Computer extracted texture features on T2w MRI to predict biochemical recurrence following radiation therapy for prostate cancer

Shoshana B. Ginsburg[a], Mirabela Rusu[a], John Kurhanewicz[b], and Anant Madabhushi[a]

[a]Case Western Reserve University, Cleveland, Ohio
[b]University of California, San Francisco, California

ABSTRACT

In this study we explore the ability of a novel machine learning approach, in conjunction with computer–extracted features describing prostate cancer morphology on pre–treatment MRI, to predict whether a patient will develop biochemical recurrence within ten years of radiation therapy. Biochemical recurrence, which is characterized by a rise in serum prostate–specific antigen (PSA) of at least 2 ng/mL above the nadir PSA, is associated with increased risk of metastasis and prostate cancer–related mortality. Currently, risk of biochemical recurrence is predicted by the Kattan nomogram, which incorporates several clinical factors to predict the probability of recurrence–free survival following radiation therapy (but has limited prediction accuracy). Semantic attributes on T2w MRI, such as the presence of extracapsular extension and seminal vesicle invasion and surrogate measurements of tumor size, have also been shown to be predictive of biochemical recurrence risk. While the correlation between biochemical recurrence and factors like tumor stage, Gleason grade, and extracapsular spread are well–documented, it is less clear how to predict biochemical recurrence in the absence of extracapsular spread and for small tumors fully contained in the capsule. Computer–extracted texture features, which quantitatively describe tumor micro–architecture and morphology on MRI, have been shown to provide clues about a tumor's aggressiveness. However, while computer–extracted features have been employed for predicting cancer presence and grade, they have not been evaluated in the context of predicting risk of biochemical recurrence. This work seeks to evaluate the role of computer–extracted texture features in predicting risk of biochemical recurrence on a cohort of sixteen patients who underwent pre–treatment 1.5 Tesla (T) T2w MRI. We extract a combination of first–order statistical, gradient, co–occurrence, and Gabor wavelet features from T2w MRI. To identify which of these T2w MRI texture features are potential independent prognostic markers of PSA failure, we implement a partial least squares (PLS) method to embed the data in a low–dimensional space and then use the variable importance in projections (VIP) method to quantify the contributions of individual features to classification on the PLS embedding. In spite of the poor resolution of the 1.5 T MRI data, we are able to identify three Gabor wavelet features that, in conjunction with a logistic regression classifier, yield an area under the receiver operating characteristic curve of 0.83 for predicting the probability of biochemical recurrence following radiation therapy. In comparison to both the Kattan nomogram and semantic MRI attributes, the ability of these three computer–extracted features to predict biochemical recurrence risk is demonstrated.

Keywords: Prostate cancer, biochemical recurrence, texture analysis, T2w MRI

Send correspondence to A.M.: anant.madabhushi@case.edu.

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under award numbers R01CA136535-01, R01CA140772-01, and R21CA167811-01; the National Institute of Diabetes and Digestive and Kidney Diseases under award number R01DK098503-02; the DOD Prostate Cancer Synergistic Idea Development Award (PC120857); the QED award from the University City Science Center and Rutgers University; the Ohio Third Frontier Technology Development Grant; and the National Science Foundation Graduate Research Fellowship. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Medical Imaging 2014: Computer-Aided Diagnosis, edited by Stephen Aylward, Lubomir M. Hadjiiski,
Proc. of SPIE Vol. 9035, 903509 · © 2014 SPIE · CCC code: 1605-7422/14/$18 · doi: 10.1117/12.2043937

1. INTRODUCTION

Approximately 240,000 new prostate cancers are diagnosed annually in the United States, and nearly 30,000 deaths are attributed to prostate cancer on an annual basis.[1] Primary localized treatment options include radical prostatectomy and radiation therapy, which can be administered as external beam radiation therapy (EBRT) or brachytherapy with or without concomitant hormonal therapy. Nevertheless, as many as one third of patients who undergo localized treatment for prostate cancer develop biochemical recurrence,[2] which, according to the ASTRO definition,[3] is characterized by a rise in serum prostate–specific antigen (PSA) of at least 2 ng/mL above the nadir PSA. Approximately one third of prostate cancers associated with biochemical recurrence will eventually metastasize.[2] Furthermore, biochemical recurrence is associated with a marginally but significantly increased risk of clinical progression and prostate cancer–related mortality.[4] Consequently, pre–treatment prediction of a patient's propensity towards biochemical recurrence following radiation therapy would be beneficial for choosing an optimal treatment strategy. Thus, patients who are likely to experience biochemical recurrence after undergoing radiation therapy may be advised to choose other treatments.

Several nomograms have been developed for predicting patient response to radiation therapy; the most widely accepted one is the Kattan nomogram.[5] The Kattan nomogram incorporates baseline PSA level, tumor stage, Gleason score, radiation dose, and the use of hormonal therapy to predict the probability of remaining recurrence–free for five years. Although these factors are good predictors of biochemical recurrence, the Kattan nomogram is limited by the presence of benign prostatic hyperplasia, which impacts pretreatment PSA levels, and inaccuracies in the determination of the Gleason score that result from biopsy sampling errors.

During the past decade, magnetic resonance imaging (MRI) has emerged as an accurate method for evaluating prostate cancer stage and grade.[6–9] Prostate MRI exams typically involve some combination of T2w MRI, which provides excellent contrast of anatomic structures; diffusion weighted imaging and dynamic contrast enhanced (DCE) MRI, which provide information about cell architecture and microvasculature; and MR spectroscopy, which provides insight into metabolic activity. Due to the ability to discern anatomical structures on T2w MRI, semantic attributes of prostate cancer, such as the presence of extracapsular extension and seminal vesicle invasion and surrogate measurements of tumor size, can be assessed on T2w MRI. These semantic attributes have all been shown to be powerful independent predictors of five–year biochemical recurrence–free survival.[10–12] Moreover, when incorporated in a predictive model together with the clinical features assessed by the Kattan nomogram, these semantic or qualitative MRI features obtained via a radiologist's interpretation can substantially augment the accuracy of the Kattan nomogram in predicting biochemical recurrence risk, leading to an area under the receiver operating characteristic (ROC) curve (AUC) of 0.78 compared to 0.61 for the Kattan nomogram.[10]

Nevertheless, semantic features represent only a small fraction of possible MRI–derived attributes that may be useful for predicting biochemical recurrence. A number of recent studies have shown that computer–extracted features, such as textural features derived from T2w MRI, are useful for detecting prostate cancer[13–17] and differentiating between high and low Gleason grade[18] prostate cancer *in vivo*.[19] It is thought that computer–extracted features are capable of quantitatively describing tumor micro–architecture and morphology, which provide clues about tumor aggressiveness. Because high grade tumors tend to be more aggressive, they are more likely to trigger biochemical recurrence after treatment. Consequently, we expect that micro–architectures in the tumor may be correlated with biochemical recurrence risk, as well as Gleason grade. Since computer–extracted MRI texture features were previously shown to effectively discriminate between prostate cancers of high and low Gleason grades,[19] we hypothesize that computer–extracted MRI texture features may capture morphometric clues for predicting biochemical recurrence as well.

