifiers. The three classifiers considered in this study, PBT, RF, and SVM, 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 classifiers is that they are easily able to incorporate information from multiple channels and data sources. RF was used as the ensemble of choice within MaWERiC due to its improved and stable performance over SVM and PBT (Table 2). RF is known to reduce data variance and therefore it is able to provide substantial performance improvement over other ensemble classifier strategies (28). RF classifiers have also shown to be relatively more stable in different levels of noise compared to other classifier ensembles (28).

It was observed that μ^{AUC} obtained by $\mathbf{h}_{PCA}^{Int}$ was statistically and significantly different from $\mathbf{h}^{MRS}, \mathbf{h}^{T_2W}, \mathbf{h}^{Int}$, and $\mathbf{h}_d^{Int}$ across all three classifiers (Table 3), although no statistically significant difference was observed across the classifiers (results not shown). These results suggest that detection performance was more a function of the choice of the feature set and/or fusion strategy (data or decision level) rather than choice of classifiers.

Although the results obtained by MaWERiC significantly outperformed many state-of-the-art feature extraction and fusion strategies for MRS and T$_2$w MRI, we acknowledge some limitations in 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 could provide clinical insights about the disease specific features, a topic that will be explored in future work. (2) Alternative wavelet based (apart from Haar and Gabor) and other feature extraction strategies (e.g., ICA (30,40)) were not considered. However, our choice of Haar wavelets for MRS and Gabor wavelets for T$_2$w MRI was motivated by previous demonstrations of their successful use in building accurate classifiers for CaP detection (22,24). (3) While PCA was used to obtain a uniform homogeneous space for representation of the different modalities, newer NLDR strategies have shown to yield better low dimensional data representations compared to PCA (33,76,77). However, these NLDR strategies are highly sensitive to 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 method of more robustly and accurately estimating spatial extent of disease on MRI is by spatially co-registering *ex vivo* whole mount radical prostatectomy sections with corresponding *in vivo* preoperative MRI. Our group has previously developed elastic registration algorithms for handling deformations between *ex vivo* histology and preoperative 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.

CONCLUSIONS

While data integration strategies for combining image based modalities have been previously presented, analogous strategies for combining imaging and non imaging are not extant in the literature (2,3,42). In this study, we presented a novel data combination method, multimodal wavelet embedding representation for data combination (MaWERiC) specifically geared towards

quantitative integration of imaging and non-imaging data. MaWERiC is comprised of two transformation modules: wavelet transformation and principal component analysis that together provide a platform for uniform and homogeneous data integration across modalities. The homogeneous low dimensional representation of disparate data sources obtained by MaWERiC was then combined in the eigen space. A random forest classifier ensemble was used in conjunction with the combined eigenvector representation of T_2w MRI and MRS channels to identify prostate cancer *in vivo*. Three-fold cross validation performed over 25 iterations and the corresponding pairwise t-test performed on a total of 36 1.5 Tesla *in vivo* T_2w MRI and MRS studies show that MaWERiC significantly outperforms (i) either modality individually, (ii) decision combination obtained by combining individual classifier decisions from both modalities, and (iii) a classifier combining metabolite peak area and ratio features from MRS and T_2w MR image intensities.

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

Acknowledgements

REFERENCES

1. Madabhushi A, Doyle S, Lee G, Basavanhally A, Monaco J, Masters S, Tomaszewski J, Feldman M. Integrated diagnostics: a conceptual framework with examples. Clin. Chem. Lab. Med. 2010; 48(7): 989–998.
2. Hong G, Zhang Y. Comparison and improvement of wavelet-based image fusion. Int. J. Remote Sens. 2008 ;29(3): 673–691.
3. Chan I, Wells Iii W, Mulkern RV, Haker S, Zhang J, Zou KH, Maier SE, Tempany CMC. Detection of prostate cancer by integration of line-scan diffusion, T_2-mapping and T_2-weighted magnetic resonance imaging; a multichannel statistical classifier. Med. Phys. 2003; 30(9): 2390–2398.
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.
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. 2010; 73(2): 352–356.
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 Int. 2001; 87(1): 66–69.
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.
8. Vos P, Hambrock T, Barensz J, and Huisman, H. Computer-assisted analysis of peripheral zone prostate lesions using T_2-weighted and dynamic contrast enhanced T_1-weighted MRI. Phys. Med. Biol. 2010; 55(6): 1719.
9. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics. CA Cancer J. Clin. 2010; 60(5): 277–300.
10. Borboroglu PG, Comer SW, Riffenburgh RH, Amling CL. Extensive Repeat Transrectal Ultrasound Guided Prostate biopsy in Patients with Previous Benign Sextant Biopsies. J. Urol. 2000; 163(1): 158–162.
11. Schiebler ML, Schnall MD, Pollack HM, Lenkinski RE, Tomaszewski JE, Wein AJ, Whittington R, Rauschning W, Kressel HY. Current role of MR imaging in the staging of adenocarcinoma of the prostate. Radiology 1993; 189(2): 339–352.
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.
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.
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.
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 oscopic imaging. Radiology 1999; 213(2): 481–488.
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.
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.
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 Biomed. 2005; 18(1): 34–43.
19. Mallat SG. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. 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]. In Proceedings of ITCC International Conference. 2003: 483–487.
21. Sebe N, Lew MS. Wavelet based texture classification. Pattern Recognition, 2000 Proceedings of 15th International Conference on Pattern Recognition. 2000; 3: 947–950 vol.943.
22. Madabhushi A, Feldman MD, Metaxas DN, Tomaszewski J, Chute D. Automated detection of prostatic adenocarcinoma from high-resolution ex vivo MRI. IEEE Trans. Med. Imaging 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. J. Electr. Electron. Eng. Aust. 1946; 93(26): 429–441.
24. Mira J, Álvarez J, Arévalo Acosta O, Santos Peñas M. Classification of Biomedical Signals Using a Haar 4 Wavelet Transform and a Hamming Neural Network. Nature Inspired Problem-Solving Methods in Knowledge Engineering., Lecture Notes in Computer Science, Volume 4528. Springer Berlin/Heidelberg; 2007; 637–646.
25. Subramani P, Sahu R, Verma S. Feature selection using Haar wavelet power spectrum. BMC Bioinformatics. 2006; 7(1): 432.
26. Duda R, Hart P, Stork D. Pattern Classification, 2nd edn. Wiley-Interscience, New York, 2000; Report nr 0471056693.
27. Hotelling H. Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 1933; 24: 417–441.
28. Breiman L. Random Forests. Mach. Learn. 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.
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. Magn. Reson. Med. 2007; 57: 150–159.
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.
32. Rohlfing T, Pfefferbaum A, Sullivan EV, Maurer CR. Information Fusion in Biomedical Image Analysis: Combination of Data vs. Combination of Interpretations. Inf. Process Med. Imaging 2005; 19: 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). Med. Phys. 2009; 36(9): 3927–3939.

