A hierarchical spectral clustering and nonlinear dimensionality reduction scheme for detection of prostate cancer from magnetic resonance spectroscopy (MRS)

Pallavi Tiwari

Department of Biomedical Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854

Mark Rosen

Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Anant Madabhushi^{a)}

Department of Biomedical Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854

(Received 10 November 2008; revised 26 June 2009; accepted for publication 26 June 2009; published 12 August 2009)

Magnetic resonance spectroscopy (MRS) has been shown to have great clinical potential as a supplement to magnetic resonance imaging in the detection of prostate cancer (CaP). MRS provides functional information in the form of changes in the relative concentration of specific metabolites including choline, creatine, and citrate which can be used to identify potential areas of CaP. With a view to assisting radiologists in interpretation and analysis of MRS data, some researchers have begun to develop computer-aided detection (CAD) schemes for CaP identification from spectroscopy. Most of these schemes have been centered on identifying and integrating the area under metabolite peaks which is then used to compute relative metabolite ratios. However, manual identification of metabolite peaks on the MR spectra, and especially via CAD, is a challenging problem due to low signal-to-noise ratio, baseline irregularity, peak overlap, and peak distortion. In this article the authors present a novel CAD scheme that integrates nonlinear dimensionality reduction (NLDR) with an unsupervised hierarchical clustering algorithm to automatically identify suspicious regions on the prostate using MRS and hence avoids the need to explicitly identify metabolite peaks. The methodology comprises two stages. In stage 1, a hierarchical spectral clustering algorithm is used to distinguish between extracapsular and prostatic spectra in order to localize the region of interest (ROI) corresponding to the prostate. Once the prostate ROI is localized, in stage 2, a NLDR scheme, in conjunction with a replicated clustering algorithm, is used to automatically discriminate between three classes of spectra (normal appearing, suspicious appearing, and indeterminate). The methodology was quantitatively and qualitatively evaluated on a total of 18 1.5 T *in vivo* prostate T2-weighted (w) and MRS studies obtained from the multisite, multi-institutional American College of Radiology (ACRIN) trial. In the absence of the precise ground truth for CaP extent on the MR imaging for most of the ACRIN studies, probabilistic quantitative metrics were defined based on partial knowledge on the quadrant location and size of the tumor. The scheme, when evaluated against this partial ground truth, was found to have a CaP detection sensitivity of 89.33% and specificity of 79.79%. The results obtained from randomized threefold and fivefold cross validation suggest that the NLDR based clustering scheme has a higher CaP detection accuracy compared to such commonly used MRS analysis schemes as z score and PCA. In addition, the scheme was found to be robust to changes in system parameters. For 6 of the 18 studies an expert radiologist laboriously labeled each of the individual spectra according to a five point scale, with 1/2 representing spectra that the expert considered normal and 3/4/5 being spectra the expert deemed suspicious. When evaluated on these expert annotated datasets, the CAD system yielded an average sensitivity (cluster corresponding to suspicious spectra being identified as the CaP class) and specificity of 81.39% and 64.71%, respectively. © 2009 American Association of Physicists in Medicine. [DOI: [10.1118/1.3180955](https://doi.org/10.1118/1.3180955)]

Key words: magnetic resonance spectroscopy, computer-aided diagnosis, prostate cancer, nonlinear dimensionality reduction, hierarchical clustering, unsupervised classification, k -means, replicated clustering, principal component analysis, locally linear embedding, graph embedding, spectral clustering, z score

I. INTRODUCTION

Prostatic adenocarcinoma (CaP) is the most commonly occurring malignancy among men with 186 320 new cases and 28 660 deaths estimated to occur in the United States in 2008 (*American Cancer Society, 2008*). Early detection of CaP offers the best hope of curing it; however, early prostate cancer is usually asymptomatic.¹ Screening of CaP is based on digital rectal examination and monitoring elevated levels of the blood serum prostate specific antigen (PSA). Definitive diagnosis of CaP involves histological examination of biopsy specimens obtained via a blinded sextant transrectal ultrasound (TRUS) directed biopsy for patients with elevated PSA levels. Since prostate ultrasound is limited in its ability to identify CaP, biopsy locations are chosen at random within the prostate sextants. Consequently, the CaP detection accuracy associated with TRUS is only 20%–25% in patients with elevated PSA levels (4–10 $\mu\text{g/ml}$).² Recently, *in vivo* endorectal T2-weighted (w) structural magnetic resonance (MR) imaging (MRI) of the prostate has allowed for greater discrimination between benign and cancerous prostatic structures as compared to TRUS.³ However, structural T2-w MRI by itself has been shown to be limited in its ability to detect small foci of carcinoma contributing to a relatively low detection specificity.³

Over the past few years, MR spectroscopic (MRS) imaging (MRSI) has emerged as a useful complement to structural MRI for potential screening of CaP.^{4,5} MRSI is a non-invasive technique used to obtain the metabolic concentrations of specific molecular markers and biochemicals in the prostate including citrate, creatine, and choline, and changes in the ratio [choline/citrate or (choline+creatine)/citrate] of which have been shown to be linked to presence of CaP.^{6–9} The spectra are obtained at either single or multiple locations from a rectangular spectral grid placed on a corresponding T2-w MR image. It has been demonstrated previously that the relative concentrations of choline, citrate, and creatine are significantly different in CaP and normal regions within the prostate.^{6,10}

The relative concentrations of choline, creatine, and citrate are obtained by calculating the area under the peak for these metabolites to assess the presence of CaP at a specific prostate location on the T2-w MRI. Identification of precise location of specific metabolites on the MR spectra is a difficult task for radiologists due to (a) a low signal to noise ratio and (b) the presence of biomedical signal artifacts associated with MR spectra such as peak overlap and peak and baseline distortion. Figure 1(a) shows a MRS grid superimposed on the corresponding T2-w MRI slice, while Fig. 1(b) shows the MR spectra acquired from individual MRS voxels from the grid shown in Fig. 1(a). Figures 1(c)–1(e) correspond to representative CaP, noisy and normal spectra obtained from three different locations (shown in red, blue, and green, respectively) within the spectral grid [Fig. 1(b)]. Note the amount of noise in the MRS in Fig. 1(d) which, in some cases, can severely limit the ability of a radiologist to accurately identify individual metabolite peaks and quantitate corresponding peak areas. Thus the usefulness of MRSI as an

FIG. 1. (a) shows a T2-w MRI slice of the prostate with a user-selected 3×7 voxel grid overlaid on the prostate; the MRS spectral grid corresponding to the T2 slice is shown in (b). An abnormal appearing spectra [red voxel in (a) and (b)] is shown in (c), while (e) shows the normal appearing spectra [green voxel in (a) and (b)] (d) shows additional spectra [blue voxel in (a) and (b)] with poor signal to noise ratio and with the baseline affected by a tail of broad upfield lipid resonance (2.0–2.5 ppm). Note that lipid contamination in edge voxels is relatively common in prostate MRSI and can affect the accuracy of the calculated metabolite ratios.

adjunct to MRI as a means of detecting, localizing, and characterizing CaP is highly dependent on the quality of the spectral examinations obtained, owing to the challenges in visually identifying spectroscopic CaP signatures (through the identification of abnormal metabolite peak area ratios) with poor data quality.¹¹

Computer-aided diagnosis (CAD) in medical imaging serves as an adjunctive method for image or data analysis, aiding the radiologist in identifying areas of disease or abnormality.¹² The low detection accuracy associated with current TRUS prostate biopsies points to a need for developing an image guided decision support system to direct needle placement in the prostate to increase CaP detection accuracy. While our group^{13,14} and others¹⁵ have begun to develop CAD systems for CaP detection from structural and functional MR imaging, corresponding developments in MRS have not been as forthcoming. This is in spite of evidence to suggest that integration of structural and metabolic imaging^{7,16} could boost diagnostic yield over that which could be obtained from any individual modality. To date, computer-based approaches for prostate MRSI analysis have focused largely on the use of semiautomated peak area inte-

gration to determine metabolite ratios, although a few researchers^{17–21} explored the use of more sophisticated techniques for automated peak finding.

