Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2011 March 4; 7963: 79630U-. doi:10.1117/12.878312.

Enhanced Multi-Protocol Analysis via Intelligent Supervised Embedding (EMPrAvISE): Detecting Prostate Cancer on Multi-Parametric MRI

Satish Viswanath^a, B. Nicolas Bloch^b, Jonathan Chappelow^a, Pratik Patel^a, Neil Rofsky^c, Robert Lenkinski^d, Elisabeth Genega^d, and Anant Madabhushi^a

^aRutgers, the State University of New Jersey, USA

^bBoston Medical Center, Boston, MA, USA

^cUT Southwestern Medical School, Dallas, Texas, USA

^dBeth Israel Deaconess Medical Center, Boston, MA USA

Abstract

Currently, there is significant interest in developing methods for quantitative integration of multi-parametric (structural, functional) imaging data with the objective of building automated meta-classifiers to improve disease detection, diagnosis, and prognosis. Such techniques are required to address the differences in dimensionalities and scales of individual protocols, while deriving an integrated multi-parametric data representation which best captures all disease-pertinent information available. In this paper, we present a scheme called Enhanced Multi-Protocol Analysis via Intelligent Supervised Embedding (EMPrAvISE); a powerful, generalizable framework applicable to a variety of domains for multi-parametric data representation and fusion. Our scheme utilizes an ensemble of embeddings (via dimensionality reduction, DR); thereby exploiting the variance amongst multiple uncorrelated embeddings in a manner similar to ensemble classifier schemes (e.g. Bagging, Boosting). We apply this framework to the problem of prostate cancer (CaP) detection on 12 3 Tesla pre-operative *in vivo* multi-parametric (T2-weighted, Dynamic Contrast Enhanced, and Diffusion-weighted) magnetic resonance imaging (MRI) studies, in turn comprising a total of 39 2D planar MR images. We first align the different imaging protocols via automated image registration, followed by quantification of image attributes from individual protocols. Multiple embeddings are generated from the resultant high-dimensional feature space which are then combined intelligently to yield a single stable solution. Our scheme is employed in conjunction with graph embedding (for DR) and probabilistic boosting trees (PBTs) to detect CaP on multi-parametric MRI. Finally, a probabilistic pairwise Markov Random Field algorithm is used to apply spatial constraints to the result of the PBT classifier, yielding a per-voxel classification of CaP presence. Per-voxel evaluation of detection results against ground truth for CaP extent on MRI (obtained by spatially registering pre-operative MRI with available whole-mount histological specimens) reveals that EMPrAvISE yields a statistically significant improvement (AUC=0.77) over classifiers constructed from individual protocols (AUC=0.62,

0.62, 0.65, for T2w, DCE, DWI respectively) as well as one trained using multi-parametric feature concatenation (AUC=0.67).

Keywords

ensemble embedding; prostate cancer; CAD; 3 Tesla; multi-protocol; multi-parametric; probabilistic boosting trees; DCE-MRI; DWI-MRI; multi-modal integration; non-rigid registration; supervised learning; T2w MRI

1. INTRODUCTION

Quantitative integration of multi-channel (modalities, protocols) information allows for construction of sophisticated meta-classifiers for identification of disease presence.^{1, 2} Such multi-channel meta-classifiers have been shown to perform significantly better compared to any individual data channel.¹ From an intuitive perspective, this is because the different channels of information each capture complementary sets of information. For example, the detection accuracy and qualitative characterization of prostate cancer (CaP) *in vivo* has been shown to significantly improve when multiple magnetic resonance imaging (MRI) protocols are considered in combination, as compared to using individual imaging protocols.³ These protocols include: (1) T2-weighted (T2w), capturing high resolution anatomical information, (2) Dynamic Contrast Enhanced (DCE), characterizing micro-vascular function via uptake and washout of a paramagnetic contrast agent, and (3) Diffusion Weighted (DWI), capturing water diffusion restriction via an Apparent Diffusion Coefficient (ADC) map. DCE and DWI MRI represent functional information, which complements structural information from T2w MRI.³

We now consider some of the most significant challenges² involved in quantitatively integrating multi-parametric (T2w, DCE, DWI) MRI to construct a meta-classifier to detect CaP. First, the issue of *data alignment* needs to be addressed, done in order to bring the multiple channels of information (T2w, DCE, and DWI MRI) into the same spatial frame of reference. This may be done via image registration techniques^{4, 5} which need to be able to account for differences in resolution amongst the different protocols. Post-alignment, the second challenge, *knowledge representation*, requires quantitative characterization of disease-pertinent information. Towards this end, textural and functional image feature extraction schemes previously developed in the context of multi-parametric MRI may be employed.^{2, 6} The final step, *data fusion*, involves some combination of extracted quantitative descriptors to construct the integrated meta-classifier. Dimensionality reduction (DR)⁷ has been shown to be useful for such quantitative fusion^{8, 9} as it allows for the construction of a lower-dimensional *embedding* space which accounts for differences in scale between the different protocols, as well as avoiding the curse of dimensionality. While the image descriptors are divorced from their physical meaning in embedding space (embedding features are not readily interpretable), relevant class-discriminatory information is largely preserved.¹⁰ This makes DR ideal for multi-parametric classification.

2. PREVIOUS RELATED WORK AND NOVEL CONTRIBUTIONS OF THIS WORK

Broadly speaking, multi-modal data fusion strategies may be categorized as *combination of data* (COD) (where the information from each channel is combined prior to classification), and *combination of interpretations* (COI) (where independent classifications based on the individual channels are combined), as shown in Figure 1. A COI approach has typically been shown to be sub-optimal as inter-protocol dependencies are not accounted for.¹ Thus, a number of COD strategies with the express purpose of building integrated quantitative meta-classifiers have recently been presented, including DR-based,¹ kernel-based¹¹ and feature-based¹² approaches.

Multi-kernel learning (MKL) schemes¹¹ represent and fuse multi-modal data based on choice of kernel. One of the challenges with MKL schemes is to identify an appropriate kernel for a particular problem, followed by learning associated weights. The most common approach for quantitative multi-parametric image data integration has involved concatenation of multi-parametric features, followed by classification in the concatenated feature space.¹² Chan et al¹³ leveraged a concatenation approach in combining texture features from multi-parametric (T2w, line-scan diffusion, T2-mapping) 1.5 T *in vivo* prostate MRI to generate a statistical probability map for CaP presence via a Support Vector Machine (SVM) classifier. More recently, a Markov Random Field-based algorithm¹⁴ as well as variants of the SVM algorithm^{15, 16} were utilized to segment CaP regions on multi-parametric MRI via concatenation of quantitative descriptors such as T2w intensity, pharmacokinetic parameters (from DCE), and ADC maps (from DWI).

Lee et al¹ proposed data representation and subsequent fusion of the different modalities in a “meta-space” constructed using DR methods such as Graph Embedding⁷ (GE). However, DR analysis of a high-dimensional feature space may not necessarily yield optimal results for multi-parametric representation and fusion due to (a) noise in the original N -D space which may adversely affect the embedding projection, or (b) sensitivity to choice of parameters being specified during DR. For example, GE is known to suffer from issues relating to the scale of analysis as well as to the choice of parameters used in the method.¹⁷ Varying these parameters can result in significantly different appearing embeddings, with no way of determining which embedding is optimal for the purposes of multi-parametric data integration and classification. There is hence a clear need for a DR scheme which is less sensitive to choice of parameters, while simultaneously providing a quantitative framework for multi-parametric data fusion and subsequent classification.

Researchers have attempted to address problems of sensitivity to noise and choice of parameters in the context of automated classification schemes via the development of classifier ensembles.^{18, 19} These algorithms combine multiple “weak” classifiers to construct a “strong” classifier which has an overall probability of error that is lower compared to any of the individual weak classifiers. Related work which applies ensemble theory in the context of DR has been presented by Hou et al,²⁰ involving a semi-supervised ensemble of DR representations within a multi-view learning framework for web data mining. Similarly,

Athisos et al²¹ employed an ensemble algorithm for nearest neighbor discovery via DR within a content retrieval system.