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.

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.

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.

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.

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.

39. Ma J, Sun Z, Dong G, Xie G. Wavelet Denoise on MRS Data Based on ICA and PCA. Advances in Neural Networks. In Proceedings of ISNN. 2005; 3: 748–753.

40. Luts J, Poullet 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. Magn. Reson. Med. 2008; 60(2): 288–298.

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.

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. 2008. ISCCSP 2008. 3rd International Symposium on Communications, Control and Signal Processing. 2008; 12–14 March: 1372–1377.

44. Pieter C Vos TH, Jelle O. Barenstz and Henkjan J Huisman. Computer-assisted analysis of peripheral zone prostate lesions using T$_2$-weighted and dynamic contrast enhanced T$_1$-weighted MRI. Phys. Med. Biol. 2010; 55(6): 1719.

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.

46. Veeramachaneni K, Osadciw LA, Varshney PK. An adaptive multimodal biometric management algorithm. IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev. 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.

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.

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.

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.

51. Coifman RR, Wickerhauser MV. Entropy-based algorithms for best basis selection. IEEE Trans. Inf. Theory 1992; 38(2): 713–718.

52. Prinzie A, Van den Poel D. Random forests for multiclass classification: Random multinomial logit. Expert. Syst. Appl. 2008; 34(3): 1721–1732.

53. Quinlan J. C4. 5: programs for machine learning. Morgan Kaufmann: Boston, 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.

55. Cortes C, Vapnik V. Support-vector networks. Mach. Learn. 1995; 20(3): 273–297.

56. Zhuowen T. Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. 10th IEEE Int. Conf. Computer Vision. ICCV 2005; 1582(2):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.

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.

59. Russ JC. The Image Processing Handbook, 5th edn. (Image Processing Handbook): CRC Press, Inc.; Boca Raton, Florida, 2006.

60. Haralick RM, Dinstein I, Shanmugam K. Textural features for image classification. IEEE Trans. Syst. IEEE Trans. Syst Man Cybern. 1973; SMC-3: 610–621.

61. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 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.

63. Polikar R.. Ensemble Based Systems in Decision Making. IEEE Circuits Syst. Mag. 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.

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.

66. Breiman L. Bagging predictors. Mach. Learn 1996; 24(2): 123–140.

67. Dietterich T. Ensemble Methods in Machine Learning. Int. Workshop Multiple Classifier Syst 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.

69. Squillaci E, Manenti G, Mancino S, Carlani 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.

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.

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 3T for Detection–Hispathologic Correlation. Radiology 2010; 255(1): 89–99.

72. Kim CK, Park BK, Kim B. Localization of prostate cancer using 3T MRI: comparison of T$_2$-weighted and dynamic contrast-enhanced imaging. J. Comput. Assist. Tomogr. 2006; 30(1): 7–11.

73. Mohak S. Feature Selection with Conjunctions of Decision Stumps and Learning from Microarray Data. IEEE Trans. Pattern Anal. Mach. Intell. accessed on 24 June 2011 [Epub ahead of print].

74. Lee SL, Kouzani AZ, Hu EJ. Random forest based lung nodule classification aided by clustering. Comput. Med. Imaging Graph. 2010; 34(7): 535–542.

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.

76. Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. Science 2000; 290(5500): 2319–2323.

77. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. Science 2000; 290(5500): 2323–2326.

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. Med. Phys. 2011; 38(4): 2005–2018.