Previous attempts at computerized analysis of MRS can be classified into two broad categories: (a) signal quantification (model dependent) and (b) pattern recognition (model independent) approaches. Commonly used quantification methods include VARPRO,¹⁷ AMARES,¹⁸ and QUEST,¹⁹ which are software utilities where the objective is to minimize the squared distance between the acquired data and a model basis function built on prior knowledge about the metabolic profiles of a typical MR spectrum. However, the performance of these quantification models is usually dependent on (i) the choice of correct number of model components, (ii) optimal choice of prior knowledge (model function), (iii) presence of noise and contributions from nonprostate spectra, (iv) peak overlap owing to contributions from multiple metabolites, and (v) baseline distortion and line broadening.²² In order to avoid the limitations of model and peak detection based approaches for MRS, recently some researchers have begun to explore domain independent techniques such as z score and principal component analysis (PCA). An excellent comprehensive comparison of quantification and pattern recognition schemes used for MRS analysis is provided in Ref. 22.

z score is a statistical quantity obtained as the ratio of the difference in each individual sample's score and the population mean to the population variance, z score analysis^{20,23} aims to quantify the totality of contributions of all metabolites in the spectral vector. In Refs. 20 and 23, z score was employed for MRS based detection of glioma and CaP, respectively. Another statistical technique, canonical correlation analysis (CCA),²⁴ based on calculating the canonical coefficients to obtain correlated linear relationships between two multidimensional variables was shown to be useful in successfully classifying prostate MRSI datasets into four classes: Aggressive tumor, tumor, mixed tissue, and healthy tissue.

Devos *et al.*²⁵ used linear discriminant analysis (LDA) and a least squares support vector machine (SVM) classifier to discriminate between different tumor classes on brain MRS. Ma and Sun²⁶ and Simonetti *et al.*²⁷ explored other linear dimensionality reduction methods such as independent component analysis and PCA, in conjunction with support vector machine (SVM) classifier, to differentiate between brain tissue classes via MRS.

Dimensionality reduction (DR) refers to the projection of high dimensional data into a reduced dimensional feature space without a significant loss in class discriminatory information. The low dimensional representation of the data is easier to visualize and DR algorithms aim to preserve object relationships so that objects that are close to one another in the high dimensional ambient space are mapped to adjacent locations in the resulting low dimensional embedding space. However, linear DR schemes such as PCA assume the original high dimensional data to be inherently linear and hence employ linear projection methods to reduce data dimensionality. Recently several nonlinear dimensionality reduction

FIG. 2. Flowchart showing various system components and methodological overview of the CAD scheme. Hierarchical spectral clustering is performed in stage 1 to automatically obtain the prostate grid, followed in stage 2 by the identification of different tissue classes via NLDR and replicated clustering.

(NLDR) algorithms have been proposed for the analysis and visualization of nonlinear data.^{29–31} The objective behind NLDR methods is to nonlinearly map objects, c , d belonging to the same object class and characterized by M -dimensional feature vectors $F(c)$, $F(d)$, to adjacent locations $S(c)$, $S(d)$ in the low dimensional embedding; $S(c)$, $S(d)$ representing the m -dimensional dominant eigenvectors corresponding to c , d ($m \ll M$). Unlike linear DR schemes that typically employ the Euclidean distance measure to estimate object distances, most NLDR schemes aim at preserving geodesic distances between objects while computing the embedding. We have previously demonstrated that the use of NLDR methods for representation of high dimensional gene- and protein-expression data, results in better classification compared to linear DR based representation (PCA and LDA).²⁸

In this paper, we present a fully automated CAD system for detecting abnormal/suspicious regions on the prostate using 1.5 T prostate MRSI data. Figure 2 illustrates the organization of our CAD scheme. Our methodology comprises of two stages. In stage 1, a novel hierarchical classification scheme is employed to recursively distinguish prostatic from extracapsular (noninformative) spectra via graph embedding (GE),^{31,32} a well known NLDR scheme, in order to hone in the region of interest (ROI) corresponding to the prostate. In a typical field of view (FOV) of an *in vivo* endorectal prostate T2-w MR image, the prostate occupies a small percentage [approximately 10% in Fig. 1(a)] of the total volume within a prostate MRI scene. Since the extracapsular spectra are most populous, the largest cluster is identified at each iteration as being noninformative and eliminated. This process is repeated until the number of MR spectra remaining is approximately equal to the number usually contained in the prostate (Θ) [Fig. 1(a)]. The removal of extracapsular spectra in stage 1 makes it easier to discriminate between suspicious appearing and normal appearing spectra within the prostate. In stage 2, a NLDR scheme is applied to nonlinearly embed the informative spectra into a reduced dimensional space. The individual prostate spectra now characterized by their low dimensional embedding coordinates are clustered into distinct classes via a “replicated clustering” scheme. All spectra are aggregated into three classes based on the assumption that they correspond to normal, CaP, and indeterminate classes.

The rest of this paper is organized as follows: In Sec. II we provide a detailed description of the feature extraction and clustering schemes employed in this paper. In Sec. III we provide the details of our CAD methodology for prostate MRS classification. In Sec. IV we explain the evaluation scheme employed in this work followed in Sec. V by results of qualitative and quantitative evaluation of our scheme on a total of 18 prostate MRI-MRS studies. Concluding remarks and future research directions are presented in Sec. VI.

II. DESCRIPTION OF FEATURE EXTRACTION AND REPLICATED CLUSTERING METHODS

II.A. Feature extraction methods

II.A.1. Linear dimensionality reduction scheme: PCA

PCA is a linear DR method widely used to visualize high dimensional data and discern object relationships in the data by finding orthogonal axes that contain the greatest amount of variance in the data.³³ These orthogonal eigenvectors corresponding to the largest eigenvalues are called “principal components.” To obtain these principal components, each data point c in set C is first centered by subtracting the mean of all the features for each observation c from its original M dimensional feature value $f_u(c)$, $u \in \{1, \dots, M\}$ as shown by

$$\bar{f}_u(c) = f_u(c) - \frac{1}{|C|} \sum_{c \in C} f_u(c), \quad u \in \{1, \dots, M\}. \quad (2.1)$$

From feature values $\bar{f}(c)$ for each $c \in C$, a new $|C| \times M$ matrix $\mathcal{Y}$ is constructed, where $|C|$ is the cardinality of set C . The matrix $\mathcal{Y}$ is then decomposed into corresponding singular values as shown by

$$\mathcal{Y} = U W_{\text{PCA}} V^T, \quad (2.2)$$

where via singular value decomposition a $|C| \times |C|$ diagonal matrix W_{PCA} containing the eigenvalues of the principal components, a $m \times |C|$ left singular matrix U , and a $M \times |C|$ matrix V are obtained. The eigenvalues in W_{PCA} represent the amount of variance for each eigenvector $\mathbf{S}_v^{\text{PCA}}$, $v \in \{1, 2, \dots, m\}$, in matrix V^T and are used to rank the corresponding eigenvectors in the order of greatest variance. Thus the first m eigenvectors that represent a prespecified percentage of the variance in the data are extracted, while the remaining Eigen vectors are discarded. Thus each data sample $c \in C$ is now described by an m -dimensional embedding vector $\mathbf{S}^{\text{PCA}}(c)$. In spite of the fact that PCA assumes that the data lie on a linear manifold, it allows for specification of the number of eigenvectors required to explain a prespecified percentage of the variance in the data.

II.A.2. z score

For a set of objects, $C^V \subset C$, $c \in C^V$, a mean vector $\mathbf{F}^\mu = [f_u^\mu | u \in \{1, \dots, M\}]$ and the corresponding standard deviation vector $\mathbf{F}^\sigma = [f_u^\sigma | u \in \{1, \dots, M\}]$, where $f_u^\mu = 1/|C^V| \sum_{c \in C^V} f_u(c)$ and $f_u^\sigma = \sqrt{1/|C^V| \sum_{c \in C^V} [f_u(c) - f_u^\mu]^2}$ are obtained. At each $c \in C$, z score ($S^z(c)$) is defined as

$$S^z(c) = \frac{\|\mathbf{F}(c) - \mathbf{F}^\mu\|_2}{\|\mathbf{F}^\sigma\|_2}, \quad (2.3)$$

where $\mathbf{F}(c)$ corresponds to the feature vector at each c and $S^z(c)$ reflects the degree to which the value of an object deviates from the normal based on a statistical linear model. A predefined threshold θ_z is then used to classify each $c \in C$ into one class or the other based on whether $S^z(c) \geq \theta_z$.

II.A.3. Nonlinear dimensionality reduction methods

In this work we consider two popular NLDR schemes, locally linear embedding (LLE) (Ref. 30) and GE (Ref. 32) for MRS analysis. We aim to demonstrate the use of NLDR schemes for representation of MRS data results in superior discrimination between CaP and non-CaP spectra compared to the use of linear DR schemes such as PCA.