However, one of the challenges associated with trying to find a set of computer–extracted features associated with risk of ten–year biochemical recurrence is dealing with the "curse of dimensionality" problem—an issue when the number of computer–extracted features overwhelms the total number of available samples. While it is important to extract a large pool of features that comprehensively describe tissue morphology and micro-architecture on MRI, many of these computer-extracted features may be redundant or may have limited value for predicting biochemical recurrence of prostate cancer. Thus, it is crucial that we identify the few MRI–derived features that are most relevant for predicting biochemical recurrence.

The traditional approach for identifying a small set of highly predictive features is feature subset selection.[20] Nevertheless, feature subset selection is a challenging problem because (a) most feature ranking processes are

unable to account for dependencies and interactive effects among features and (b) both feature ranking and greedy feature selection algorithms tend to be highly unstable, as several feature subsets may lead to the same high classification accuracy.[20] An alternate method for identifying top–performing features would involve constructing a classifier using all of the features and then identifying the few features that have the greatest positive impact on this classifier. Unfortunately, constructing a classifier with hundreds of features is unfeasible due to the Hughes effect,[21] which dictates that predictive power reduces as dimensionality increases. Conversely, dimensionality reduction approaches (e.g., principal component analysis) involve transforming the high–dimensional features into a series of eigenvectors, which are then used for classification. Because very few eigenvectors are needed to describe most of the variance in the original features, dimensionality reduction succeeds in reducing the dimension of the data so that constructing a generalizable classifier with high predictive power becomes feasible. The most salient drawback of dimensionality reduction, however, is that these eigenvectors typically have no direct biological meaning. However, in predictive modeling in biomedical domains, understanding which features are predictive can be as important as overall prediction accuracy.

The primary goal of this work is to construct a predictive model that uses a small number of computer–extracted T2w MRI features to predict biochemical recurrence risk following radiation therapy. Toward this end, we extract a combination of first–order statistical features, which highlight regions with large changes in signal intensity; non–steerable gradient features, which highlight tumor boundaries; co–occurrence features, which highlight homogeneous regions of low signal intensity in aggressive cancer; and Gabor wavelet features, which consider multi–scale and multi–orientation architectures within the tumor. Then, partial least squares (PLS) is implemented to reduce the dimensionality of the high–dimensional MRI features. We leverage the variable importance in projections (VIP) scoring system[22] to score and rank each computer–extracted feature according to the extent of its contribution to accurate prediction of biochemical recurrence in the PLS embedding subspace. Finally, the few features associated with the highest VIP scores are used to construct a predictive model that is compared to the Kattan nomogram in terms of its ability to accurately distinguish between patients who will develop biochemical recurrence and those who will remain recurrence–free. A summary of our approach is illustrated in Figure 1.

The remainder of this paper is organized as follows. Section 2 elaborates on previous work in computer–aided characterization of prostate cancer and highlights the novel contributions of our work. In Section 3 we discuss our methodology for feature extraction and feature selection via the VIP method, and in Section 4 we focus on our experimental design and methods for evaluating the features identified by VIP. Section 5 discusses our experimental results, and Section 6 contains our concluding remarks.

2. PREVIOUS WORK AND NOVEL CONTRIBUTIONS

A number of studies have explored methods for computer–aided detection (CAD) and localization of prostate cancer on MRI.[9,13–17,23–26] Most of these studies exploit multiparametric MRI, including T2w, dynamic contrast–enhanced (DCE), and diffusion–weighted (DW) MRI protocols.[9,23] For example, several groups[24,25] have reported that combining apparent diffusion coefficient (ADC) maps obtained from DW MRI with T2w MRI improves prostate cancer detection when compared to T2w MRI alone, and Ampeliotis et al.[26] reported a statistically significant improvement in cancer detection accuracy when signal intensities from DCE and T2w MRI were combined, compared to the use of individual MRI protocols. Furthermore, a number of recent studies have shown that computer–extracted features describing textures on MRI are useful for detecting prostate cancer.[13–17] These studies use a combination of texture features extracted from T2w MRI and quantitative features extracted from MR spectroscopy, DW MRI, or DCE MRI, and they report improved prostate cancer detection and localization accuracy compared to MRI signal intensities. While most of these studies exploit multiparametric MRI, which is more informative of prostate cancer presence than T2w MRI alone, some have shown that it is possible to construct effective CAD classifiers for prostate cancer localization using T2w MRI alone.[16]

More recently a handful of studies have explored using multiparametric MRI for predicting prostate cancer grade, in addition to location and extent.[6–9] Prostate cancers associated with higher Gleason grades have been shown to be correlated with lower T2w MRI signal intensities[6] and lower ADC values on DW MRI,[7] and quantitative features extracted from ADC maps and DCE MRI are moderately correlated with Gleason scores.[8,9] Furthermore, several clinical observational studies have explored the use of pre–treatment MRI to