In this paper, we present a novel solution to better represent and fuse multi-parametric data via a new DR scheme that we refer to as *ensemble embedding*. The spirit behind our technique is to construct a single stable embedding by generating and combining multiple uncorrelated, independent embeddings derived from the multi-parametric feature space. Our rationale for adopting this approach is that the result of ensemble embedding will better preserve class-discriminatory information as compared to any of the individual embeddings used in its construction. We have previously demonstrated preliminary results for a similar scheme⁸ applied to uni-modal data analysis; where multiple embeddings were combined to analyze textural descriptors of *in vivo* T2w MRI data for the presence of CaP. In contrast, our current work is intended to provide a generalized framework for multi-parametric data analysis, while additionally providing theoretical intuition for this approach.

The application of our ensemble embedding framework (termed Enhanced Multi-Protocol Analysis via Intelligent Supervised Embedding or EMPrAvISE) for multi-parametric data representation and fusion is shown in the context of integrating prostate T2w, DCE and DWI MRI for CaP detection. EMPrAvISE is intended to inherently account for (1) differences in dimensionalities between individual protocols (via DR), (2) noise and parameter sensitivity issues with DR-based representation (via the use of an ensemble of embeddings), and (3) inter-protocol dependencies in the data (via intelligent ensemble embedding construction). First, a multi-attribute, higher order mutual information (MI)-based elastic registration scheme (entitled MACMI)⁴ is used to bring the different MRI (T2w, DCE, DWI) protocols into spatial alignment. MACMI is also used to map pathologist-annotated regions of CaP from available *ex vivo* whole-mount radical prostatectomy specimens onto *in vivo* multi-parametric MRI data, to obtain a surrogate ground truth CaP extent on MRI. The information available from each protocol is then characterized via a number of quantitative descriptors,⁶ via application of different feature extraction schemes. Rather than make use of a direct concatenation of all the multi-parametric image features, we utilize an ensemble of embedding representations of the multi-parametric feature data.⁸ The final resulting representation is then used to train a probabilistic boosting tree (PBT) classifier in order to detect CaP presence on a per-voxel basis from multi-parametric MRI. We qualitatively and quantitatively compare CaP detection results obtained via EMPrAvISE against classifier results obtained via individual protocols as well as multi-protocol feature concatenation, on a per-voxel basis. Figure 2 illustrates the different steps comprising EMPrAvISE.

3. THEORY FOR ENSEMBLE EMBEDDING

3.1 Intuition for an Ensemble Embedding approach to Represent and Fuse Multi-Parametric Data

In this section we shall describe some of the theory and properties underlying ensemble embedding; specifically motivating its use within EMPrAvISE for multi-parametric data representation and fusion. Our intent is to analytically demonstrate that ensemble embedding will (1) preserve object-class adjacency from the original high-dimensional feature space as

best possible, and (2) construct a low-dimensional data representation with lower error compared to any single application of DR to the high-dimensional feature space.

3.2 Preliminaries

We first introduce some preliminary notation and definitions (Table 1). An object shall be referred to by its label c and is defined as a point in an N -dimensional space $\mathbb{R}^N$. It is represented by an N -tuple $\mathbf{F}(c)$ comprising its unique N -dimensional co-ordinates. In a subspace $\mathbb{R}^n \subset \mathbb{R}^N$ such that $n \ll N$, this object c in a set C is represented by an n -tuple of its unique n -dimensional coordinates $\mathbf{X}(c)$. $\mathbb{R}^n$ is also known as the *embedding* of objects $c \in C$ and is always calculated via some projection of $\mathbb{R}^N$.

The notation Λ^{cd} , henceforth referred to as the *pairwise relationship*, will represent the relationship between two objects $c, d \in C$ with corresponding vectors $\mathbf{F}(c), \mathbf{F}(d) \in \mathbb{R}^N$. Similarly, the notation δ^{cd} will be used to represent the pairwise relationship between two objects $c, d \in C$ with embedding vectors $\mathbf{X}(c), \mathbf{X}(d) \in \mathbb{R}^n$. We assume that this relationship satisfies the three properties of a metric (e.g. Euclidean distance). Finally, a triplet of objects $c, d, e \in C$ is referred to as a *unique triplet* if $c \neq d, d \neq e$, and $c \neq e$. Unique triplets will be denoted simply as (c, d, e) .

3.3 Definitions

Definition 1. The function defined on a unique triplet (c, d, e) is called a *triangle relationship*, $\triangle(c, d, e)$, if when $\Lambda^{cd} < \Lambda^{ce}$ and $\Lambda^{cd} < \Lambda^{de}$, then $\delta^{cd} < \delta^{ce}$ and $\delta^{cd} < \delta^{de}$.

For objects $c, d, e \in C$ whose relative pairwise relationships in $\mathbb{R}^N$ are preserved in $\mathbb{R}^n$, the triangle relationship $\triangle(c, d, e) = 1$. For ease of notation, the triangle relationship $\triangle(c, d, e)$ will be referred to as $\triangle$ for the rest of this section. Note that for a set of R unique objects (R

$= |C|$, $|\cdot|$ is cardinality of a set), $Z = \frac{R!}{3!(R-3)!}$ unique triplets may be formed.

Definition 2. Given Z unique triplets $(c, d, e) \in C$ and an embedding $\mathbb{R}^n$ of all objects $c, d, e \in C$, the associated embedding strength $\psi^{ES}(\mathbb{R}^n) = \frac{\sum_C \triangle(c, d, e)}{Z}$.

The embedding strength $\psi^{ES}(\mathbb{R}^n)$ is hence the fraction of unique triplets $(c, d, e) \in C$ for which $\triangle(c, d, e) = 1$. We refer to a *true embedding* $\mathbb{R}^{\hat{n}}$ as one for which $\triangle(c, d, e) = 1$, for all unique triplets $(c, d, e) \in C$. $\mathbb{R}^{\hat{n}}$ hence perfectly preserves all pairwise relationships (denoted as $\delta^{\hat{cd}}$ for all objects $c, d \in C$) from $\mathbb{R}^N$. We note that there may be multiple $\mathbb{R}^{\hat{n}}$ that can be calculated from a single $\mathbb{R}^N$; one may choose any one of them to calculate $\delta^{\hat{cd}}$.

Note that the most optimal true embedding will be the original $\mathbb{R}^N$ itself, i.e. $\delta^{\hat{cd}} = \Lambda^{cd}$. However, as $\mathbb{R}^N$ may not be ideal for classification (due to the curse of dimensionality), we are attempting to approximate a true embedding as best possible in n -D space. Practically speaking, most any $\mathbb{R}^n$ will be associated with some degree of error compared to the original $\mathbb{R}^N$. We define the mean squared error (MSE) in the pairwise relationship between every pair of objects $c, d \in C$ in any $\mathbb{R}^n$ with respect to the true pairwise relationships in $\mathbb{R}^{\hat{n}}$ as,

$$\varepsilon_x = E_{cd}(\hat{\delta}^{cd} - \delta^{cd})^2. \quad (1)$$

where E_{cd} is the expectation of the squared error in the pairwise relationships in $\mathbb{R}^n$, calculated over all pairs of objects $c, d \in C$. Assuming a uniform distribution, we can calculate the probability of $(c, d, e) = 1$ for any unique triplet $(c, d, e) \in C$ in any $\mathbb{R}^n$ as,

$$p(\Delta|c, d, e, \mathbb{R}^n) = \frac{\sum_c \Delta(c, d, e)}{Z}. \quad (2)$$

Definition 3. A strong embedding, $\mathbb{R}^{\ddot{n}}$, is an $\mathbb{R}^n$ for which $\psi^{ES}(\mathbb{R}^n) > \theta$.