Graph embedding. The aim of graph embedding³² is to find an embedding vector $\mathbf{S}^{\text{GE}}(c_i)$, $\forall c_i \in C$, $i \in \{1, \dots, |C|\}$, such that the relative ordering of the distances between objects in high dimensional space is maximally preserved in the lower dimensional space. Thus, if locations $c_i, c_j \in C$, $i, j \in \{1, \dots, |C|\}$ are adjacent in the high dimensional feature space, then $\|\mathbf{S}^{\text{GE}}(c_i) - \mathbf{S}^{\text{GE}}(c_j)\|_2$ should be small, where $\|\cdot\|_2$ represents the Euclidean norm. This will only be true if the distances between all $c_i, c_j \in C$ are preserved in the low dimensional mapping of the data. To compute the optimal embedding, an adjacency matrix $W_{\text{GE}} \in \mathbb{R}^{|C| \times |C|}$ is first defined as

$$W_{\text{GE}}(i, j) = e^{-\|\mathbf{F}(c_i) - \mathbf{F}(c_j)\|_2}, \quad \forall c_i, c_j \in C, i, j \in \{1, \dots, |C|\}. \quad (2.4)$$

$\mathbf{S}^{\text{GE}}(c_i)$ is then obtained via maximization of the following objective function:

$$E(\mathcal{X}_{\text{GE}}) = 2\gamma \text{tr} \left[\frac{\mathcal{X}_{\text{GE}}(D - W_{\text{GE}})\mathcal{X}_{\text{GE}}^T}{\mathcal{X}_{\text{GE}} D \mathcal{X}_{\text{GE}}^T} \right], \quad (2.5)$$

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

Locally linear embedding (LLE). LLE (Ref. 30) operates by assuming that objects within a local neighborhood in a high dimensional feature space are linearly related. Consider the set of high dimensional feature vectors $\mathcal{F} = \{\mathbf{F}(c_1), \mathbf{F}(c_2), \dots, \mathbf{F}(c_n)\}$, $\forall c_i \in C$, $i \in \{1, \dots, n\}$. LLE aims to map the set $\mathcal{F}$ to the corresponding set $\mathcal{X}_{\text{LLE}} = \{\mathbf{S}^{\text{LLE}}(c_1), \mathbf{S}^{\text{LLE}}(c_2), \dots, \mathbf{S}^{\text{LLE}}(c_n)\}$ of embedding coordinates. Let $d^{(1)}, \dots, d^{(K)}$ be the K nearest neighbors of c_i and let $\eta^K(c_i)$ be the indices of the location of the K -nearest neighbors (K -NN) of $c_i \in C$. The feature vector $\mathbf{F}(c_i)$ and its K -NN's $\{\mathbf{F}(d^{(1)}), \mathbf{F}(d^{(2)}), \dots, \mathbf{F}(d^{(K)})\}$ are assumed to lie on a

patch of the manifold that is locally linear, allowing us to use the Euclidean metric to determine distance between neighbors. Each $F(c_i)$ can then be approximated by a weighted sum of its K -NN. The optimal reconstruction weights are given by the sparse matrix W_{LLE} (subject to the constraint $\sum_j W_{LLE}(i,j)=1$) that minimizes

$$E_1(W_{LLE}) = \sum_{i=1}^n \left\| F(c_i) - \sum_{r=1}^K W_{LLE}(i,r) F(d^{(r)}) \right\|_2. \quad (2.6)$$

Having determined the weighting matrix W_{LLE} , the next step is to find a low dimensional representation of the points in $\mathcal{F}$ that preserves this weighting. Thus, for each $F(c_i)$ approximated as the weighted combination of its K -NN, its projection $S^{LLE}(c_i)$ will be the weighted combination of the projections of these same K -NN. The optimal $\mathcal{X}_{LLE}$ in the least squares sense minimizes

$$E_2(\mathcal{X}_{LLE}) = \sum_{i=1}^n \left\| S^{LLE}(c_i) - \sum_{j=1}^n W_{LLE}(i,j) S^{LLE}(c_j) \right\|_2^2 \\ = \text{tr}(\mathcal{X}_{LLE} L \mathcal{X}_{LLE}^T), \quad (2.7)$$

where $\mathcal{X}_{LLE} = [S^{LLE}(c_1), S^{LLE}(c_2), \dots, S^{LLE}(c_n)]$, $L = (I - W_{LLE})(I - W_{LLE}^T)$, and I is the identity matrix. The minimization of Eq. (2.7) subject to the constraint $\mathcal{X}_{LLE} \mathcal{X}_{LLE}^T = I$ (a normalization constraint that prevents the solution $\mathcal{X}_{LLE} = \mathbf{0}$) is an eigenvalue problem whose solutions are the eigenvectors of the Laplacian matrix L . Since the rank of L is $n-1$, the first eigenvector is ignored and the second smallest eigenvector represents the best one-dimensional projection of all the samples. The best two-dimensional projection is given by the eigenvectors with the second and third smallest eigenvalues, and so forth.

II.B. Replicated k -means clustering in the reduced feature space

For the DR schemes, $\phi \in \{\text{PCA}, \text{LLE}, \text{GE}\}$, unsupervised replicated clustering is used to classify all objects $c \in C$ into one of the k classes based on $S^\phi(c)$, the low dimensional representation of $F(c)$. Replicated clustering is a variant of the popular k -means³⁴ clustering scheme. The k -means algorithm is initialized by randomly partitioning the data into k clusters and computing the cluster center for each partition. The distance of each point from each of the k centroids is computed and each object is reassigned to the closest cluster centroid to minimize the intraclass variance. This random initialization may lead to local minima leading, in turn, to different clustering results. The motivation behind replicated clustering is to make the final aggregation results from k -means more deterministic. In replicated clustering, multiple weak clusterings of the data are generated. The optimal clustering solution is then chosen among the various weak clusterings as the one with the least intraclass variance. Below we briefly describe the various steps involved in this algorithm.

Step 1. At each of T iterations, k -means is applied to cluster all objects $c \in C$ into one of the k classes $V_t^1, V_t^2, \dots, V_t^k$,

FIG. 3. Illustration of the spatial relationship between MRS metavoxels $c \in C$ and T2-w MRI voxels $g \in G$. The spectral grid C comprising of 28 metavoxels has been overlaid on a T2-w MRI prostate slice and is shown in white. Note the region outlined in red on C corresponds to the area occupied by a metavoixel, but may contain multiple MRI voxels (highlighted in red).

$t \in \{1, \dots, T\}$, where each $c \in C$ is characterized by a high dimensional feature vector $F(c)$. For each $c \in V_t^q$, $q \in \{1, 2, \dots, k\}$, $t \in \{1, 2, \dots, T\}$ the centroid of each cluster is determined as

$$F_t^q = \frac{1}{|V_t^q|} \sum_{c \in V_t^q} F(c). \quad (2.8)$$

Step 2. At each iteration $t \in \{1, \dots, T\}$, the average Euclidean distance between each $F(c) \in V_t^q$ and corresponding cluster center F_t^q , $t \in \{1, \dots, T\}$, $q \in \{1, 2, \dots, k\}$, is then determined as

$$d_t^q = \frac{1}{|V_t^q|} \sum_{c \in V_t^q} \|F(c) - F_t^q\|. \quad (2.9)$$

The average intracluster distance over all k clusters is then obtained as

$$\mu_t^d = \frac{1}{k} \sum_q d_t^q. \quad (2.10)$$

Step 3. Finally, the clusterings $\hat{V}^q$, $q \in \{1, 2, \dots, k\}$ within a specific iteration $t \in \{1, \dots, T\}$ are identified as the stable clustering result for which μ_t^d is minimum over all t .

Note that replicated clustering identifies stable clusterings as those that minimize intraclass variance. Note further that while we are not explicitly seeking to increase intercluster distance, our empirical results suggest that replicated k -means clustering tends to also push the cluster centers farther apart.

III. METHODOLOGY

In Sec. III A we provide a brief description of the notation employed in this paper. In Sec. III B we provide a brief description of the data sets considered in this study. Details concerning determination of the approximate ground truth for spatial extent of CaP on MRI are provided in Sec. III C. Methodological details regarding the two-stage CAD system for CaP detection on MRS are detailed in Secs. III D and III E, respectively.