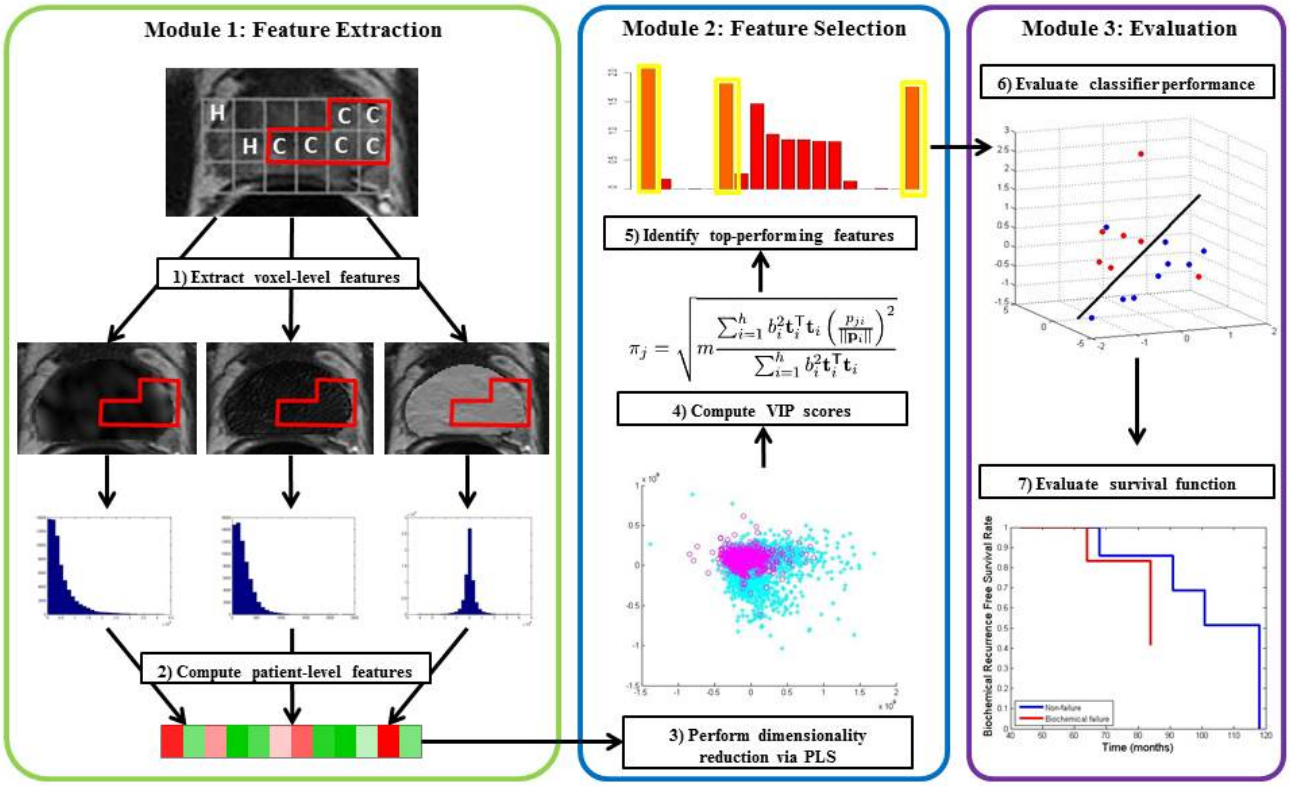


Figure 1: Flowchart illustrating methodology used in this paper for extraction, scoring, ranking, and evaluation of computer–extracted features for predicting biochemical recurrence risk.

predict biochemical recurrence following treatment.[10,12,27] For example, predictive models that exploit attributes on T2w MRI, such as tumor size, presence and extent of extracapsular spread, and presence or absence of seminal vesicle invasion, were shown to outperform the Kattan nomogram in predicting progression to biochemical recurrence following radiation therapy.[10] However, none of these studies used computer–extracted MRI features to predict risk of biochemical recurrence.

Although computer–extracted texture features have previously been shown to be useful for prostate cancer detection, localization, and grading,[6–9,9,13–17,23–26] to the best of our knowledge no one has considered looking at the association between computer–extracted texture features on pre–treatment T2w MRI and subsequent biochemical recurrence. The underlying hypothesis of our study is that computer–extracted texture features, which quantitatively describe tumor micro–architecture and morphology on T2w MRI, can provide clues about a tumor's aggressiveness and hence its likelihood to recur after treatment. This hypothesis is tested on sixteen patients with biopsy–confirmed prostate cancer who underwent EBRT or brachytherapy and subsequently participated in at least five years of follow–up. Among these patients, seven experienced biochemical recurrence within ten years of treatment, while the remaining nine remained recurrence–free for at least five years. Our aim is to identify a set of computer–extracted features from T2w MRI that can complement the performance of the Kattan nomogram, which predicts biochemical recurrence risk based on clinical variables alone. More specifically, we seek computer–extracted features that can be gleaned from old 1.5 Tesla T2w MRI, which suffers from poor spatial resolution, because of the long term outcome information available along with these fifteen–year–old MRI exams.

Feature	No.	Parameters	Description	Motivation
First–order statistical	8	window size	Mean, standard deviation, and range of intensities	Localize regions with significant changes in signal intensity
Non–steerable gradient	13	window size	Image convolutions with Sobel and Kirsch operators	Accurately detect region boundaries
Co-occurrence	13	distance, window size	Statistical features computed from joint pdf of intensity value co–occurrences	Differentiate homogeneous regions of low SI in aggressive cancer from high SI in normal prostate
Gabor wavelet	54	orientation, wavelength	Multi–orientation features computed from Gaussian function convolved with sinusoid	Quantify local and global features qualitatively assessed by radiologists analyzing cancer

Table 1: Description of features computed on a per–voxel basis from T2w MRI and the motivation for using them for predicting biochemical recurrence of prostate cancer. SI = signal intensity.

3. FEATURE EXTRACTION AND SELECTION

3.1 Feature Extraction from T2w MRI

We define a voxel in the T2w MR image as $c \in \mathbf{C}$, where $\mathbf{C}$ is a 3–dimensional grid of MRI voxels. Each $c \in \mathbf{C}$ is associated with a label $\mathbf{l}(c) \in \{0, 1\}$, where $\mathbf{l}(c) = 1$ if voxel c is cancerous and $\mathbf{l}(c) = 0$ otherwise. Let $\mathbf{C}^1 = \{c : \mathbf{l}(c) = 1\}$. Each voxel in $\mathbf{C}^1$ is also associated with the feature vector $\mathbf{F}(c)$, which is comprised of computer–extracted features extracted from T2w MRI. A brief description of the image texture features extracted in this work, as well as our motivation for using them to predict biochemical recurrence, is provided in Table 1.

Following voxel–level extraction of texture features from cancerous prostate voxels (see Figure 2), patient–level features were subsequently computed by calculating the mean, median, standard deviation, skewness, and kurtosis of the distributions of each of these features. Thus, for $j \in \{1, ..., 88\}$ the mean value of the jth feature was calculated as $\frac{1}{|\mathbf{C}^1|} \sum_{c \in \mathbf{C}^1} \mathbf{F}_j(c)$, where $| * |$ denotes cardinality. In additional to the mean, the median, standard deviation, skewness and kurtosis were similarly computed for each of the 88 features, together providing a single, patient–level feature vector $\mathbf{x}$ containing 440 features.

3.2 Dimensionality Reduction via Partial Least Squares

In order to reduce the dimensionality of the data, we use partial least squares (PLS),[28] a popular linear dimensionality reduction method. In contrast to the more well–known principal components analysis (PCA), an unsupervised dimensionality reduction scheme that seeks to maximize data covariance, PLS is a supervised dimensionality reduction method that maximizes data covariance specifically as it relates to a response variable. Given an $n \times m$ data matrix $\mathbf{X}$ and an $n \times 1$ outcome vector $\mathbf{y}$, PLS forms the following model:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T, \ \mathbf{y} = \mathbf{U}\mathbf{q}^T, \tag{1}$$

where $\mathbf{T}$ is an $n \times h$ matrix of X–scores, $\mathbf{P}$ is an $m \times h$ matrix of X–loadings, $\mathbf{U}$ is an $n \times h$ matrix of Y–scores, and $\mathbf{q}$ is a $1 \times h$ vector of Y–loadings, and $h << n$. The X–loadings that comprise $\mathbf{P}$ express the relationships between the raw data $\mathbf{X}$ and the transformed data, or scores $\mathbf{T}$, and the Y–loadings in $\mathbf{q}$ are the regression coefficients from regressing the Y–variables onto the Y–scores $\mathbf{U}$. PLS finds rotation directions to maximize $\mathbf{T}^T\mathbf{U}$. As a result, when the columns of $\mathbf{P}$ are orthogonalized, generating an orthogonal transformation matrix $\mathbf{W}$, the columns of $\mathbf{W}$ are precisely the eigenvectors of $\mathbf{X}^T\mathbf{Y}\mathbf{Y}^T\mathbf{X}$.