A strong embedding $\mathbb{R}^{\ddot{n}}$ will accurately preserve the triangle relationship for more than some fraction θ of the unique triplets $(c, d, e) \in C$ that exist. An embedding $\mathbb{R}^n$ which is not a strong embedding is referred to as a *weak embedding*, denoted as $\mathbb{R}^{\dot{n}}$. In this work we utilize classification accuracy to approximate embedding strength. We have demonstrated that the embedding strength of any $\mathbb{R}^n$ increases monotonically with its classification accuracy (not shown for the sake of brevity). Therefore, we may say that strong embedding will have a higher classification accuracy compared to a weak embedding

We can calculate multiple uncorrelated (i.e. independent) embeddings from a single $\mathbb{R}^N$ which may be denoted as $\mathbb{R}_m^n, m \in \{1, \dots, M\}$, where M is total number of possible uncorrelated embeddings. Note that both strong and weak embeddings will be present among all of the M possible embeddings. All objects $c, d \in C$ can then be characterized by corresponding embedding vectors $\mathbf{X}_m(c), \mathbf{X}_m(d) \in \mathbb{R}_m^n$ with corresponding pairwise relationship δ_m^{cd} . Given multiple δ_m^{cd} , we can form a distribution $p(X = \delta_m^{cd})$, over all M embeddings. Our hypothesis is that the maximum likelihood estimate (MLE) of $p(X = \delta_m^{cd})$, denoted as $\tilde{\delta}^{cd}$, will approximate the true pairwise relationship δ^{cd} for objects $c, d \in C$.

Definition 4. An embedding $\mathbb{R}^n$ is called an ensemble embedding, $\mathbb{R}^{\tilde{n}}$, if for all objects $c, d \in C$, $\delta^{cd} = \tilde{\delta}^{cd}$.

We denote the ensemble embedding vectors for all objects $c \in C$ by $\tilde{\mathbf{X}}(c) \in \mathbb{R}^{\tilde{n}}$. Additionally, from Equation 2, $p(\Delta|c, d, e, \mathbb{R}^{\tilde{n}})$ represents the probability that the triangle relationship (c, d, e) will be satisfied for $\mathbb{R}^{\tilde{n}}$. Proposition 1 below aims to demonstrate that for $\mathbb{R}^{\tilde{n}}$ to be a strong embedding, it must be constructed from a combination of multiple strong embeddings $\mathbb{R}^{\ddot{n}}$.

Proposition 1. Given M identical, independent embeddings $\mathbb{R}_m^n, m \in \{1, \dots, M\}$, with a constant $p(\Delta|c, d, e, \mathbb{R}_m^n)$ that $(c, d, e) = 1$ for all $(c, d, e) \in C$, $\lim_{m \rightarrow \infty} \psi^{ES}(\mathbb{R}^{\tilde{n}}) \rightarrow 1$.

The proof may be derived using the Binomial theorem (omitted for the sake of brevity). Proposition 1 reflects two important, necessary properties of ensemble embedding: (1) that some minimum number of strong embeddings $\mathbb{R}^{\ddot{n}}$ must be considered for $\mathbb{R}^{\tilde{n}}$ to become a strong embedding, (2) the strength of the ensemble embedding $\psi(\mathbb{R}^{\tilde{n}})$ will increase significantly as we include more strong embeddings $\mathbb{R}^{\ddot{n}}$ in calculating $\mathbb{R}^{\tilde{n}}$.

While Proposition 1 can be demonstrated for the combination of identical strong embeddings, it may further be extended to combining uncorrelated, independent embeddings which are strong (but are not necessarily identical), in a manner similar to classifier ensemble schemes. Proposition 2 aims to demonstrate that $\mathbb{R}^n$ will have a lower inherent error in its pairwise relationships compared to the uncorrelated, independent strong constituent embeddings $\mathbb{R}_k^n, k \in \{1, \dots, K\}$.

Given K observations $\delta_k^{cd}, k \in \{1, \dots, K\}$, we first define the pairwise relationship in $\mathbb{R}^{\tilde{n}}$ as $\tilde{\delta}^{cd} = E_K(\delta_k^{cd})$, where E_K is the expectation of δ_k^{cd} over K observations. The MSE in $\tilde{\delta}^{cd}$ with respect to the true pairwise relationships in $\mathbb{R}^n$ may be defined as (similar to Equation 1),

$$\varepsilon_{\tilde{x}} = E_{cd}(\hat{\delta}^{cd} - \tilde{\delta}^{cd})^2, \quad (3)$$

where E_{cd} is the expectation of the squared error in the pairwise relationships in $\mathbb{R}^{\tilde{n}}$ calculated over over all pairs of objects $c, d \in C$. It is clear that if for all $c, d \in C$ that $\tilde{\delta}^{cd} = \delta^{cd}$, then $\mathbb{R}^{\tilde{n}}$ is also a true embedding. From Equation 1, we can also calculate the expected MSE over all K embeddings as,

$$\varepsilon_{K,X} = E_K[\varepsilon_X] = E_K \left[E_{cd}(\hat{\delta}^{cd} - \delta_k^{cd})^2 \right]. \quad (4)$$

Proposition 2. *Given K uncorrelated, independent strong embeddings, $\mathbb{R}_k^n, k \in \{1, \dots, K\}$, $\varepsilon_{K,X} \leq \varepsilon_X$.*

The proof may be demonstrated in a manner similar to that shown in [19], where Breiman showed that this result was true in the context of weak classifiers (omitted for the sake of brevity). Proposition 2 implies that $\mathbb{R}^{\tilde{n}}$ will never have a higher error than the maximum error associated with any individual strong constituent embedding $\mathbb{R}_k^n, k \in \{1, \dots, K\}$.

4. METHODOLOGY AND ALGORITHMS FOR *EMPRAVISE*

4.1 Creating n -dimensional data embeddings

One of the requirements for an ensemble embedding is the calculation of multiple uncorrelated embeddings $\mathbb{R}^n$ from the high-dimensional feature space $\mathbb{R}^N$. This is also true of ensemble classifiers such as Boosting¹⁸ and Bagging¹⁹ which require multiple uncorrelated, weak classifications of the data to be generated prior to combination. Similar to Bagging, we make use of a *feature space perturbation* technique to generate uncorrelated embeddings. This is implemented (as shown in the algorithm below) by first creating M bootstrapped feature subsets of V features each from $\mathbb{R}^N$ (each subset $\mathcal{F}_m, m \in \{1, \dots, M\}$ containing $\binom{N}{V}$ features). The feature space associated with each subset $\mathcal{F}_m$ is then embedded into an n -D space via Graph Embedding.⁷ The rationale for this approach is that the resulting $\mathbb{R}_m^n, m \in \{1, \dots, M\}$, obtained in this manner will be independent, uncorrelated embeddings.