III.A. Notation

We represent the 3D prostate T2-w MRI scene by $\mathcal{G} = (G, f)$, where G is a 3D grid of voxels $g \in G$ and $f(g)$ is a function that assigns an intensity value to every $g \in G$. We

TABLE I. List of commonly used notation and symbols in this paper.

Symbol	Description
$\mathcal{G}$	3D MRI scene
G	3D grid of MRI voxels
g	Voxel location in $G, g \in G$
$f_u(c)$	MR signal intensity at c
$f(g)$	MR intensity value at g
M	Number of original high dimensions
ϕ	DR method, $\phi \in \{\text{PCA}, \text{LLE}, \text{GE}\}$
i, j	Two adjacent spatial locations on an image
$\mathcal{R}$	Maximum CaP diameter
C_p	Potential cancer space, $C_p \subset C$
$\hat{V}_{q,q}$	Stable clusters obtained for $q \in \{1, 2, \dots, k\}$
$\Delta X, \Delta Y$	Size of metavoxel in X and Y dimensions
α	Threshold parameter for z score, $\alpha \in [0, 1]$
C_s	Set of precise spatial locations of CaP
$C_{a,o}$	Spatial locations of C_a outside C_p
$SN(C_a^{CaP})$	Sensitivity value associated with CaP class
W_ϕ	Distance matrix for $\phi \in \{\text{GE}, \text{LLE}, \text{PCA}\}$
κ	Locally linear parameter for $\phi \in \{\text{LLE}\}$
$\mathcal{C}$	3D MR spectral scene
C	3D grid of metavoxels
c	A metavoxel in $C, c \in C$
u	Frequency index
$F(c)$	Vector of spectral content at c
m	Number of reduced dimensions, $m \ll M$
$S^\phi(c)$	Low dimensional embedding vector at c
$S^z(c)$	z scores at each $c \in \tilde{C}_T$
K_s	Contiguous slices with CaP presence
N_g	Number of CaP voxels in $C_p, N_g \in C_p$
V_T	Prostate grid
Θ	Threshold for prostate size
θ_z	z score threshold
C_a^q	Set of spatial locations for clusters $\hat{V}_{q,q}, q \in \{1, 2, \dots, k\}$
$C_{a,i}$	Spatial locations of C_a inside C_p
$SP(C_a^{CaP})$	Specificity value associated with CaP class
η	Confidence estimate associated with SN and SP
u	Dimension parameter for $\phi \in \{\text{LLE}, \text{GE}, \text{PCA}\}$

also define a spectral scene $\mathcal{C} = (C, F)$ where C is a 3D grid of metavoxels, $c \in C$. Each metavoxel c is associated with a corresponding M -dimensional spectral vector $F(c) = [f_u(c) | u \in \{1, \dots, M\}]$, where $f_u(c)$ represents the MRS signal intensity at each c along the frequency domain. Figure 3 shows the spatial relationship between the MR spectral metavoxel $c \in C$ and T2-w MRI voxel $g \in G$. Note that the distance between any two adjacent metavoxels $c_i, c_j \in C$, $\|c_i - c_j\|_2$, where $\|\cdot\|_2$ denotes the L_2 norm, $i, j \in \{1, \dots, |C|\}$, and $|C|$ is the cardinality of C , is roughly 13 times the distance between any two adjacent MRI voxels $g_i, g_j \in G$, where $i, j \in \{1, \dots, |G|\}$. A list of commonly used notations and symbols in this paper is given in Table I.

III.B. MRSI data description

A total of 18 deidentified and anonymized prostate cancer MRI/MRS datasets from the ACRIN trial was chosen randomly for this pilot study. All exams were performed on a

1.5 T magnet (GE Medical Systems, Milwaukee, WI). Oblique axial T2-w spectra (TR 4000–6000 ms, TE 90–120 ms, slice thickness of 3 mm, acquisition matrix of 256×192 , and FOV of 12–14 cm) were first obtained. Using the PROSE (Prostate spectroscopy and imaging examination) software package (www.gehealthcare.com/user/mr/applications/products/prose.html), MRSI (TR 1000 ms, TE 130 ms, spectral width of 1000 Hz, and number of points is 512) was then prescribed from the oblique axial T2-w images, with outer voxel suppression and oblique suppression planes manually set to exclude the majority of fat about the prostate capsule. A FOV of 11 cm was used for the 16×16 MRSI grid (voxel dimension of $6.75 \times 6.75 \times 3$ mm³). MRS was obtained as a PRESS (point resolved spectroscopy sequence) sequence.³⁵ The MRS spectral grid was contained in DICOM image sets, from which the 16×16 grid containing 256 complex spectra per slice was obtained using IDL 6.4 (ITT Visual Information Systems).

III.C. Determining approximate ground truth for spatial extent of CaP on MRI-MRS

Following radical prostatectomy, the gland was fixed in formalin and sectioned per institutional routine (whole mount or standard sections). Sections were then embedded in formalin, and duplicate slides for each block were prepared for central review. Slides were stained with Hematoxylin and Eosin and reviewed by a single central pathologist for areas of cancer. MRI-pathology correlation was established through a joint review session of trial imagers and pathologists, who determined slice-by-slice histology-MRI concordance. In order to ensure uniform spectral and histologic sextant assignment, sextant boundaries (i.e., between apex and midgland, and between midgland and base) were based on the MRI slice assignments that had previously been determined by each site radiologist. Using these sextant boundaries, and the best approximation of MRI-histologic concordance, the presence and the diameter of CaP in each sextant were established with maximum tumor diameter, denoted as $\mathcal{R}$, recorded for all positive sextants.

III.C.1. Potential cancer space

In order to quantitatively evaluate CAD performance in terms of performance metrics such as sensitivity and specificity, the precise spatial location of the target class within C is required. Unfortunately, this information was not readily available for the studies considered in this project. Hence, we define a probabilistic ground truth for CaP that involves first defining a *potential cancer space* representing a spatial region (C_p) within C within which the tumor is embedded. To appreciate the need for C_p , let us assume the ideal case scenario (Fig. 4) where the precise spatial location of CaP is known *a priori* and is denoted by the set C_s . If C_a denotes the set of spatial locations corresponding to CaP identified by the CAD system, the true positive (TP) area could be calculated as $|C_s \cap C_a|$. Similarly, the false positive (FP) area for the CAD system is $|C_a - (C_s \cap C_a)|$ and false negative (FN) area is $|C_s - (C_s \cap C_a)|$. For the problem we are considering,

FIG. 4. An illustration of the precise ground truth location (C_s) on the prostate and spatial location of the class (C_a) identified as CaP by a CAD system. Note that in this case sensitivity, specificity of CaP detection via CAD can be determined precisely since C_s is known exactly. C_p represents the potential cancer space that needs to be defined when C_s is not available and contains within it N_g CaP metavoxels.

C_s is unavailable and hence the need for C_p which is defined as the set of spatial locations within which a total number of N_g CaP locations are contained. The true number of metavoxels c within C_p that represent CaP can be calculated as

$$N_g = K_s \times \left\lceil \frac{(\mathcal{R}^2)}{\Delta X \Delta Y} \right\rceil, \quad (3.1)$$

where K_s represents the number of contiguous MR sections containing CaP, $\lceil \cdot \rceil$ refers to the ceiling operation, and ΔX and ΔY refer to the size of the metavoxel c in the X and Y dimensions. Thus for a MRS scene $\mathcal{C}$, with known cancer in left midgland (LM), the prostate being contained in a 3×6 grid and the prostate midgland region extending over two contiguous slices, the total number of CaP metavoxels $|C_p|$ is 18 ($3 \times 3 \times 2$). The 3×6 prostate grid is divided into two equal right and left halves. Given that the tumor has a maximum diameter of 13.75 mm in LM, with $\Delta X, \Delta Y = 6.875$, $N_g = 8$ metavoxels corresponding to CaP within C_p . Both C_p and N_g are integral to defining probabilistic estimates of CAD sensitivity, specificity, and positive predictive value (PPV), details of which are provided in Sec. IV.

III.C.2. Expert annotations

We also endeavored to obtain a more precise estimate of CaP ground truth for a subset of the studies via expert annotation of individual MR spectra. For 6 of the 18 ACRIN studies, an expert radiologist laboriously annotated each individual spectra with labels 1–5; with 1 and 2 being spectra that appeared to be normal and spectra 3–5 being suspicious appearing. For this expert annotated dataset, our CAD system attempted to discriminate all spectra as being either normal or suspicious appearing (i.e., distinguish between spectra labeled as 1/2 and those labeled as 3/4/5).