3.3 Feature Selection

In order to identify a handful of features that have the greatest influence in the PLS model, the variable importance on projection (VIP) method[22] was employed. VIP provides a score for each feature so that the features can be ranked according to their predictive power in the PLS model. The VIP for the jth feature, π_j, is defined as follows:

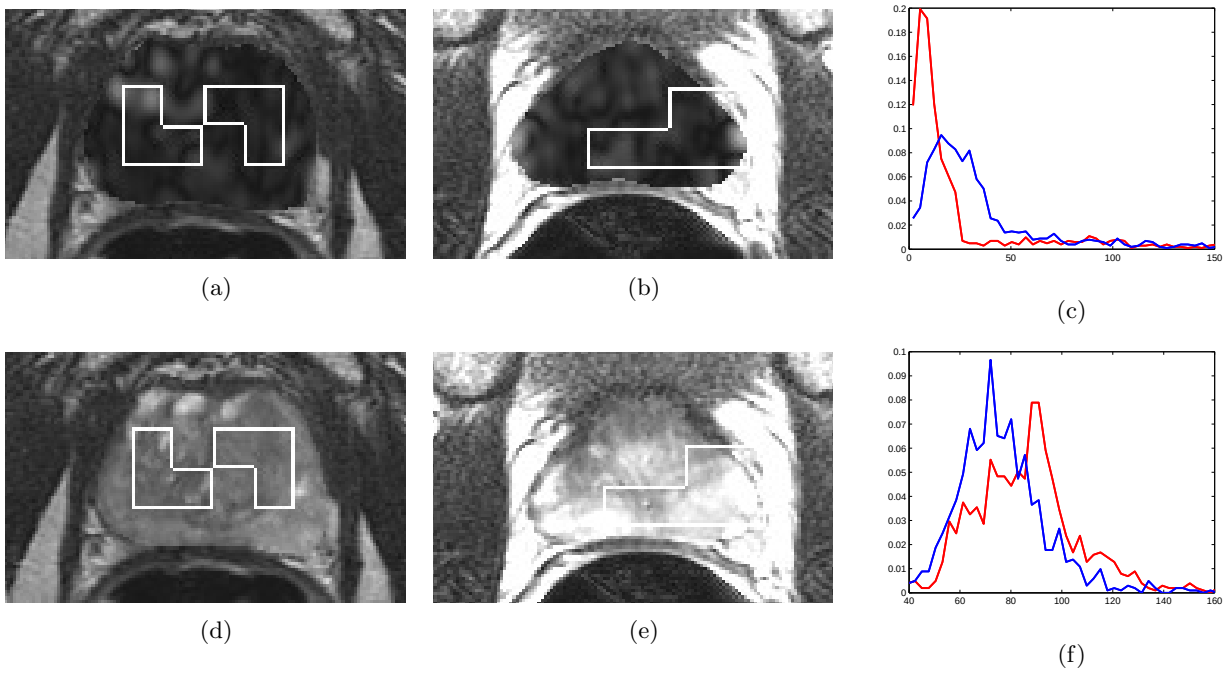


Figure 2: The top–performing Gabor feature selected by VIP (top) and a first–order statistical feature (bottom) is shown for (a) and (d) a recurrence–free patient and (b) and (e) a patient who develops biochemical recurrence with the ground truth cancer extent bounded in white. The histograms in (c) and (f) show the distributions of these feature in cancerous regions of the prostate for the recurrence–free patient (blue) and the patient who experiences biochemical recurrence (red). Note that significant differences in both the skewness and kurtosis of the distributions of the Gabor wavelet feature can be appreciated in (c) but not in (f).

$$\pi_j = \sqrt{m \frac{\sum_{i=1}^{h} b_i^2 \mathbf{t}_i^T \mathbf{t}_i \left(\frac{w_{ji}}{||\mathbf{w}_i||} \right)^2}{\sum_{i=1}^{h} b_i^2 \mathbf{t}_i^T \mathbf{t}_i}}. \tag{2}$$

where m is the number of features in the original, high–dimensional feature space and b_i are regression coefficients that relate the transformed data $\mathbf{T}$ to the outcome vector $\mathbf{y}$. The degree to which a feature contributes to classification in the PLS transformed space is directly proportional to its associated VIP score. Thus, features with VIP scores near 0 have little predictive power, and the features with the highest VIP scores are identified as most useful for discriminating between the two classes.

4. EXPERIMENTAL DESIGN

4.1 Description of Data

Sixteen patients with biopsy–confirmed prostate cancer who underwent external beam radiation therapy or brachytherapy and subsequently participated in at least five years of follow–up were retrospectively included in this study. Biochemical recurrence after radiation therapy was defined, according to the ASTRO definition, as a rise of 2 ng/mL or more above the nadir PSA.[3] Using this criterion, seven patients experienced biochemical recurrence within 10 years of the end of treatment, while the remaining nine patients remained recurrence–free for at least five years. The average time until biochemical recurrence was 6.8 years, and the mean length of follow–up time in the non–failure patients was 7.0 years. Among these sixteen patients two patients who developed biochemical recurrence and three who remained recurrence–free underwent hormonal therapy in addition to radiation therapy.

Semantic MRI Attributes	Kattan Nomogram Features
	Pre-treatment PSA
Tumor size	Tumor stage
Extracapsular spread	Gleason score
Seminal vesicle invasion	Radiation dose
	Hormonal treatment

Table 2: Lists of semantic MRI attributes and clinical features assessed via the Kattan nomogram.

Prior to treatment each patient had been clinically referred for a prostate cancer MR staging exam for improved therapeutic selection; this exam included acquisition of T2w MRI and MR spectroscopy. MRI was performed by using a 1.5 T whole–body MRI unit (Signa; GE Medical Systems, Milwaukee, Wisconsin). The patients were imaged while in the supine position by using a body coil for signal excitation and a pelvic phased–array coil (GE Medical Systems) combined with a balloon–covered expandable endorectal coil (Medrad, Pittsburgh, PA) for signal reception. Data sets were acquired as $16 \times 8 \times 8$ phase–encoded spectral arrays (1024 voxels) by using a nominal spectral resolution of 0.24–0.34 cm^3, 1000/130, and a 17–min acquisition time. MR spectroscopy was not used in our study for biochemical recurrence risk prediction; however, MR spectra were used to determine cancer presence and extent on MRI (see Section 4.2). Consequently, three–dimensional MR spectroscopic imaging data were processed and aligned with the corresponding T2w imaging data using a combination of in–house software and Interactive Display Language (Research Systems, Boulder, Colorado) software tools.[29] The raw spectral data were apodized with a 1–Hz Gaussian function and Fourier transformed in the time domain and in three spatial domains.