Graph Embedding⁷ involves eigenvalue decomposition of a confusion matrix $\mathcal{W} \in \mathbb{R}^{|C| \times |C|}$, representing the adjacencies between all objects $c \in C$ in high-dimensional feature space. The result of GE, $\mathbf{X}(c)$, is obtained from the maximization of the function

$$S(\mathcal{X}) = 2\gamma \times \text{tr} \left[\frac{\mathcal{X}(\mathcal{D} - \mathcal{W})\mathcal{X}^\top}{\mathcal{X}\mathcal{D}\mathcal{X}^\top} \right], \text{ where tr is the trace operator, } \mathcal{X} = [\mathbf{X}(c_1), \mathbf{X}(c_2), \dots, \mathbf{X}(c_q)], q = |C| \text{ and } \gamma = q - 1. \mathcal{D} \text{ is a diagonal matrix where the diagonal element is defined as } \mathcal{D}(i, i) = \sum_j \mathcal{W}(i, j). \text{ Eigenvectors corresponding to the smallest } n \text{ Eigenvalues of } (\mathcal{D} - \mathcal{W}) \mathcal{X}^\top = \lambda \mathcal{D} \mathcal{X} \text{ are calculated. The matrix } \mathcal{X} \text{ of the first } n \text{ Eigenvectors is constructed, and } \forall c \in C, \mathbf{X}(c) \text{ is defined as row } i \text{ of } \mathcal{X}, \text{ such that } \mathbf{X}(c) = [e_v(c) | v \in \{1, \dots, n\}] \in \mathbb{R}^n.$$

Algorithm CreateWeakEmbed

Input: $\mathbf{F}(c) \in \mathbb{R}^N$ for all objects $c \in C, n$

Output: $\mathbf{X}_m(c) \in \mathbb{R}_m^n, m \in \{1, \dots, M\}$

Data Structures: Feature subsets $\mathcal{F}_m$, total number of subsets M , number of features in each subset V

begin

0. for $m = 1$ to M do

1. Select $V < N$ features from $\mathbb{R}^N$, forming subset $\mathcal{F}_m$;

2. Calculate $\mathbf{X}_m(c) \in \mathbb{R}_m^n$, for all $c \in C$ via GE of the feature space associated with each $\mathcal{F}_m$;

3. endfor

end

4.2 Selection of strong embeddings

Having generated M uncorrelated embeddings, we now calculate their corresponding embedding strengths $\psi(\mathbb{R}_m^n), m \in \{1, \dots, M\}$. Embedding strength was approximated by supervised classification accuracy, denoted as ψ^{Acc} . Embeddings for which $\psi^{Acc}(\mathbb{R}_m^n) > \theta$ are then selected as strong embeddings, where θ is a pre-specified threshold.

4.3 Constructing the ensemble embedding

Given K selected embeddings $\mathbb{R}_k^n, k \in \{1, \dots, K\}$, we quantify pairwise relationships between all the objects in each $\mathbb{R}_k^n$ via Euclidean pairwise distances. This yields δ_k^{cd} for all objects $c, d \in C, k \in \{1, \dots, K\}$, stored in a confusion matrix W_k for each $\mathbb{R}_k^n$. Corresponding entries across all W_k (after any necessary normalization) are used to estimate $\tilde{\delta}^{cd}$ (via maximum likelihood estimation), and stored in $\tilde{W}$. In our implementation, we have used the median as the maximum likelihood estimator as (1) the median is less corruptible to outliers, (2) the median and the expectation are interchangeable if one assumes a normal distribution. We apply multi-dimensional scaling²² (MDS) to construct $\mathbb{R}^{\tilde{n}}$ while preserving the pairwise distances in $\tilde{W}$, for all objects $c \in C$.

Note that once the ensemble embedding representation $\mathbb{R}^{\tilde{n}}$ has been constructed, we may construct a classifier to distinguish the different object classes within $\mathbb{R}^{\tilde{n}}$.

4.4 Algorithm

Algorithm *EMPrAvISE*

Input: $\mathbf{F}(c) \in \mathbb{R}^N$ for all objects c, n, M, V, θ
Output: $\tilde{\mathbf{X}}(c) \in \mathbb{R}^{\tilde{n}}$

begin

0. Construct feature space $\mathbf{F}(c) \in \mathbb{R}^N, \forall c \in C$ (via feature extraction);

1. *for* $m = 1$ *to* M *do*

2. Calculate $\mathbf{X}_m(c) = \text{CreateWeakEmbed}(\mathbf{F}(c) | \mathcal{F}_m, M, V), \forall c \in C$, hence yielding $\mathbb{R}_m^n$;

3. $k=0$;

4. Calculate $\psi^{Acc}(\mathbb{R}_m^n)$ (based on classification accuracy);

5. *if* $\psi^{Acc}(\mathbb{R}_m^n) > \theta$

6. $k++$;

7. $W_k(i, j) = \|\mathbf{X}_m(c) - \mathbf{X}_m(d)\|_2 \forall c, d$ with indices i, j ;

8. *endif*

9. *endfor*

10. $\tilde{W}(i, j) = \text{MEDIAN}_k [W_k(i, j)] \forall c, d$;

11. Apply MDS to $\tilde{W}$ to obtain $\mathbb{R}^{\tilde{n}}$;

12. Train a classifier on $\tilde{\mathbf{X}}(c) \in \mathbb{R}^{\tilde{n}}, \forall c \in C$, to distinguish object-class categories;

end

5. EMPRAVISE FOR PROSTATE CANCER DETECTION USING MULTI-PARAMETRIC MRI

5.1 Data Acquisition

A total of 12 pre-operative *in vivo* patient studies were obtained using a 3 Tesla Genesis Signa MRI machine at the Beth Israel Deaconess Medical Center. Each of the patients was diagnosed with CaP via examination of needle core biopsies, and scheduled for a radical prostatectomy. Prior to surgery, MR imaging was performed using an endo-rectal coil in the axial plane and included T2w, DCE, and DWI protocols. The DCE-MR images were acquired during and after a bolus injection of 0.1 mmol/kg of body weight of gadopentetate dimeglumine using a 3-dimensional gradient echo sequence with a temporal resolution of 1 min 35 sec. Two pre-contrast and 5 post-contrast sequential acquisitions were obtained. DWI imaging had B-values of 0 and 1000, with the number of directions imaged being 25, based on which an ADC map was calculated.

Prostatectomy specimens were subsequently sectioned and stained with Haematoxylin and Eosin (H & E) and examined by a trained pathologist to accurately delineate presence and extent of CaP. 39 corresponding whole mount histological sections (WMHS) and T2w MRI slices were automatically identified from these 12 studies, via a recently developed group-

wise matching scheme.²³ The slice correspondences were then validated by a pathologist and radiologist working together.

5.2 Inter-protocol alignment of T2w, DCE, DWI MRI

T2w and ADC (from DWI) must be brought into spatial alignment with DCE MRI (denoted $\mathcal{C}^{T1,t} = (C, f^{T1,t})$, where $f^{T1,t}(c)$ assigns an intensity value to every voxel $c \in C$ at time point t , $t \in \{1, \dots, 6\}$), in order to facilitate analysis of all the data within the same frame of reference. This is done via volumetric affine registration,⁴ hence correcting for inter-acquisition movement and resolution differences between the MRI protocols. Stored DICOM* image header information was used to determine relative voxel locations and sizes as well as slice correspondences between T2w, DCE, and ADC imagery.

Post inter-protocol registration, we obtain the T2w MR image $\mathcal{C}^{T2} = (C, f^{T2})$ and the corresponding ADC map $\mathcal{C}^{ADC} = (C, f^{ADC})$ in alignment with images in $\mathcal{C}^{T1,t}$. Therefore for every voxel $c \in C$, $f^{T2}(c)$ is the T2w MR image intensity value and $f^{ADC}(c)$ is the corresponding ADC value. We analyzed all MRI data at the DCE-MRI resolution (256×256 voxels). Known MRI intensity artifacts such as MR intensity inhomogeneity and non-standardness were then corrected for.²⁴ Figure 3 shows representative results of inter-protocol registration. Note the similarity in spatial alignment and resolution in Figures 3(c)–(e).

5.3 Multi-modal registration of WMHS and MRI to obtain “ground truth” CaP extent

Registration of images from different modalities such as WMHS and MRI is complicated on account of the vastly different image characteristics of the individual modalities.⁴ For example, the appearance of tissue and anatomical structures (e.g. hyperplasia, urethra, ducts) on MRI and histology are significantly different.²⁵ These differences are further exacerbated due to histological processing on WMHS (uneven tissue fixation, gland slicing and sectioning result in duct dilation and tissue loss) and the use of an endo-rectal coil on MRI (causing gland deformation). This may cause registration based on traditional intensity-based similarity measures, such as MI, to fail.⁴ We have previously complemented intensity information with features derived by transformations of these intensities to drive multi-modal registration.⁵

In [4], Chappelow et al leveraged the availability of multiple imaging protocols (T2w, DCE, DWI) to introduce complementary sources of information for registration via a novel image similarity measure, Multi-Attribute Combined MI (MACMI).⁴ MACMI was found to be capable of simultaneously encoding the information from multiple protocols within a multivariate MI formulation. It therefore has the ability to handle images that significantly vary in terms of intensities and deformation characteristics, such as for *in vivo* MRI and *ex vivo* WMHS. Additionally, it involves a simple optimization procedure whereby a sequence of individual image transformations is determined.