III.D. Stage 1: Localization of prostate using hierarchical spectral clustering of MRS

For the FOV for studies considered in this work, a majority of the spectra lay outside the prostate. The motivation behind stage 1 of our algorithm is to exploit the differences in spectral characteristics of spectra inside and outside the prostate to hone in the prostate ROI automatically. This is done by identifying and eliminating the dominant cluster at every iteration until a cluster size threshold is attained. The

nonlinear DR scheme graph embedding³² is employed to project all spectra into a reduced dimensional embedding $S^{GE}(c)$, followed by replicated k -means clustering to aggregate all $c \in C$ into two clusters $\hat{V}^1, \hat{V}^2$ corresponding to prostatic and extracapsular spectral classes. At each iteration $t \in \{1, \dots, T\}$, a subset of voxels $\tilde{C}_t$ in C is obtained by eliminating the noninformative extracapsular spectra identified as the dominant cluster ($\hat{V}^{dom}$). The approximate number of prostate spectra (Θ) of a MRS grid is learned during the offline training phase. The automatic cascaded scheme stops when the number of remaining spectra in the MRS grid is approximately equal to Θ . The result of the HierarclustMRS algorithm is a spectral grid ($\tilde{C}_T$) containing all the prostate spectra.

Algorithm HierarclustMRS

Input: $F(c)$ for all $c \in C, \Theta, C$.

Output: $\tilde{C}_T$.

begin

0. Initialize $\tilde{C}_0 = C, t = 0$;

1. *while* $|\tilde{C}_t| > \Theta$ *do*

2. Apply Graph Embedding (Ref. 32), to $F(c)$, for all $c \in \tilde{C}_t$ to obtain $S_t^{GE}(c)$;

3. Apply replicated k -means clustering on $S_t^{GE}(c)$ to obtain two stable clusters $\hat{V}_t^1, \hat{V}_t^2$;

4. Identify larger cluster $\hat{V}_t^{dom} = \arg\{\max_w |\hat{V}_t^w|\}$, where $w \in \{1, 2\}$;

5. Create set $\tilde{C}_{t+1} \subset \tilde{C}_t$ by eliminating all $c \in \hat{V}_t^{dom}$ from $\tilde{C}_t$;

6. $t = t + 1$;

7. *endwhile*;

8. $\tilde{C}_T = \tilde{C}_t$;

end

Note that in general the algorithm is terminated when the total number of spectra is marginally greater than (or equal to) Θ , which usually occurs within two to three iterations.

III.E. Stage 2: CaP identification on MRS via NLDR

Following stage 1, we attempt to apply more sophisticated analysis to the spectra in $\tilde{C}_T$ to be able to discriminate between different tissue classes in the prostate. Apart from NLDR schemes (LLE and graph embedding) that were considered, two other feature extraction schemes (z score and PCA) were also evaluated in terms of their ability to discriminate between different classes of the prostate spectra (abnormal appearing, normal appearing and indeterminate). Following feature extraction, replicated k -means clustering for the DR schemes (LLE, graph embedding, and PCA) and thresholding for z score was applied to obtain hard classification of the spectra into CaP and non-CaP categories. For z score evaluation, a predefined threshold $\mathcal{F}$ is used to classify each $c \in \tilde{C}_T$ as abnormal or normal appearing based on whether $S^z(c) > \mathcal{F}$ or $S^z(c) < \mathcal{F}$. Three commonly employed DR methods (explained previously in Sec. III) are applied to the MR spectra in $\tilde{C}_T$ so that for any $c \in \tilde{C}_T$, the high dimensional ambient feature vector $F(c)$ is mapped to $S^\phi(c)$, where

$\phi \in \{PCA, LLE, GE\}$. Replicated clustering is then employed to cluster each $S^\phi(c)$, $\forall c \in \tilde{C}_T$, into one of the three possible classes, $\hat{V}^{\phi,1}, \hat{V}^{\phi,2}, \hat{V}^{\phi,3}$ corresponding to abnormal appearing, normal appearing, or an indeterminate class, with the intermediate class corresponding to undefined, equivocal, or noisy spectra. For the manually annotated datasets, replicated clustering was used to identify all analyzable spectra as being either normal or suspicious appearing (two clusters).

IV. EVALUATION METHODS

IV.A. Identification of cancer cluster

As mentioned previously in Sec. III C 1, a potential CaP space C_P is defined within which the number of CaP locations N_g is determined. Following replicated clustering of $S^\phi(c)$, $c \in \tilde{C}_T$, we need to identify which of the k clusters $\hat{V}^1, \hat{V}^2, \dots, \hat{V}^k \in \tilde{C}_T$ correspond to the CaP class. We represent the corresponding sets of spatial locations for clusters $\hat{V}^1, \hat{V}^2, \dots, \hat{V}^k$ as $C_a^1, C_a^2, \dots, C_a^k$, respectively. With respect to C_P (as illustrated in Fig. 5), some part of C_a^q may be within ($C_{a,i}^q, q \in \{1, 2, \dots, k\}$) or outside ($C_{a,o}^q, q \in \{1, 2, \dots, k\}$) C_P . Thus $C_a^q = C_{a,i}^q \cup C_{a,o}^q$, $q \in \{1, 2, \dots, k\}$. The TP, FP, true negative (TN), and FN ratios associated with each class C_a^q , $q \in \{1, 2, \dots, k\}$ with respect to C_P and N_g are then obtained. The following heuristic algorithm is then used to identify which of $C_a^q, q \in \{1, 2, \dots, k\}$ represents the CaP class (C_a^{CaP}).

Algorithm IdentifyCaPCluster

Input: $C_a^q, q \in \{1, 2, \dots, k\}, C_P, N_g, C$.

Output: C_a^{CaP}

begin

0. for $q=1$ to k do;

1. $C_{a,i}^q = C_a^q \cap C_P$;

2. $C_{a,o}^q = C_a^q - C_{a,i}^q$;

3. If $|C_{a,i}^q| \geq N_g$ then

$TP^q = N_g, FP^q = |C_a^q| - N_g, FN^q = 0, TN = |C - C_a^q|$;

4. else $TP^q = |C_{a,i}^q|, FP = |C_{a,o}^q|, FN = N_g - |C_{a,i}^q|, TN = |C - C_{a,o}^q| - N_g$;

5. endif;

6. endfor;

7. $q^* = \arg\{\max_q [\frac{TP^q}{FP^q}]\}$; $C_a^{\text{CaP}} = C_a^{q^*}$

end

Thus Algorithm IdentifyCaPCluster identifies the CaP cluster as the one that maximizes true positive fraction while simultaneously minimizing the false positive fraction.

IV.B. Performance evaluation metrics

Having identified C_a^{CaP} , the corresponding CaP detection sensitivity and specificity values are determined as

$$SN(C_a^{\text{CaP}}) = \frac{TP_a}{TP_a + FN_a} \times 100, \quad (4.1)$$

$$SP(C_a^{\text{CaP}}) = \frac{TN_a}{TN_a + FP_a} \times 100. \quad (4.2)$$

FIG. 5. Illustration of the potential ground truth space (C_P) containing N_g metavoxels corresponding to CaP. C_a represents the CaP segmentation obtained by the MRS CAD analysis scheme and $C_{a,o}$ and $C_{a,i}$ represent those regions of C_a that lie outside and within C_P respectively.

Given that $SN(C_a^{\text{CaP}})$ and $SP(C_a^{\text{CaP}})$ are estimates of the sensitivity and specificity (given the probabilistic estimates of the CaP ground truth extent), we also compute a confidence estimate (η) associated with the performance measures as a function of $|C_P|$ and N_g . η is determined as

$$\eta = \frac{N_g}{|C_P|} \times 100. \quad (4.3)$$

Thus for each study for which we only have C_P and N_g available as surrogates of CaP ground truth, we report η along with $SN(C_a^{\text{CaP}})$ and $SP(C_a^{\text{CaP}})$ for those studies.