4.2 Annotation of Ground Truth Cancer Extent

MR spectral voxels were annotated by an expert radiologist with more than 25 years of experience on a 5–point scale adapted from the standardized 5–point scale developed by Jung et al.,[30] where each spectrum is defined as either (1) definitely benign, (2) probably benign, (3) equivocal, (4) probably malignant, or (5) definitely malignant. While the scale described by Jung et al. was based on metabolic ratios of MR spectra alone, the spectral annotations in this work were performed by incorporating the presence and strength of hypointensities on T2w MRI as well. MRS voxels annotated as "likely benign" or "probably benign" were considered to be "benign" for our analyses, while voxels annotated as "probably malignant" or "likely malignant" were considered cancerous. MR spectroscopy voxels annotated as "equivocal" were excluded from further analysis. Following annotation of ground truth cancer extent on MR spectroscopy, T2w MRI was brought into spatial alignment with MR spectroscopy using information in the image descriptor files. Thus, the texture features described in Section 3.1 were extracted only from T2w MRI voxels labeled as cancerous, and non–canceorus regions were ignored in subsequent analyses.

4.3 Feature Selection via VIP

In order to identify the three computer–extracted features that are most predictive of biochemical recurrence, PLS was employed to reduce the data dimensionality, and a VIP score was subsequently computed for each of the high–dimensional features according to equation (2). In order to ensure generalizability of the selected feature set, leave–one–out cross–validation was performed. At each step of this cross–validation procedure, the data from all but one subject was used to construct the PLS embedding and subsequently to compute feature–wise VIP scores. After the cross–validation procedure was completed, the top three computer–extracted features were identified by voting across all cross–validation iterations, and the most commonly selected features were thus identified.

4.4 Experiment 1: Evaluation of VIP for Feature Selection

In order to evaluate whether the VIP scheme is indeed selecting features with high predictive power, the top 1–10 features with the highest VIP scores were employed to construct logistic regression models that predict a patient's risk of developing biochemical recurrence. Based on the results obtained from these predictive models, receiver operating characteristic (ROC) curves representing the tradeoff between classifier sensitivity and specificity were

VIP: Top 4 Features	VIP Scores	mRMR: Top 4 Features	VIP Scores
Gabor feature 9 skewness	1.30 ± 0.31	Gabor feature 14 SD	1.04 ± 0.09
Gabor feature 8 skewness	1.28 ± 0.29	Gabor feature 51 kurtosis	0.77 ± 0.28
Gabor feature 44 kurtosis	1.18 ± 0.22	Gabor feature 52 skewness	0.65 ± 0.19

Table 3: Top three features selected by the VIP feature selection scheme and by mRMR and their associated VIP scores. SD = standard deviation.

generated. The area under the ROC curve (AUC) was used to comparatively evaluate classifier accuracy in conjunction with different feature subsets. The AUC associated with each of the logistic regression classifiers was evaluated using a leave–one–out cross–validation procedure to ensure generalization of the classifier and to obtain a robust estimate of AUC.

Additionally, in order to compare VIP with state–of–the–art feature selection algorithms in terms of its ability to identify features with high predictive power, we compared VIP with minimum–redundancy–maximum–relevance (mRMR),[31] a popular information theoretic feature selection scheme. AUC values associated with logistic regression classifiers constructed both in conjunction with the top 1–10 features selected by mRMR and with the top 1–10 features selected by VIP were obtained using leave–one–out cross–validation.

4.5 Experiment 2: Comparative Evaluation of Classifier for Biochemical Recurrence Risk Prediction

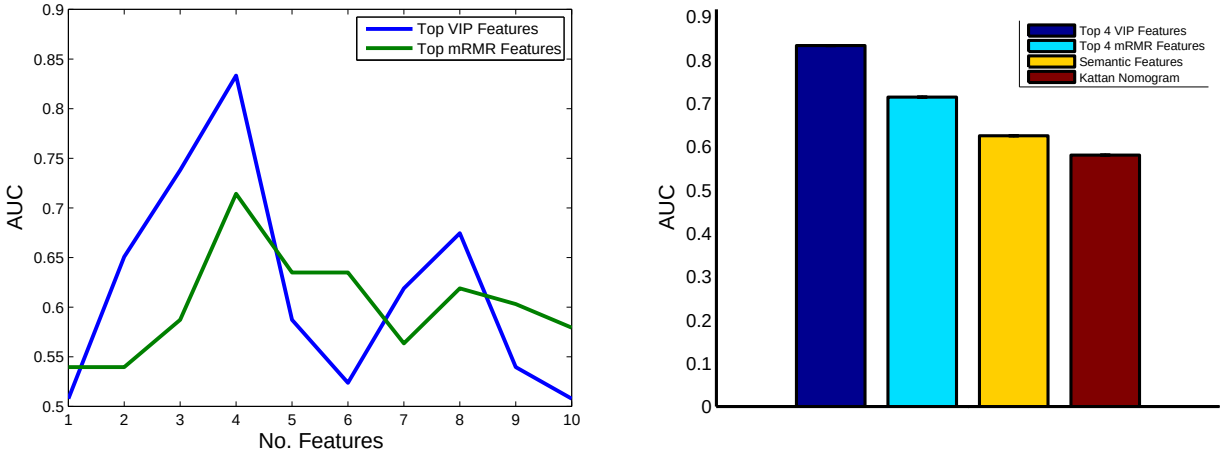
The top three computer–extracted features obtained by the VIP scheme were further compared to (a) semantic MRI attributes (see Table 2) and (b) the Kattan nomogram in terms of their ability to accurately predict biochemical recurrence following radiation therapy. The Kattan nomogram incorporates clinical features (see Table 2) to predict the probability of remaining recurrence–free for five years. To evaluate the predictive power of semantic attributes on MRI, a logistic regression classifier was trained in a leave–one–out cross–validation procedure to predict biochemical recurrence based on the following semantic attributes: number of cancerous MRS voxels, a surrogate measurement of tumor size; presence or absence of extracapsular spread; and presence or absence of seminal vesicle invasion. The AUC associated with this predictive model, as well as the area under the ROC curve drawn based on the survival probabilities obtained using the Kattan nomogram, was thus obtained.