*<http://medical.nema.org/>

We implemented MACMI within an elastic registration framework, whereby the similarity measure is used to drive a set of free form deformations (FFDs) defined with a hierarchical grid size. This allows for local image transformations across multiple image resolutions. We denote the transformed WMHS $\mathcal{C}^H = (C, f^H)$, in alignment with $\mathcal{C}^{T1,t}$, $\mathcal{C}^{T2}$, $\mathcal{C}^{ADC}$. CaP extent on $\mathcal{C}^H$ is then mapped onto the DCE coordinate frame C , yielding the set of CaP voxels $G(C)$ (surrogate ground truth CaP extent). We thus assign a label to each voxel $c \in G(C)$, $Y(c) = 1$, with $Y(c) = 0$ otherwise.

Figure 3(a) shows the original WMHS image (identified as being in correspondence with the T2w image in Fig 3(c)), while Figures 3(c)–(e) show the corresponding results of spatially registering the WMH in Figure 3(a) with the corresponding MRI protocols (T2w, DCE, DWI) via MACMI. As a result of image registration (Figure 3(b)), we can map the CaP extent (outlined in white on Fig 3(b)) from WMHS onto the corresponding multi-parametric MRI (CaP extent outlined in red on Figures 3(c)–(e)).

5.4 Multi-parametric feature extraction

The visual appearance of CaP on the different MRI protocols is summarized in Table 2 (based on radiologist and quantitative CAD-derived descriptors). A total of 5 image texture features were calculated from each of $\mathcal{C}^{T2}$ as well as $\mathcal{C}^{ADC}$. These include first and second order statistical features, as well as non-steerable gradient features. The extracted texture features and the corresponding intensity values were concatenated to form the feature

vectors $\mathbf{F}^{T2}(c) = [f^{T2}(c), f_\phi^{T2}(c) | \phi \in \{1, \dots, 5\}]$ (from $\mathcal{C}^{T2}$) and

$\mathbf{F}^{ADC}(c) = [f^{ADC}(c), f_\phi^{ADC}(c) | \phi \in \{1, \dots, 5\}]$ (from $\mathcal{C}^{ADC}$), associated with every voxel $c \in C$. Representative feature images derived from $\mathcal{C}^{T2}$ and $\mathcal{C}^{ADC}$ are shown in Figures 3(f) and (g).

The wash-in and wash-out of the contrast agent within the gland is characterized by varying intensity values across the time-point images $\mathcal{C}^{T1,t}$, $t \in \{1, \dots, 7\}$ (Figure 3(h)). This time-point information is directly concatenated to form a single feature vector $\mathbf{F}^{T1}(c) = [f^{T1,t}(c) | t \in \{1, \dots, 6\}]$ associated with every voxel $c \in C$.

Every voxel $c \in C$ was thus characterized by a number of different multi-parametric feature vectors (summarized in Table 3). For the purposes of comparing *EMPrAvISE* with an alternative data representation scheme, a multi-attribute vector $\mathbf{F}^{Feats}(c)$ is also constructed by directly concatenating the individual T2w, DCE, and ADC attributes.

5.5 Constructing the ensemble embedding representation of multi-parametric MRI data

The algorithm *EMPrAvISE* was applied to the feature vector $\mathbf{F}^{Feat}(c) \in \mathbb{R}^N, N = 18, |\mathbb{R}^N| = |C|$, i.e. for all voxels $c \in C$. We denote $\mathcal{F}$ as the superset of all multi-parametric features, such that $|\mathcal{F}| = N$. Note that $\mathcal{F} = \mathcal{F}_{T2} \cup \mathcal{F}_{T1} \cup \mathcal{F}_{ADC}$ where $\mathcal{F}_{T2}, \mathcal{F}_{T1}, \mathcal{F}_{ADC}$ are feature sets associated with the individual T2w, DCE, ADC protocols respectively. Feature space perturbation was implemented by first forming M bootstrapped subsets of features $\mathcal{F}_m \subset \mathcal{F}$. These features were randomly drawn from $\mathcal{F}$ such that (1) $|\mathcal{F}_u| = |\mathcal{F}_v| = V$, (2) $\mathcal{F}_u \cap \mathcal{F}_v = \emptyset$, (3) each of N features appears in at least one $\mathcal{F}_m$, and (4) one feature from each of $\mathcal{F}_{T2}, \mathcal{F}_{T1}, \mathcal{F}_{ADC}$ appears in each $\mathcal{F}_m$, where $u, v, m \in \{1, \dots, M\}$. The feature space associated

with each feature subset $\mathcal{F}_m$ was then embedded in n -D space via GE,⁷ yielding M corresponding weak embeddings $\mathbb{R}_m^n$.

The corresponding M embedding strengths, $\psi^{Acc}(\mathbb{R}_m^n)$, were then calculated based on the supervised classification accuracy of a probabilistic boosting tree classifier (PBT)²⁶ (additional details in Section 5.6), using labels $Y(c)$, $\forall c \in C$. A leave-one-out cross-validation approach was utilized in the training and evaluation of this PBT classifier.

Embeddings with $\psi^{Acc}(\mathbb{R}_m^n) > \theta$ were then selected as strong, and combined as described in Section 4.3. The final result of *EMPrAvISE* is the ensemble embedding vector $\mathbf{F}^{Em}(c) = [\tilde{e}_v(c)]_{v \in \{1, \dots, n\}} \in \mathbb{R}^{\tilde{n}}$, $\forall c \in C$ (n , the intrinsic dimensionality, is estimated via the technique presented in [27]).

5.6 Classification of multi-parametric MRI via PBTs

A voxel-level probabilistic boosting tree classifier (PBT) classifier was constructed for each feature set, $\mathbf{F}^\beta(c)$, $\beta \in \{T1, T2, ADC, Feats, Em\}$, $\forall c \in C$, considered in Table 3. The PBT algorithm has recently demonstrated success in the context of multi-modal data analysis²⁸ as it leverages a powerful ensemble classifier (Adaboost) in conjunction with the robustness of decision tree classifiers²⁶ to allow for the computation of weighted probabilistic decisions for difficult to classify samples. The PBT classifier comprises the following main steps,

1. A tree structure of length L is iteratively generated in the training stage, where each node of the tree is boosted with T weak classifiers.
2. The hierarchical tree is obtained by dividing new samples into two subsets of $\tilde{\mathbf{F}}_{Right}^\beta$ and $\tilde{\mathbf{F}}_{Left}^\beta$ and recursively training the left and right sub-trees using Adaboost.¹⁸
3. To solve for over-fitting, an error parameter ε is introduced such that samples falling in the range $[0.5 - \varepsilon, 0.5 + \varepsilon]$ are assigned to both subtrees such that $p(Y(c)=1|\mathbf{F}^\beta(c)) \rightarrow \hat{\mathbf{F}}_{Right}^\beta(c)$, and $p(Y(c)=0|\mathbf{F}^\beta(c)) \rightarrow \hat{\mathbf{F}}_{Left}^\beta(c)$. The function $\mathbf{h}^\beta(c) = p(Y(c)|\mathbf{F}^\beta(c))$ represents the posterior class conditional probability of sample c belonging to class $Y(c) \in \{0, 1\}$, given the feature vector $\mathbf{F}^\beta(c)$, $\beta \in \{T1, T2, ADC, Feats, Em\}$.
4. The PBT algorithm stops when the misclassification error (of Adaboost) hits a pre-defined threshold.