V. RESULTS

V.A. Qualitative results

V.A.1. Stage 1: Qualitative evaluation of the hierarchical clustering scheme

Figure 6 shows the qualitative results of the hierarchical cascade scheme for distinguishing prostatic from extracapsular spectra. Figure 6(a) represents spatial maps of the spectral grid $\tilde{C}_0$ (16×16 spectral voxels) superimposed on the corresponding T2-w MRI scene for one patient study. Every $c \in \tilde{C}_0$ in Fig. 6(a) is assigned one of two colors (blue and red) corresponding to spectra identified by the algorithm as prostatic or extracapsular. Note that the dominant cluster [spatial locations in red in Fig. 6(a)] has been eliminated in the second iteration [$\tilde{C}_1$ (16×8 spectral voxels)] [Fig. 6(b)]. The final spectral grid [$\tilde{C}_2$ in Fig. 6(c)] is obtained after elimination of extracapsular spectra (red locations) during the third iteration of the cascade. Figures 6(d)–6(f) represent the embedding plots [where each original spectral vector $F(c)$, $c \in C$ is plotted in 3D eigenvector space using the three dominant embedding values as coordinates] from $\tilde{C}_0$ (16×16 spectral voxels) (d) to $\tilde{C}_2$ (7×4 spectral voxels) (f) for one study at three different levels of the cascade. Note that at the end of the third iteration, the prostate ROI has been accurately identified and the spectral grid accurately overlaid on the prostate. Further note that in Figs. 6(a)–6(c), the spectral grid with the pronounced boundary indicates the ROI during the current iteration.

Fig. 6. Spectral grids for a single 2D slice of a T2-w MRI scene for a patient at (a) the first cascade level $\tilde{C}_0$, (b) second cascade level $\tilde{C}_1$, and (c) third cascade level $\tilde{C}_2$. Note that the size of the grid reduces from 16×16 metavoxels in (a) to 7×4 in (c) by elimination of extracapsular spectra in the dominant cluster (red). The corresponding clustered embedding plots at each of the cascade levels are also shown in (d)–(f), corresponding, in turn, to the metavoxel grids shown in (a)–(c).

V.A.2. Stage 2: Evaluation of feature extraction schemes for CaP detection

The identification of the prostate grid in stage 1 allows for the resolvability of the three MR spectral classes (abnormal appearing, normal appearing, and indeterminate). The differences between these three spectral classes within the prostate spectra cluster at the higher levels in the cascade ($\tilde{C}_1, \tilde{C}_2, \tilde{C}_3$) become discriminable only after the removal of extracapsular spectra. Note that replicated clustering was used to identify only two clusters (suspicious and normal appearing spectra) for the six manually annotated studies. Figure 7 shows the qualitative results of the four feature extraction schemes employed in this work for CaP detection for two different patient studies, each row in Fig. 7 corresponding to a different study. The three colors assigned to the spectral voxels in Fig. 7 correspond to the three clusters obtained via replicated clustering on the reduced dimensional spectra $S^{\phi(c)}$, $\phi \in \{\text{PCA}, \text{LLE}, \text{GE}\}$ for $c \in C$. For the z score scheme, each metavoxel was classified as belonging to one of the two classes [red and blue in Figs. 7(a) and 7(d)], reflecting abnormal and normal appearing spectra. The white box superposed in Figs. 7(a)–7(f) shows the potential cancer space (C_P) for corresponding slices. In each of Figs. 7(a)–7(f), the red cluster was identified as comprising abnormal using the IdentifyCaPcluster algorithm (sec. IV) and following feature extraction and replicated clustering. Figures 7(a) and 7(d) show the results obtained via z score; while Figs. 7(b) and 7(e) show the results of PCA in identifying abnormal appearing, normal appearing, and indeterminate spectra. Figures

7(c) and 7(f) show the corresponding results for graph embedding and LLE, respectively. For the first study (first row in Fig. 7), graph embedding [Fig. 7(c)] appears to yield a near perfect CaP detection in terms of sensitivity and specificity as only CaP voxels are identified within the white cancer grid (C_P). The corresponding results for z score [Fig. 7(a)] and PCA [Fig. 7(b)] both yield poor detection sensitivity and specificity. Similarly, for the second study shown in Fig. 7, LLE [Fig. 7(f)] appears to yield higher CaP detection sensitivity and specificity compared to z score [Fig. 7(d)], and PCA [Fig. 7(e)]. Figure 8 shows an example of the MR spectral grid with classification labels obtained from graph embedding and replicated clustering plotted back on the individual spectra; the spectra in red corresponding to those identified as abnormal appearing, blue spectra corresponding to normal appearing, and green spectra corresponding to indeterminate.

V.B. Quantitative results

V.B.1. Quantitative evaluation of stage 1: Hierarchical clustering

At the end of stage 1, the largest rectangular box $\tilde{C}_T$ that contains all prostate spectra is then overlaid on the T2-w image. Note that in quantitative evaluation of stage 1, the precise spatial extent of the prostate is all that is needed. This ground truth is ascertained by manual placement of a spectral grid ($\tilde{C}_g^T$) on the prostate by an expert radiologist. Table II

FIG. 7. Qualitative comparison of prostate MRS analysis via four different feature extraction schemes employed for two different studies; each row corresponding to a different study. (a) and (d) represent the suspicious appearing (red voxels) and benign appearing (blue voxels) spectra identified via z score, and (b) and (e) illustrate the clustering results obtained via PCA [red (suspicious appearing), blue (benign appearing), green (intermediate)]. Results for the NLDR schemes (c) GE and (f) LLE are also shown [the red, blue, and green colors having the same meaning as for PCA in (b) and (e)]. The white box superposed on (a)–(f) shows the locations of the potential cancer space (C_p). In each of (a)–(f) the cluster with the red metavoxels was the one identified as the CaP class based on the IdentifyCaPcluster algorithm (Sec. V A).

shows the average sensitivity, specificity, and PPV in automated identification of the prostate grid $\tilde{C}_T$ with respect to $\tilde{C}_g^T$ and averaged over 18 studies.

V.B.2. Quantitative evaluation of stage 2: Identifying suspicious appearing spectra

Evaluation of z score via receiver operating characteristic (ROC) curve analysis. In order to define the optimal threshold θ^z for performing a z score-based classification of each $c \in C$ as normal or suspicious, a set of cancerous voxels obtained via annotation by an expert radiologist, $C^V \subset C$, was defined during an offline training phase. For each $c \in C^V$, a corresponding $S^z(c)$ was obtained which was then used to define the average μ^z and standard deviation σ^z of z score values for CaP. The threshold θ^z was then defined as $\mu^z \pm \alpha \sigma^z$, where $\alpha \in [0, 1]$. The value of α was uniformly varied between $[0, 1]$ and a corresponding values for θ^z obtained. At each value of θ^z , each metavoxel $c \in \tilde{C}_T$ is identified as belonging to CaP if $S^z(c) > \theta^z$, normal otherwise. Thus at each θ^z , the corresponding sensitivity and specificity of CaP detection via z score is computed [Eqs. (4.12) and (4.13)] with respect to C_p and N_g . A curve is then fit to the sensitivity, specificity values to obtain the ROC curve. The optimal threshold $\hat{\theta}^z$ was determined as the operating point on the ROC curve; the location on the ROC curve closest to 100% sensitivity, specificity. Average sensitivity and specificity at the operating point of ROC curve were found to be 72.85% and 65.45%, respectively.

Evaluation of DR methods via cross validation using

probabilistic ground truth for CaP extent. The ability of a classifier to distinguish between object classes in embedding spaces obtained via DR methods is dependent on to the choice of the number of dimensions (v) of the embedding space in which the data are represented. For the NLDR methods, the low dimensional data representations obtained via LLE (Ref. 30) is also a function of κ , the parameter controlling the size of the local neighborhood within which linearity is assumed. In order to evaluate the parameter sensitivity of different DR methods, the robustness of the DR methods over different values of κ and v was quantitatively evaluated. LLE was evaluated by varying $\kappa \in \{6, 7, \dots, 15\}$ and $v \in \{3, 4, \dots, 10\}$, a total of 80 different combinations of parameter values. PCA and graph embedding were evaluated for eight different values of $v \in \{3, 4, \dots, 10\}$.