In addition to comparing AUC values associated with these methods, Kaplan–Meier survival analysis was also performed. Kaplan–Meier survival probabilities, calculated as the ratio of the number of subjects who remain event–free to the total number of study subjects, are useful for measuring the fraction of subjects who remain recurrence–free at any given time after treatment. Because Kaplan–Meier survival analysis inherently accounts for censored data, it is able to account for subjects in our study who were lost to follow–up prior to ten years following treatment. Kaplan–Meier survival curves that stratify patients based upon their risk of developing biochemical recurrence were estimated and compared for the top three VIP–selected computer–extracted features and the Kattan nomogram.

5. EXPERIMENTAL RESULTS AND DISCUSSION

5.1 Experiment 1: Evaluation of VIP for Feature Selection

The three computer–extracted features that contribute most to accurate prediction of biochemical recurrence following radiation therapy in the PLS embedding space are the skewness and kurtosis of the distributions of three Gabor wavelet features (see Table 3). When these three features were used in conjunction with a logistic regression classifier to predict biochemical recurrence risk, an AUC of 0.74 was obtained; when the top four features were used, an AUC of 0.83 was achieved (see Figure 3(a)). In Figure 2 grayscale representations of the Gabor filter associated with the highest VIP score illustrate over–expression in a patient who develops biochemical recurrence within four years compared to a patient who remains recurrence–free for ten years. While the distributions of this Gabor feature in these two patients have very different shapes, and hence different values for skewness and kurtosis, the shapes of the distributions of an arbitrarily–selected first–order statistical feature are very similar for both patients (see Figure 2).



(a) (b)

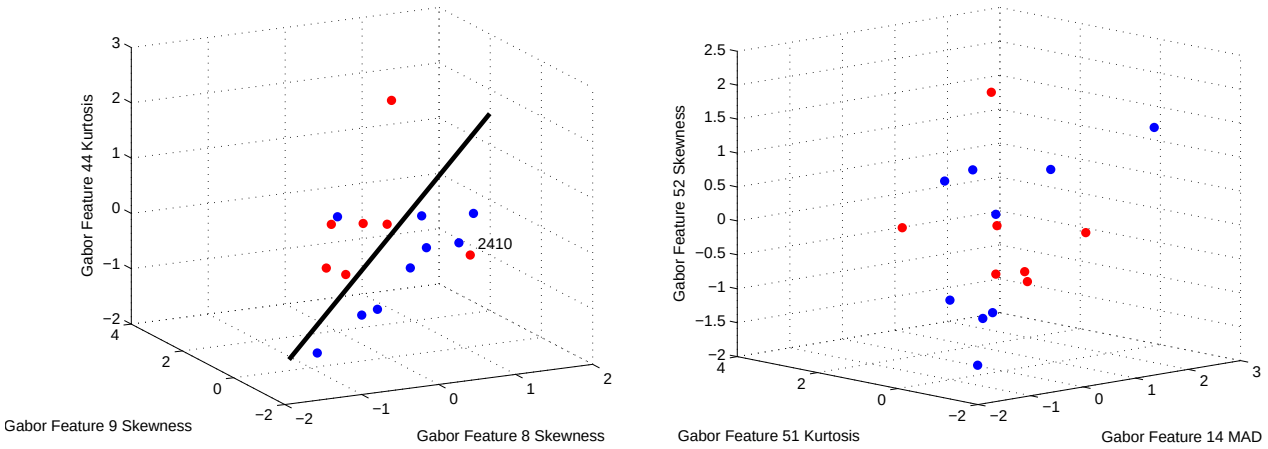
Figure 3: (a) AUC is compared for various numbers of computer–extracted features selected by VIP and mRMR. (b) AUC is compared for the top–performing features obtained via the VIP and mRMR feature selection algorithms, as well as for semantic MRI attributes and the Kattan nomogram.

A scatterplot of the three computer–extracted features with the highest VIP scores shows excellent separation between subjects who will develop biochemical recurrence and those who remain recurrence–free, as a plane drawn through the scatterplot would lead to misclassifications of only one subject in each class (see Figure 4(a)). This suggests that PLS, a linear dimensionality reduction method, is sufficient for this data and that nonlinear methods for dimensionality reduction are not necessary in this context. Interestingly, patient 2410, the only subject who developed biochemical recurrence but whose datapoint is present among the cluster of recurrence–free subjects, did not experience biochemical recurrence until 7.6 years after radiation therapy. This is the longest span of time until biochemical recurrence seen in our patient cohort.

It is notable that the computer–extracted features identified by the VIP scheme were all Gabor wavelet features. In fact, the features selected via the mRMR feature selection routine were also exclusively Gabor wavelet features. These findings suggest the importance of Gabor wavelet features in predicting progression to biochemical recurrence following radiation therapy for prostate cancer. Gabor filters[32] provide multi–scale, multi–orientation features that reflect localized frequency characteristics within the tumors and that may hint to micro–architectural patterns of more aggressive tumors that are more likely to recur. The ability of Gabor wavelet features to capture micro–architectural patterns that characterize cancer has been previously documented, as Gabor wavelet features were identified as important for detection and localization of prostate tumors on T2w MRI[16,33] and ADC maps.[33]

5.1.1 Comparison of VIP and mRMR

Like the computer–extracted features selected by VIP, the top three features selected by mRMR are also exclusively Gabor wavelet features, including the skewness and kurtosis of Gabor filter distributions. A scatterplot of the three top features selected by mRMR, obtained by voting across all cross–validation iterations, is shown in Figure 4(b). Note that these three features provide good separability between the two classes and appear to be fairly separable in this subspace, albeit by nonlinear methods. When the top three features were used in conjunction with a logistic regression classifier to predict biochemical recurrence, an AUC of 0.59 was obtained; when the top four features were used, an AUC of 0.71 was achieved—considerably lower than the AUC values provided by the VIP feature selection scheme (see Figure 3(a)). The improved performance of VIP, when compared with mRMR, in selecting features that are predictive of biochemical recurrence may be attributed to the fact that mRMR performed feature selection in the high–dimensional feature space and was therefore encumbered by the high data dimensionality. In contrast, the VIP scheme overcame the "curse of dimensionality" by first performing dimensionality reduction via PLS and subsequently performing feature selection in the PLS embedding space.



(a) (b)

Figure 4: Scatterplot of the top–performing Gabor features selected by (a) VIP and (b) mRMR.

Thus, VIP may have been more capable of identifying the most predictive features in the face of the high feature dimensionality.