During testing, the conditional probability of the object c is calculated at each node based on the learned hierarchical tree. A discriminative model was obtained at the top of the tree by combining the probabilities associated with propagation of the object at various nodes, yielding a posterior conditional probability belonging to the cancer class, $\mathbf{h}^\beta(c) = p(Y(c) = 1 | \mathbf{F}^\beta(c)) \in [0, 1]$, $\beta \in \{T1, T2, ADC, Feats, Em\}$, for every voxel $c \in C$.

5.7 Incorporating spatial constraints via Markov Random Fields

We have previously demonstrated the use of a novel probabilistic pairwise Markov model (PPMMs) to detect CaP lesions on prostate histopathology,²⁹ via the incorporation of spatial

constraints to a classifier output. PPMs formulate Markov priors in terms of probability densities, instead of the typical potential functions,³⁰ facilitating the creation of more sophisticated priors. We make use of this approach to similarly impose spatial constraints to the classifier output (per-voxel), with the objective of accurately segmenting CaP lesions on MRI.

6. EXPERIMENTAL RESULTS AND DISCUSSION

6.1 Performance Evaluation Measures

We define $\mathbf{h}_\rho^\beta(c)$ as the binary prediction result for classifier $\mathbf{h}^\beta(c)$ at each threshold $\rho \in [0, 1]$, such that $\mathbf{h}_\rho^\beta(c)=1$ when $\mathbf{h}^\beta(c) \geq \rho$, 0 otherwise; $\forall \beta \in \{T1, T2, ADC, Feats, Em\}$. For every scene $\mathcal{C}$, threshold ρ , and classifier $\mathbf{h}^\beta(c)$, the set of voxels identified as CaP is denoted $\Omega_\rho^\beta(C)=\{c|\mathbf{h}_\rho^\beta(c)=1\}$, $c \in C$, $\forall \beta \in \{T1, T2, ADC, Feats, Em\}$. We then perform ROC analysis by calculating the sensitivity (SN) and specificity (SP) of $\Omega_\rho^\beta(C)$ with respect to the corresponding ground truth CaP extent $G(C)$, at every $\rho \in [0, 1]$.

A leave-one-out cross validation strategy over the 39 slices was used to evaluate the performance of each of the classifiers constructed (Table 3). An ROC curve is generated for each slice, each curve then corresponding to a single run of leave-one-out cross validation. We then average these ROC curves by first fitting a smooth polynomial through each of the resulting 39 ROC curves. Mean and standard deviation of Area Under the ROC (AUC) values are then calculated. The operating point ϑ on the ROC curve is defined as value of ρ which yields detection SN , SP that is closest to 100% sensitivity and 100% specificity (the top left corner of the graph).

6.2 Experiment 1: Comparison of EMPrAvISE against individual feature based classifiers

We first compared $\mathbf{h}^{Em}$ (via *EMPrAvISE*) against classifiers constructed using the different uni-modal feature sets corresponding to T2w, DCE, and DWI MRI data ($\mathbf{h}^{T2}$, $\mathbf{h}^{T1}$, $\mathbf{h}^{ADC}$). As may be gleaned from Table 4(b), $\mathbf{h}^{Em}$ yields a higher classification accuracy and AUC compared to $\mathbf{h}^{T2}$, $\mathbf{h}^{T1}$, $\mathbf{h}^{ADC}$.

6.3 Experiment 2: Comparison of EMPrAvISE against multi-modal classifier strategies

In this experiment, we compared the performance of $\mathbf{h}^{Em}$ with $\mathbf{h}^{Feats}$. Qualitative comparisons of the probability heatmaps so obtained are shown in Figure 5 (where red corresponds to a higher probability of CaP presence and blue corresponds to lower CaP probabilities). The ground truth spatial extent of CaP obtained by mapping disease extent from WMH onto MR imaging is outlined in red on Figures 5(a) and (d). It can be seen that $\mathbf{h}^{Em}$ (Figures 5(c) and (f)) demonstrates significantly more accurate and specific predictions of CaP presence compared to $\mathbf{h}^{Feats}$ (Figures 5(b) and (e)). This is also reflected in the quantitative evaluation, with $\mathbf{h}^{Em}$ resulting in an AUC of 0.73 (purple curve, Figure 4(a)) compared to an AUC of 0.67 for $\mathbf{h}^{Feats}$ (black curve, Figure 4(a)). Additionally, we see that classification based on multi-parametric integration ($\mathbf{F}^{Feats}$, $\mathbf{F}^{Em}$) outperforms classification based on the individual protocols ($\mathbf{F}^{T1}$, $\mathbf{F}^{T2}$, $\mathbf{F}^{ADC}$). Our quantitative results corroborate

findings in the clinical literature which suggest that the combination of multiple imaging protocols yield superior diagnostic accuracy compared to any single protocol.^{3, 31, 32}

6.4 Experiment 3: Markov Random Fields in conjunction with EMPrAvISE

Figure 6 illustrates results of applying MRFs to the probability heatmaps obtained via *EMPrAvISE* ($\mathbf{h}^{Em}$) to yield $\mathbf{h}_{MRF}^{Em}$. At the operating point of the ROC curve, $\Omega_{\theta}^{Em}(C)$ can be seen to have a number of extraneous regions (Figures 6(c) and (g)). In contrast, $\Omega_{MRF,\theta}^{Em}(C)$ results in a more accurate and specific CaP detection result (Figures 6(d) and (h)). Also shown are RGB colormap representations based on scaling the values in $\tilde{\epsilon}_1(c)$, $\tilde{\epsilon}_2(c)$, $\tilde{\epsilon}_3(c)$ (from $\mathbf{F}^{Em}(c)$) into the RGB colorspace (Figures 6(a), (e)). Similarly colored regions are those that are similar in the ensemble embedding space $\mathbb{R}^n$. Note relatively uniform coloring within ground truth CaP areas in Figures 6(a) and (e), suggesting that *EMPrAvISE* is able to accurately represent the data in a reduced dimensional space while preserving disease-pertinent information.

The ROC curves in Figure 4(a) further demonstrate the improvements in CaP detection accuracy via $\mathbf{h}_{MRF}^{Em}$ (light blue curve, AUC = 0.77). These improvements in AUC and classification accuracy were found to be statistically significant ($p < 0.01$) in a paired two-tailed Students' t -test across the 39 leave-one-out cross-validation runs (Table 4), with the null hypothesis being that no improvement was offered by $\mathbf{h}_{MRF}^{Em}$.

7. CONCLUDING REMARKS

In this paper we presented EMPrAvISE, a novel multi-parametric data representation and integration framework. EMPrAvISE makes use of dimensionality reduction and a supervised ensemble of embeddings to (1) accurately capture the maximum available class information from the data, and (2) account for differing dimensionalities and scales in the data. The spirit behind using an ensemble of embeddings is to exploit the variance among multiple uncorrelated embeddings in a manner similar to ensemble classifier schemes. We have demonstrated the application of EMPrAvISE to the detection of prostate cancer on 3 Tesla *in vivo* multi-parametric (T2w, DCE, DWI) MRI. The low-dimensional data representation via EMPrAvISE was found to be superior for classification as compared to (1) the individual protocols, and (2) concatenation of multi-parametric features. We made use of a probabilistic pairwise Markov Random Field algorithm to complement the result of EMPrAvISE (AUC = 0.77) via the incorporation of spatial constraints. Sources of error within our study may exist due to (1) approximate calculation of slice correspondences between MRI and WMHS, and (2) registration-induced errors in the mapping of ground truth CaP extent from WMHS onto MRI. Therefore, our results could prove more (or less) accurate than reported, based on the margin of error in these 2 methods. However, we also note that there is currently no exact, error-free method to determine the ground truth CaP extent on MRI. Future work will hence focus on validation of our approach on a larger cohort of data. We also intend to explore the application of both EMPrAvISE and ensemble embedding in the context of other domains.