For each of $\phi \in \{\text{LLE, GE, PCA}\}$, threefold and fivefold cross validation averaged over 20 iterations were also performed to obtain average sensitivity and specificity in terms of CaP detection performance for all 18 studies. Threefold cross validation was performed by randomly choosing three datasets and calculating average CaP detection sensitivity and specificity across three studies on the 80-dimensional parameter space. The parameter set ($v_{\max}$ for PCA, GE and $v_{\max}, \kappa_{\max}$ for LLE) with maximum sensitivity and specificity in this space were then identified as optimal values and used for the CAD scheme on the remaining 15 studies. The average CaP detection sensitivity and specificity on these 15 studies was then recorded. On the next trial, 3 random training studies from 18 were again selected and used to optimize the parameters and evaluation again was done on the remain-

FIG. 8. MRS spectral grid plotted with the classification labels (three colors correspond to three different clusters) obtained from graph embedding and replicated clustering. The red spectra correspond to those identified as abnormal appearing, the blue correspond to normal appearing, and the green correspond to indeterminate spectra (possibly noise in this case).

ing 15 studies. This entire process was repeated for a total of 20 times. The mean μ^ϕ and standard deviation σ^ϕ in CaP detection sensitivity and specificity across these 20 iterations were reported in Table III(a) for $\phi \in \{\text{LLE}, \text{GE}, \text{PCA}\}$. A similar routine was employed when performing fivefold cross validation. The corresponding results are reported in Table III(b). All NLDR schemes employed in this work were found to have higher sensitivity of CaP detection compared to PCA; graph embedding performing the best with a sensitivity of almost 90%. In Figs. 9 and 10 barplots are shown representing CaP detection sensitivity and specificity (μ^ϕ, σ^ϕ) for each of the 18 studies considered in this work via threefold and fivefold cross validation, respectively, over 20 iterations for GE. The confidence estimates (η) associated with the sensitivity and specificity measurements of each

study are also shown. Note that while η is low for a majority of the studies, for studies 11 and 18, the confidence associated with CaP detection sensitivity and specificity for graph embedding were almost 90% and 80%, respectively.

Evaluation of all feature extraction methods against expert annotated spectra. Table IV shows the average sensitivity and specificity results obtained for $\phi \in \{\text{LLE}, \text{GE}, \text{PCA}\}$,

TABLE II. Average sensitivity, specificity, PPV values for automated identification of prostate grid using hierarchical spectral clustering averaged over 18 studies.

Sensitivity	Specificity	PPV
97.66%	98.87%	89.29%

TABLE III. This shows the average and standard deviation in CaP detection sensitivity and specificity for different feature extraction methods over $\kappa \in \{5, \dots, 15\}$ and $v \in \{3, \dots, 10\}$ for 18 different studies via (a) threefold cross validation, and (b) fivefold cross validation.

Method	Sensitivity	Specificity
(a)		
GE	89.33 ± 1.87	79.79 ± 2.24
LLE	83.70 ± 4.50	81.04 ± 4.60
PCA	78.53 ± 3.20	83.97 ± 2.81
(b)		
GE	89.09 ± 2.10	79.22 ± 2.10
LLE	84.93 ± 4.42	81.32 ± 4.05
PCA	78.78 ± 4.31	84.11 ± 2.46

FIG. 9. Barplot showing average and standard deviation in CaP detection sensitivity and specificity for the individual 18 studies averaged over 20 iterations of threefold cross validation via graph embedding. The coefficient of confidence (η) associated with the estimation of the performance measures is also shown in green.

z } for six studies for which each spectra had been labeled according to the five point scale³⁶ by an expert radiologist. For all methods, we made the assumption that spectra labeled as 1/2 were all normal appearing and the spectra labeled 3/4/5 were suspicious appearing. For each of $\phi \in \{\text{LLE}, \text{GE}, \text{PCA}, z\}$, the IdentifyCaPCluster algorithm was employed to identify the cluster most likely to correspond to CaP spectra. Since we only had six studies available, the mean and standard deviation in CaP detection sensitivity and specificity across all the expert annotated datasets were determined (based on previously determined $v_{\max}$ and $\kappa_{\max}$) and shown in Table IV.

VI. CONCLUDING REMARKS

In this paper we have presented a novel application of nonlinear dimensionality reduction and hierarchical clustering for automated identification of (a) the prostate ROI based on the classification of MR spectral data alone and (b) suspicious appearing spectra within the prostate ROI. A total of 18 MRI/MRS studies from the ACRIN trial was considered for the evaluation of four different feature extraction algorithms (PCA, z score, LLE, GE) in conjunction with replicated clustering in terms of their ability to identify suspicious appearing spectra. Owing to the fact that only limited knowledge regarding precise spatial extent of CaP was available for the studies considered in this work, we defined a probabilistic ground truth estimate for CaP and a confidence coefficient to assess the degree of certainty associated with the CAD performance measures reported. The high confidence estimates associated with two of the studies (study 11 and 18) seem to suggest that the consistently high CaP detection sensitivity and specificity measurements for the other studies are not erroneous. For a subset (6) of the 18 studies, a radi-

TABLE IV. This shows the average and standard deviation in CaP detection sensitivity and specificity for different feature extraction methods obtained via comparison with expert annotations for six ACRIN studies.

Method	Sensitivity	Specificity
GE	79.74 ± 13.18	65.49 ± 7.58
LLE	81.36 ± 5.69	64.71 ± 9.84
PCA	76.68 ± 14.07	48.26 ± 12.95
z score	62.70 ± 18.67	50.73 ± 7.39

FIG. 10. Barplot showing the average and standard deviation in CaP detection sensitivity and specificity for the individual 18 studies averaged over 20 iterations for fivefold cross validation via graph embedding. The coefficient of confidence (η) associated with the estimation of the performance measures is also shown in green.

ologist laboriously annotated each of the spectra according to the five point scale. For these studies, we assumed that the spectra corresponding to 3/4/5 were cancerous. In comparing the four feature extraction schemes on the 18 datasets with partial CaP ground truth estimates, as well as the six studies for which expert annotations was available, the NLDR schemes (GE and LLE) consistently outperformed PCA and z score in terms of both CaP detection sensitivity and specificity. In addition, the nonlinear DR schemes were found to be relatively robust to change in the value of the system parameters (κ, v). The use of the replicated clustering scheme helped overcome the instability associated with k -means clustering, yielding consistently stable clusters. Our scheme is also highly efficient with stage 1 only requiring an average of 11.24 s to analyze a $256 \times 256 \times 8$ MRI/MRS 3D 1.5 T scene on a Pentium IV, 2 Gbyte RAM Intel processor machine. Stage 2 took an average of 5.23 s for automated classification of MR spectra as being suspicious, normal appearing, or indeterminate.

Future work will involve integration of our automated CAD MRS scheme with T2-w MRI to incorporate both structural and functional information for more accurate identification of CaP. We also aim to perform more rigorous analysis of the scheme on a larger cohort of data. The availability of more precise knowledge of spatial location of CaP on the MR imagery will help to further confirm and validate the efficacy of our methods.

While replicated clustering is more stable than k -means, it still requires specification of the value of k . In this work we assumed $k=3$ based on the assumption that all the spectra were either normal appearing, suspicious appearing, or indeterminate. While Jung *et al.*,³⁶ identified five distinct classes of prostate MR spectra (five point scale), in our experiments (owing perhaps to the quality of data) we were unable to find five unique clusters. A future avenue of exploration in the future will be to look at fully unsupervised clustering schemes (e.g., mean shift) which do not require prespecification of the number of data clusters.

ACKNOWLEDGMENTS

This work was made possible via grants from the Wallace H. Coulter Foundation, New Jersey Commission on Cancer Research, National Cancer Institute (Grant Nos.

R01CA136535-01, ARRA-NCI-3 R21CA127186-02S1, R21CA127186-01, and R03CA128081-01), the Society for Imaging Informatics in Medicine (SIIM), The Cancer Institute of New Jersey, and the Life Science Commercialization Award from Rutgers University.

^{a)} Author to whom correspondence should be addressed: 599 Taylor Road, Piscataway, NJ 08854. Telephone: 732-445-4500 (ext. 6213). Fax: 732-445-3753. Electronic mail: anantm@rci.rutgers.edu

¹W. Catalona, D. Smith, T. Ratliff, K. Dodds, D. Coplen, J. Yuan, J. Petros, and G. Andriole, "Measurement of prostate-specific antigen in serum as a screening test for prostate cancer," *N. Engl. J. Med.* **324**, 1156–1161 (1991).

²P. Borboroglu, S. Comer, R. Riffenburgh, and C. Amling, "Extensive repeat transrectal ultrasound guided prostate biopsy in patients with previous benign sextant biopsies," *J. Urol. (Paris)* **163**(1), 158–162 (2000).

³Kyle K. Yu, and H. Hricak, "Imaging prostate cancer," *Radiol. Clin. North Am.* **38**(1), 59–85 (2000).