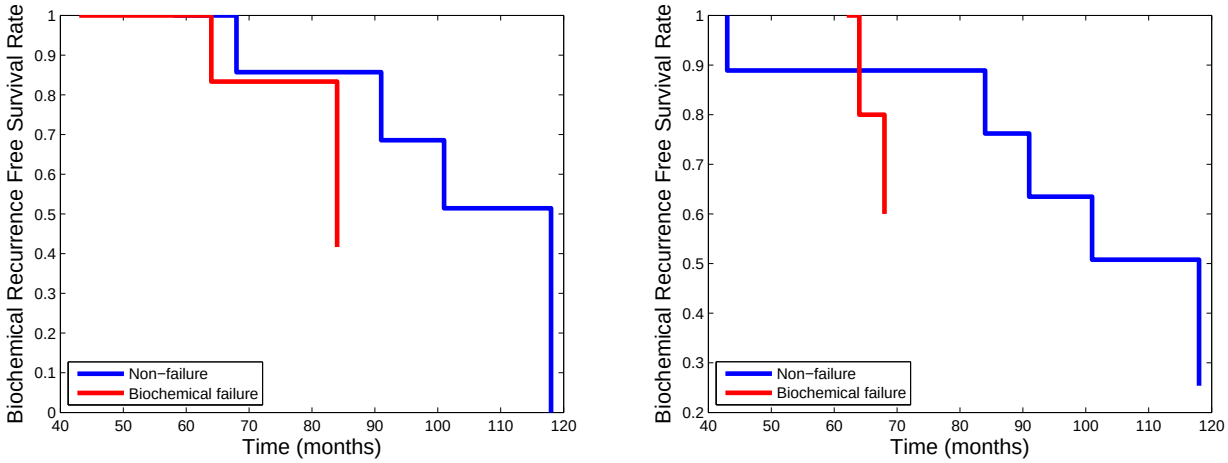
5.2 Experiment 2: Comparative Evaluation of Classifier for Biochemical Recurrence Risk Prediction

In comparison to the top–performing computer–extracted MRI features, semantic attributes on MRI and the Kattan nomogram provided significantly lower AUC values of 0.63 and 0.58, respectively (see Figure 3(b)). Kaplan–Meier survival curves that stratify the patients based upon their risk of developing biochemical recurrence are shown in Figure 5 for the three computer–extracted features selected by VIP and for the Kattan nomogram. It is clear from Figure 5 that VIP–selected features provide better separation between patients who will develop biochemical recurrence and those who will remain recurrence–free than the Kattan nomogram, although the differences between the two survival curves are not statistically significant. These results suggest that a few computer–extracted texture features provide more accurate prediction of biochemical recurrence risk than both the Kattan nomogram and semantic attributes on MRI.

6. CONCLUDING REMARKS

The objective of this work was to identify computer–extracted features from T2w MRI that predict biochemical recurrence risk with high accuracy and to construct a predictive model that uses these features to predict biochemical recurrence following radiation therapy. On a cohort of sixteen patients who underwent radiation therapy, we extracted a combination of texture features from cancerous regions on T2w MRI, implemented PLS to reduce the dimensionality of these MRI–derived features, and leveraged the VIP scoring system to identify the features that are most predictive of biochemical recurrence risk. In comparison to both the Kattan nomogram and semantic MRI features, the three computer–extracted features identified by the VIP scheme, in conjunction with a logistic regression model predicting biochemical recurrence risk, led to more accurate differentiation between patients who later developed biochemical recurrence and those who remained recurrence–free.

Nevertheless, there are several limitations to our study. Firstly, our entire cohort of patients was small, consisting of only sixteen subjects: seven patients who developed biochemical recurrence and nine who remained recurrence–free. The small sample size may have significant implications in terms of generalization of the computer–extracted features identified by the VIP scheme to predict risk of biochemical recurrence in a larger or different population. Secondly, many of the patients who remained recurrence–free had limited follow–up. Patients were excluded from our study if they were not followed–up regularly for five years after completion of radiation therapy, but we cannot be certain that patients who were followed up for between 5–10 years truly



(a) (b)

Figure 5: Kaplan–Meier survival curves obtained via (a) the three computer–extracted features selected by VIP and (b) the Kattan nomogram. Differences in survival curves between patients who experienced biochemical failure and those who did not were not statistically significant.

remained recurrence–free during this time. Thirdly, predicting biochemical recurrence necessitated using older data associated with long term outcome information; consequently, the imaging data used in our study was approximately fifteen years old, had poor spatial resolution, and included only T2w MRI. The fact that our methods outperform the Kattan nomogram in spite of the poor data quality suggest that more recently acquired MRI, which generally includes multiparametric images with improved spatial and temporal resolution, will prove to be even more useful for predicting biochemical recurrence following treatment.

A final limitation of our study is the limited ability of clinical biochemical recurrence, as defined by PSA criteria, to predict long–term patient outcome or eventual metastasis.[34] For example, two patients in our cohort who remained recurrence–free, as defined by PSA criteria and the ASTRO definition, for five years after external beam radiation therapy later underwent brachytherapy although they remained recurrence–free, and a third patient who underwent a prostate biopsy nearly seven years after completing radiation therapy was found to have a Gleason grade 6 tumor but no evidence of clinical biochemical recurrence. These caveats, which are artifacts of the ASTRO definition's limitations in predicting disease progression, may have negatively impacted the generalizability of our predictive model.

Although our study was small and the data limited, our finding that computer–extracted texture features are more useful for predicting the likelihood of developing biochemical recurrence following radiation therapy than the Kattan nomogram and semantic features on MRI is significant. Our results imply that computer–extracted features are successfully able to capture micro–architectures in the tumors that provide morphometric clues for predicting biochemical recurrence. Furthermore, our success in mining the low–resolution T2w MRI data for clues to predict progression to biochemical recurrence suggests that computer–extracted features from contemporary MRI data, which involves high–resolution, multi–parametric MRI protocols, may lead to models for predicting biochemical recurrence with even higher accuracy. Moreover, the fact that computer–extracted features appear to be more predictive of biochemical recurrence than the Kattan nomogram and semantic MRI attributes suggests that computer–extracted features provide information above and beyond what is provided by tumor stage, grade, and size and the presence of extracapsular spread or seminal vesicle invasion. Thus, incorporating computer–extracted features from T2w MRI in a nomogram to predict biochemical recurrence may be highly beneficial and lead to improved nomograms that predict a patient's risk of biochemical recurrence with greater accuracy.

REFERENCES

[1] American Cancer Society. Cancer Facts and Figures 2013. Atlanta: American Cancer Society, 2013.

[2] Darish, O.M. and Raj, G.V. Management of biochemical recurrence after primary localized therapy for prostate cancer. *Frontiers in Oncology* **2**, 1-6 (2012).

[3] Roach, M., Hanks, G., Thames H., et al. Defining biochemical failure following radiotherapy with or without hormonal therapy in men with clinically localized prostate cancer: recommendations of the RTOG–ASTRO Phoenix Consensus Conference. *Int J Radiat Oncol Biol Phys* **65**, 965-74 (2006).

[4] Uchio, E.M., Aslan, M., Wells, C.K., et al. Impact of biochemical recurrence in prostate cancer among US veterans. *Arch Intern Med* **170**:1390-5 (2010).

[5] Kattan, M.W., Zelefsky, M.J., Kupelian P.A., et al. Pretreatment nomogram for predicting the outcome of three–dimensional conformal radiotherapy in prostate cancer. *J Clin Oncol* **18**, 33529 (2000).