ACKNOWLEDGMENTS

This work was made possible via grants from the Wallace H. Coulter Foundation, National Cancer Institute (Grant Nos. R01CA136535, R01CA140772, and R03CA143991), Department of Defense Prostate Cancer Research Program (W81XWH-08-1-0072), and The Cancer Institute of New Jersey. The authors would like to thank Dr. James Monaco and Dr. Gaoyu Xiao for useful discussions and implementations used in this paper.

REFERENCES

1. Lee G, Doyle S, Monaco J, Madabhushi A, Feldman MD, Master SR, Tomaszewski JE. A knowledge representation framework for integration, classification of multi-scale imaging and non-imaging data: Preliminary results in predicting prostate cancer recurrence by fusing mass spectrometry and histology. *Proc. ISBI*. 2009:77–80.
2. Viswanath S, Bloch B, Rosen M, Chappelow J, Toth R, Rofsky N, Lenkinski RE, Genega E, Kalyanpur A, Madabhushi A. Integrating structural and functional imaging for computer assisted detection of prostate cancer on multi-protocol in vivo 3 Tesla MRI. *SPIE Medical Imaging : Computer-Aided Diagnosis*. 2009; 7260:72603I.
3. Kitajima K, Kaji Y, Fukabori Y, Yoshida KI, Suganuma N, Sugimura K. Prostate cancer detection with 3 T MRI: Comparison of diffusion-weighted imaging and dynamic contrast-enhanced MRI in combination with T2-w imaging. *J Magn Reson Imaging*. 2010; 31(3):625–631. [PubMed: 20187206]
4. Chappelow J, Bloch B, Genega E, Rofsky N, Lenkinski R, Tomaszewski J, Feldman M, Rosen M, Madabhushi A. Elastic Registration of Multimodal Prostate MRI and Histology via Multi-Attribute Combined Mutual Information. *Medical Physics*. 2010 **Accepted**.
5. Chappelow J, Bloch B, Rofsky N, Genega E, Lenkinski R, DeWolf W, Viswanath S, Madabhushi A. COLLINARUS: Collection of Image-derived Non-linear Attributes for Registration Using Splines. *Proc. SPIE*. 2009; 7259
6. 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.
7. Shi J, Malik J. Normalized Cuts and Image Segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 2000; 22(8):888–905.
8. Viswanath S, Rosen M, Madabhushi A. A consensus embedding approach for segmentation of high resolution in vivo prostate magnetic resonance imagery. *Proc. SPIE*. 2008; 6915 69150U-12.
9. Viswanath S, Bloch B, Genega E, Rofsky N, Lenkinski R, Chappelow J, Toth R, Madabhushi A. A Comprehensive Segmentation, Registration, and Cancer Detection Scheme on 3 Tesla In Vivo Prostate DCE-MRI. *Proc. MICCAI*. 2008:662–669.
10. Lee G, Rodriguez C, Madabhushi A. Investigating the Efficacy of Nonlinear Dimensionality Reduction Schemes in Classifying Gene- and Protein-Expression Studies. *Computational Biology and Bioinformatics, IEEE Transactions on*. 2008; 5(3):1–17.
11. Lanckriet GR, Deng M, Cristianini N, Jordan MI, Noble WS. Kernel-based data fusion and its application to protein function prediction in yeast. *Pac Symp Biocomput*. 2004:300–311. [PubMed: 14992512]
12. Verma R, Zacharaki E, Ou Y, Cai H, Chawla S, Lee S, Melhem E, Wolf R, Davatzikos C. Multiparametric Tissue Characterization of Brain Neoplasms and Their Recurrence Using Pattern Classification of MR Images. *Academic Radiology*. 2008; 15(8):966–977. [PubMed: 18620117]
13. Chan I, Wells W III, Mulkern R, Haker S, Zhang J, Zou K, Maier S, Tempany C. Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-w magnetic resonance imaging; a multichannel statistical classifier. *Medical Physics*. 2003; 30(6):2390–2398. [PubMed: 14528961]
14. 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. *Medical Imaging, IEEE Transactions on*. 2009; 28(6):906–915.
15. Artan Y, Haider MA, Langer DL, van der Kwast TH, Evans AJ, Yang Y, Wernick MN, Trachtenberg J, Yetik IS. Prostate cancer localization with multispectral MRI using cost-sensitive

- support vector machines and conditional random fields. *Image Processing, IEEE Transactions on*. 2010; 19(9):2444–2455.
16. Ozer S, Langer DL, Liu X, Haider MA, van der Kwast TH, Evans AJ, Yang Y, Wernick MN, Yetik IS. Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI. *Medical Physics*. 2010; 37(4):1873–1883. [PubMed: 20443509]
 17. Zelnik-Manor, L.; Perona, P. *Advances in Neural Information Processing Systems*. Vol. 17. MIT Press; 2004. Self-tuning spectral clustering; p. 1601-1608.
 18. Freund, Y.; Schapire, R. *Proc. 2nd European Conf. Computational Learning Theory*. Springer-Verlag; 1995. A decision-theoretic generalization of on-line learning and an application to boosting; p. 23-37.
 19. Breiman L. Bagging predictors. *Machine Learning*. 1996; 24(2):123–140.
 20. Hou C, Zhang C, Wu Y, Nie F. Multiple view semi-supervised dimensionality reduction. *Pattern Recognition*. 2009; 43(3):720–730.
 21. Athitsos V, Alon J, Sclaroff S, Kollios G. Boostmap: An embedding method for efficient nearest neighbor retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2008; 30(1): 89–104. [PubMed: 18000327]
 22. Venna J, Kaski S. Local multidimensional scaling. *Neural Networks*. 2006; 19(6):889–899. [PubMed: 16787737]
 23. Xiao G, Bloch B, Chappelow J, Genega E, Rofsky N, Lenkinski R, Tomaszewski J, Feldman M, Rosen M, Madabhushi A. Determining histology-MRI slice correspondences for defining MRI-based disease signatures of prostate cancer. *Computerized Medical Imaging and Graphics*. 2010 **In Press, Corrected Proof**.
 24. Madabhushi A, Udupa JK. New methods of MR image intensity standardization via generalized scale. *Medical Physics*. 2006; 33(9):3426–3434. [PubMed: 17022239]
 25. Bartolozzi C, Menchi I, Lencioni R, Serni S, Lapini A, Barbanti G, Bozza A, Amorosi A, Manganelli A, Carini M. Local staging of prostate carcinoma with endorectal coil MRI: correlation with whole-mount radical prostatectomy specimens. *European Radiology*. 1996; 6:339–345. [PubMed: 8798003]
 26. Tu Z. Probabilistic Boosting-Tree: Learning Discriminative Models for Classification, Recognition, and Clustering. *Proc. IEEE ICCV*. 2005:1589–1596.
 27. Levina E, Bickel P. Maximum likelihood estimation of intrinsic dimension. *Adv. NIPS*. 2005; 17:777–784.
 28. Tiwari P, Rosen M, Reed G, Kurhanewicz J, Madabhushi A. Spectral embedding based probabilistic boosting tree (ScEPTre): classifying high dimensional heterogeneous biomedical data. *Proc. MICCAI*. 2009; 12:844–851.
 29. Monaco J, Tomaszewski J, Feldman M, Hagemann I, Moradi M, Mousavi P, Boag A, Davidson C, Abolmaesumi P, Madabhushi A. High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models. *Medical Image Analysis*. 2010; 14(4):617–629. [PubMed: 20493759]
 30. Potts R. Some generalized order-disorder transformations. *Mathematical Proceedings of the Cambridge Philosophical Society*. 1952; 48(01):106–109.
 31. 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]
 32. Chen M, Dang HD, Wang JY, Zhou C, Li SY, Wang WC, Zhao WF, Yang ZH, Zhong CY, Li GZ. Prostate cancer detection: comparison of T2-weighted imaging, diffusion-weighted imaging, proton magnetic resonance spectroscopic imaging, and the three techniques combined. *Acta Radiol*. 2008; 49(5):602–610. [PubMed: 18568549]

Figure 1.
Summary of multi-modal data fusion approaches.