⁴J. Scheidler, H. Hricak, D. Vigneron, K. Yu, D. Sokolov, R. Huang, C. Zaloudek, S. Nelson, P. Carroll, and J. Kurhanewicz, "Prostate cancer: Localization with three-dimensional proton MR spectroscopic imaging-clinicopathologic study," *Radiology* **213**, 473–480 (1999).

⁵P. Carroll, F. Coakley, and J. Kurhanewicz, "Magnetic resonance imaging and spectroscopy of prostate cancer," *Rev. Neurol.* **8**(1), S4–S10 (2006).

⁶G. Ullrich, Mueller-Lisse, M. Swanson, D. Vigneron, H. Hricak, A. Bessette, R. Males, P. Wood, S. Noworolski, S. Nelson, I. Barken, P. Carroll, and J. Kurhanewicz, "Time-dependent effects of hormone-deprivation therapy on prostate metabolism as detected by combined magnetic resonance imaging and 3D magnetic resonance spectroscopic imaging," *Magn. Reson. Med.* **46**, 49–57 (2001).

⁷J. Kurhanewicz, M. Swanson, S. Nelson, and D. Vigneron, "Combined magnetic resonance imaging and spectroscopic imaging approach to molecular imaging of prostate cancer," *J. Magn. Reson. Imaging* **16**, 451–463 (2002).

⁸A. Heerschap, G. J. Jager, M. van der Graaf, J. O. Barentsz, J. Rosetten, G. De La Oosterhof, E. Ruijter, and J. Ruijs, "In vivo proton MR spectroscopy reveals altered metabolite content in malignant prostate tissue," *Anticancer Res.* **17**, 1455–1460 (1997).

⁹J. Kurhanewicz, D. B. Vigneron, H. Hricak, P. Narayan, P. Carroll, and S. J. Nelson, "Three-dimensional H-1 MR spectroscopic imaging of the in situ human prostate with high (0.24-0.7-cm³) spatial resolution," *Radiology* **198**(3), 795–805 (1996).

¹⁰K. L. Zakian, S. Eberhardt, H. Hricak, A. Shukla-Dave, S. Kleinman, M. Muruganandham, K. Sircar, M. W. Kattan, V. E. Reuter, P. T. Scardino, and J. A. Koutcher, "Transition zone prostate cancer: Metabolic characteristics at H MR spectroscopic imaging—initial results," *Radiology* **229**, 241–247 (2003).

¹¹A. Wetter, T. Engl, D. Nadjmabadi, K. Fliessbach, T. Lehnert, J. Gurung, W. Beecken, and J. Vogl, "Combined MRI and MR spectroscopy of the prostate before radical prostatectomy," *AJR, Am. J. Roentgenol.* **187**, 724–730 (2006).

¹²A. Madabhushi and D. Metaxas, "Ultrasound techniques in digital mammography and their application in breast cancer diagnosis," *Medical Imaging Systems Technology: Analysis and Computational Methods* (World Scientific, Singapore, 2005), pp. 119–150.

¹³A. Madabhushi, M. Feldman, D. Metaxas, J. Tomaszewski, and D. Chute, "Automated detection of prostatic adenocarcinoma from high-resolution ex vivo MRI," *IEEE Trans. Med. Imaging* **24**(12), 1611–1625 (2005).

¹⁴S. Viswanath, M. Rosen, and A. Madabhushi, "A consensus embedding approach for segmentation of high resolution in vivo prostate magnetic resonance imagery," *Society of Photo-Optical Instrumentation Engineers Medical Imaging Conference (SPIE 2008)*, San Diego (SPIE, 2008), Vol. 6915, 69150U.

¹⁵I. Chan, W. Wells, R. Mulkern, S. Haker, J. Zhang, K. Zou, S. Maier, and C. Tempany, "Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging: A multichannel statistical classifier," *Med. Phys.* **30**(9), 2390–2398 (2003).

¹⁶J. Kurhanewicz, D. B. Vigneron, R. G. Males, M. G. Swanson, K. K. Yu, and H. Hricak, "The prostate: MR imaging and spectroscopy. Present and future," *Radiology* **38**(1), 115–138 (2000).

¹⁷J. W. C. Van der Veen, R. De Beer, P. R. Luyten, and D. Van Ormondt, "Accurate quantification of in vivo 31P NMR signals using the variable projection method and prior knowledge," *Magn. Reson. Med.* **6**(1), 92–98 (1988).

¹⁸L. Vanhamme, A. Boogaart, and S. Huffel, "Improved method for accurate and efficient quantification of MRS data with use of prior knowledge," *J. Magn. Reson.* **129**(1), 35–43 (1998).

¹⁹H. Ratiney, M. Sdika, Y. Coenradie, S. Cavassila, D. van Ormondt, and D. Graveron-Demilly, "Time-domain semiparametric estimation based on a metabolite basis set," *Pediatr. Nephrol.* **18**(1), 1–13 (2005).

²⁰K. Zakian, K. Sircar, H. Hricak, H. Chen, A. Dave, S. Eberhardt, M. Muruganandham, L. Ebor, M. Kattan, V. Reuter, P. Scardino, and J. Koutcher, "Correlation of proton MR spectroscopic imaging with gleason score based on sept-section pathologic analysis after radical prostatectomy," *Radiology* **234**, 804–814 (2005).

²¹P. Pels, E. Ozturk-Isik, M. Swanson, L. Vanhamme, J. Kurhanewicz, S. Nelson, and S. Huffel, "Quantification of prostate MRSI data by model-based time domain fitting and frequency domain analysis," *Pediatr. Nephrol.* **19**, 188–197 (2006).

²²M. Kelm, B. Menze, C. Zechmann, K. Baudendistel, and F. Hanprecht, "Automated estimation of tumor probability in prostate MRSI: Pattern recognition vs. quantification," *Magn. Reson. Med.* **57**, 150–159 (2006).

²³H. Quan and S. Bao, "z score analysis for H-MRSI data of glioma and prostate cancer," *Asia-Oceania Federation of Organizations for Medical Physics*, 2004.

²⁴T. Laudadio, P. Pels, L. Lathauwer, P. Hecke, and S. Huffel, "Tissue segmentation and classification of MRSI data using canonical correlation analysis," *Magn. Reson. Med.* **54**, 1519–1529 (2005).

²⁵A. Devos, L. Lukas, J. A. K. Suykens, L. Vanhamme, A. R. Tate, F. A. Howe, C. Majas, A. Moreno-Torres, M. van der Graaf, C. Ars, and S. Van Huffel, "Classification of brain tumours using short echo time 1H MR spectra," *J. Magn. Reson. Imaging* **17**, 164–175 (2004).

²⁶J. Ma and J. Sun, "MRS classification based on independent component analysis and support vector machines," *IEEE Fifth International Conference on Hybrid Intelligent Systems (IEEE, Nov. 2005)*, pp. 81–84.

²⁷A. W. Simonetti, W. J. Melssen, F. Szabo de Edelenyi, J. J. van Asten, A. Heerschap, and L. M. Buydens, "Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification," *NMR Biomed.* **18**, 34–43 (2005).

²⁸G. Lee, A. Madabhushi, and C. Rodriguez, "Investigating the efficacy of nonlinear dimensionality reduction schemes in classifying gene and protein expression studies," *IEEE/ACM Trans. Comput. Biol. Bioinf.* **5**, 368–384 (2008).

²⁹J. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science* **290**, 2319–2322 (2000).

³⁰S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science* **290**, 2323–2326 (2000).

³¹A. Madabhushi, J. Shi, M. Rosen, J. Tomaszewski, and M. Feldman, "Graph embedding to improve supervised classification: Detecting prostate cancer," *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2005)*, LNCS, 2005, Vol. 3749, pp. 729–737.

³²J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(8), 888–905 (2000).

³³H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.* **24**, 417–441 (1933).

³⁴J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, 1967), Vol. 1, pp. 281–297.

³⁵P. A. Bottomley, "Spatial localization in NMR spectroscopy in vivo," *Ann. N.Y. Acad. Sci.* **508**, 333–348 (1997).

³⁶J. Jung, F. Coakley, D. Vigneron, M. Swanson, A. Qayyum, V. Weinberg, K. Jones, P. Carroll, and J. Kurhanewicz, "Prostate depiction at endorectal MR spectroscopic imaging: investigation of a standardized evaluation system," *Radiology* **233**, 701–708 (2004).