[6] Wang, L., Hricak, H., Kattan, M.W., et al. Prediction of seminal vesicle invasion in prostate cancer: Incremental value of adding endorectal MR imaging to the Kattan nomogram. *Radiology* **242**, 182-8 (2007).

[7] Gibbs, P., Liney, G.P., Pickles, M.D., et al. Correlation of ADC and T2 measurements with cell density in prostate cancer at 3.0 tesla. *Invest Radiol* **44**, 572-6 (2009).

[8] Oto, A., Yang, C., Kayhan, A., et al. Diffusion–weighted and dynamic contrast–enhanced MRI of prostate cancer: Correlation of quantitative MR parameters with Gleason score and tumor angiogenesis. *Am J Roentgenol* **197**, 1382-90 (2011).

[9] Peng, Y., Jiang, Y., Yang, C., et al. Quantitative analysis of multiparametric prostate MR images: Differentiation between prostate cancer and normal tissue and correlation with Gleason score—a computer–aided diagnosis development study. *Radiology* **267**, 787-96 (2013).

[10] Westphalen, A.C., Koff, W.J., Coakley, F.V., et al. Prostate cancer: prediction of biochemical failure after external–beam radiation therapy—Kattan nomogram and endorectal MR imaging estimation of tumor volume. *Radiology* **261**(2), 477-86 (2011).

[11] Fuchsjager, M.H., Pucar, D., Zelefsky, M.J., et al. Predicting post–external beam radiation therapy PSA relapse of prostate cancer using pretreatment MRI. *Int J Radiat Oncol Biol Phys* **78**(3), 743-50 (2010).

[12] Riaz, N., Afaq, A., Akin, O., et al. Pretreatment endorectal coil magnetic resonance imaging findings predict biochemical tumor control in prostate cancer patients treated with combination brachytherapy and external–beam radiotherapy. *Int J Radiat Oncol Biol Phys* **84**, 707-11 (2012).

[13] Chan, I., Wells, W., Mulkern, R.V., et al. Detection of prostate cancer by integration of line–scan diffusion, T2–mapping and T2weighted magnetic resonance imaging; a multichannel statistical classifier. *Med Phys* **30**, 2390-8 (2003).

[14] Niaf, E., Rouviere, O., Mege–Lechevallier, F., et al. Computer–aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. *Phys Med Biol* **57**, 3833-51 (2012).

[15] Tiwari, P., Viswanath, S., Kurhanewicz, J., et al. Multimodal wavelet embedding representation for data combination (MaWERiC): Integrating magnetic resonance imaging and spectroscopy for prostate cancer detection. *NMR Biomed* **25**, 607-19 (2011).

[16] Viswanath, S.E., Bloch, N.B., Chappelow, J.C., et al. Central gland and peripheral zone prostate tumors have significantly different quantitative imaging signatures on 3 tesla endorectal, in vivo T2weighted MR imagery. *J Magn Reson Imaging* **36**, 213-24 (2012).

[17] Vos, P.C., Hambrock, T., Barentsz, J.O., Huisman, H.J. Computer–assisted analysis of peripheral zone prostate lesions using T2weighted and dynamic contrast enhanced T1–weighted MRI. *Phys Med Biol* **55**, 1719 (2010).

[18] Epstein, J.I., Allsbrook, W.C., Amin, M.B., Egevad, L.L.; ISUP Grading Committee. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason grading of prostatic carcinoma. *Am J Surg Pathol* **29**,1228-42 (2005).

[19] Tiwari, P., Kurhanewicz, J., Madabhushi, A. Multi–kernel graph embedding for detection, Gleason grading of prostate cancer via MRI/MRS. *Med Image Anal* **17**, 219-35 (2013).

[20] Yan, H., Yuan, X., Yan, S., Yang, J. Correntropy based feature selection using binary projection. *Pattern Recogn* **44**, 2834-42 (2011).

[21] Hughes, G.F. On the mean accuracy of statistical pattern recognizers. *IEEE Trans Info Theory* **IT–14**, 55-63 (1968).

[22] Chong, I.G. and Jun, C.H. Performance of some variable selection methods when multicollinearity is present. *Chemometr Intell Lab* **78**, 103-12 (2005).

[23] Hambrock, T., Vos, P.C., Hulsbergen–van de Kaa, C.A., et al. Prostate cancer: Computer–aided diagnosis with multiparametric 3–T MR imaging—effect on observer performance. *Radiology* **266**, 521-30 (2013).

[24] Mazaheri, Y., Hricak, H., Fine, S.W., et al. Prostate tumor volume measurement with combined T2–weighted imaging and diffusion–weighted MR: correlation with pathologic tumor volume. *Radiology* **252**, 449-57 (2009).

[25] Shimofusa, R., Fujimoto, H., Akamata, H., et al. Diffusion–weighted imaging of prostate cancer. *J Comput Assist Tomogr* **35**, 223-8 (2011).

[26] Ampeliotis, D., Antonakoudi, A., Berberidis, K., et al. A computer–aided system for the detection of prostate cancer based on magnetic resonance image analysis. *Communications, Control and Signal Processing*, 13727 (2008).

[27] Park, S.Y., Kim, C.K., Park, B.K., et al. Prediction of biochemical recurrence following radical prostatectomy in men with prostate cancer by diffusion–weighted magnetic resonance imaging: Initial results. *Eur Radiol* **21**, 1111-8 (2011).

[28] Esbensen, K. Multivariate data analysis in practice: An introduction to multivariate data analysis and experimental design. CAMO: Norway; 2004.

[29] Tran, T.K., Vigneron, D.B., Sailasuta, N., et al. Very selective suppression pulses for clinical MRSI studies of brain and prostate cancer. *Magn Reson Med* **43**, 23-33 (2000).

[30] Jung, J.A., Coakley, F.V., Vigneron, D.B., et al. Prostate depiction at endorectal MR spectroscopic imaging: Investigation of a standardized evaluation system. *Radiology* **233**, 7018 (2004).

[31] Peng, H., Long, F., Ding, C. Feature selection based on mutual information criteria of max–dependency, max–relevance, and min–redundancy. *IEEE Trans Pattern Anal Mach Intell* **27**, 1226-38 (2005).

[32] Turner, M.R. Texture discrimination by Gabor functions. *Biol Cybern* **55**, 71-82 (1986).

[33] Ginsburg, S., Tiwari, P., Kurhanewicz, J., Madabhushi, A. Variable ranking with PCA: Finding multiparametric MR imaging markers for prostate cancer diagnosis and grading. In proc. Prostate Cancer Imaging Workshop, MICCAI (2011).

[34] Punnen, S., Cooperberg, M.R., D'Amico, A.V., et al. Management of biochemical recurrence after primary treatment of prostate cancer: A systematic review of the literature. *Eur Urol* **64**, 905-15 (2013).