Figure 2.
Flowchart showing different system components and overall organization of EMPraVISE.

Figure 3.

Images chosen as being in slice correspondence for (a) original WMHS and (c) T2w MR image. CaP outline on (a) is in blue (by a pathologist). (b) Overlay of deformed WMHS image $\mathcal{C}^H$ (via MACMI) onto $\mathcal{C}^{T2}$, allowing mapping of CaP extent (outlined in white). Corresponding co-registered multi-parametric MR images shown for (c) $\mathcal{C}^{T2}$, (d) $\mathcal{C}^{T1.5}$, and (e) $\mathcal{C}^{ADC}$, with mapped CaP extent from (b) outlined in red. Representative texture features (derived within the prostate ROI alone) are also shown for (f) $\mathcal{C}^{T2}$ and (g) $\mathcal{C}^{ADC}$. Note the improvement in image characterization of CaP compared to original intensity information in (c) and (e), respectively. (h) Corresponding time-intensity curves for CaP (red) and benign (blue) regions are shown based on DCE MRI data. Note the differences in the uptake and wash-out characteristics between the red and blue curves.

Classifier	AUC	Accuracy
$\mathbf{h}^{T2}$	0.62 ± 0.22	0.58 ± 0.19
$\mathbf{h}^{T1}$	0.62 ± 0.14	0.61 ± 0.12
$\mathbf{h}^{ADC}$	0.65 ± 0.21	0.62 ± 0.19
$\mathbf{h}^{Feats}$	0.67 ± 0.21	0.63 ± 0.19
$\mathbf{h}^{Em}$	0.73 ± 0.13	0.70 ± 0.10
$\mathbf{h}_{MRF}^{Em} (\mathbf{h}^{Em} + \text{MRF})$	0.77 ± 0.16	0.76 ± 0.12

(b)

Figure 4.
(a) Average ROC curves across 39 leave-one-out cross validation runs. Different colored ROC curves correspond to different classifiers. The best performing classifier was $\mathbf{h}_{MRF}^{Em}$ (c), shown in light blue. (b) Summary of average and standard deviation of AUC and accuracy values for different classifiers averaged over the 39 leave-one-out cross-validation runs, for the different classifier strategies in Table 3.

Figure 5.

Representative results are shown for 2D slices from 2 different studies (on each row). (a), (d) CaP extent outline (in red) delineated on WMHS-T2w MRI overlay (via MACMI). Probability heatmaps are shown for (b), (e) $\mathbf{h}^{Feats}$, and (c), (f) $\mathbf{h}^{Em}$. On each probability heatmap, red corresponds to a higher probability of CaP presence, and the mapped CaP extent (from WMHS) is delineated in green. Note that *EMPrAvISE* ((c), (f)) is far more accurate, with significantly fewer false positives and false negatives compared to either of (b), (e).

Figure 6.

(a), (e) RGB representation of the ensemble embedding (calculated via *EMPrAvISE*) with the CaP ground truth region superposed in black (obtained via registration with corresponding WMHS). (b), (f) Probability heatmap for $\mathbf{h}^{Em}$, where red corresponds to a higher probability for presence of CaP. Note the significantly higher accuracy and specificity of CaP segmentation results via application of MRFs in (d), (h) $\Omega_{MRF, \vartheta}^{Em}(C)$ compared to (c), (g) $\Omega_{\vartheta}^{Em}(C)$ (obtained by thresholding the heatmaps in (b), (f) at the operating point threshold ϑ).

Table 1

Summary of notation used in Section 2.

$\mathbb{R}^N$	High-dimensional space	$\mathbb{R}^n$	Embedding space
c, d, e	Label of object in set C	R	Number of objects in C
$\mathbf{F}(c)$	High-dimensional feature vector	$\mathbf{X}(c)$	Embedding vector
Λ^{cd}	Pairwise relationship in $\mathbb{R}^N$	δ^{cd}	Pairwise relationship in $\mathbb{R}^n$
(c, d, e)	Triangle relationship	$\psi(\mathbb{R}^n)$	Embedding strength
$\mathbb{R}^{\hat{n}}$	True embedding	$\delta^{\hat{c}\hat{d}}$	Pairwise relationship in $\mathbb{R}^{\hat{n}}$
$\mathbb{R}^{\ddot{n}}$	Strong embedding	$\mathbb{R}^{\dot{n}}$	Weak embedding
$\mathbb{R}^{\tilde{n}}$	Ensemble embedding	$\delta^{\tilde{c}\tilde{d}}$	Pairwise relationship in $\mathbb{R}^{\tilde{n}}$

Table 2

Qualitative CaP appearance on multi-parametric MRI and corresponding quantitative features used.

	Qualitative appearance of CaP	Quantitative features extracted
T2w	low T2w signal intensity in peripheral zone	1st order statistics, Kirsch/Sobel (gradients) 2nd order co-occurrence (Haralick)
DCE	distinctly quicker contrast enhancement for CaP compared to benign	Multi-time point intensity information
DWI	significantly low ADC compared to benign	ADC values, gradients 1st and 2nd order statistics

Table 3

Different feature datasets and corresponding classifier strategies considered in this work for multi-parametric data analysis.

Description		Data vectors	Classifier
Single Protocol	T2w	$\mathbf{F}^{T2}(c) = [f^{T2}(c), f_{\phi}^{T2}(c) \phi \in \{1, \dots, 5\}]$	$\mathbf{h}^{T2}(c)$
	DCE	$\mathbf{F}^{T1}(c) = [f^{T1,t}(c) t \in \{1, \dots, 6\}]$	$\mathbf{h}^{T1}(c)$
	ADC	$\mathbf{F}^{ADC}(c) = [f^{ADC}(c), f_{\phi}^{ADC}(c) \phi \in \{1, \dots, 5\}]$	$\mathbf{h}^{ADC}(c)$
Multi-parametric	Features	$\mathbf{F}^{Feat}(c) = [\mathbf{F}^{T2}(c), \mathbf{F}^{T1}(c), \mathbf{F}^{ADC}]$	$\mathbf{h}^{Feat}(c)$
	EMPrAvISE	$\mathbf{F}^{Em}(c) = [\tilde{\epsilon}_v(c) v \in \{1, \dots, n\}]$	$\ \mathbf{h}^{Em}(c), \mathbf{h}_{MRF}^{Em}$

Table 4

p values for a paired Students t-test comparing the improvement in CaP detection performance (in terms of AUC and accuracy) of $\mathbf{h}_{MRF}^{Em}$ over $\mathbf{h}^{T2}$, $\mathbf{h}^{T1}$, $\mathbf{h}^{ADC}$, $\mathbf{h}^{Feats}$, and $\mathbf{h}^{Em}$ respectively. Improvements in accuracy and AUC for $\mathbf{h}_{MRF}^{Em}$ were found to be statistically significantly better ($p < 0.01$) compared to each of $\mathbf{h}^{T2}$, $\mathbf{h}^{T1}$, $\mathbf{h}^{ADC}$, $\mathbf{h}^{Feats}$, and $\mathbf{h}^{Em}$ respectively; the null hypothesis being that no improvement was seen via $\mathbf{h}_{MRF}^{Em}$ in each comparison.

	$\mathbf{h}^{T2}/\mathbf{h}_{MRF}^{Em}$	$\mathbf{h}^{T1}/\mathbf{h}_{MRF}^{Em}$	$\mathbf{h}^{ADC}/\mathbf{h}_{MRF}^{Em}$	$\mathbf{h}^{Feats}/\mathbf{h}_{MRF}^{Em}$	$\mathbf{h}^{Em}/\mathbf{h}_{MRF}^{Em}$
AUC	2.15e-07	1.40e-05	1.33e-04	5.86e-06	2.43e-04
Accuracy	9.64e-08	3.16e-08	1.89e-05	3.32e-05	3.32e-05