# Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel statistical classifier

Ian Chan and William Wells III
*Surgical Planning Laboratory, Department of Radiology, Division of MRI, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115*
*and MIT Artificial Intelligence Laboratory, Cambridge, Massachusetts 02139*

Robert V. Mulkern
*Children's Hospital Department of Radiology, Harvard Medical School, Boston, Massachusetts 02115*

Steven Haker and Jianqing Zhang
*Surgical Planning Laboratory, Department of Radiology, Division of MRI, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115*

Kelly H. Zou
*Surgical Planning Laboratory, Department of Radiology, Division of MRI, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115*
*and Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts 02115*

Stephan E. Maier
*Brigham and Women's Hospital Department of Radiology, Harvard Medical School, Boston, Massachusetts 02115*

Clare M. C. Tempany[a)]
*Surgical Planning Laboratory, Department of Radiology, Division of MRI, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115*

(Received 16 October 2002; accepted for publication 26 May 2003; published 22 August 2003)

A multichannel statistical classifier for detecting prostate cancer was developed and validated by combining information from three different magnetic resonance (MR) methodologies: T2-weighted, T2-mapping, and line scan diffusion imaging (LSDI). From these MR sequences, four different sets of image intensities were obtained: T2-weighted (T2W) from T2-weighted imaging, Apparent Diffusion Coefficient (ADC) from LSDI, and proton density (PD) and T2 (T2 Map) from T2-mapping imaging. Manually segmented tumor labels from a radiologist, which were validated by biopsy results, served as tumor ''ground truth.'' Textural features were extracted from the images using co-occurrence matrix (CM) and discrete cosine transform (DCT). Anatomical location of voxels was described by a cylindrical coordinate system. A statistical jack-knife approach was used to evaluate our classifiers. Single-channel maximum likelihood (ML) classifiers were based on 1 of the 4 basic image intensities. Our multichannel classifiers: support vector machine (SVM) and Fisher linear discriminant (FLD), utilized five different sets of derived features. Each classifier generated a summary statistical map that indicated tumor likelihood in the peripheral zone (PZ) of the prostate gland. To assess classifier accuracy, the average areas under the receiver operator characteristic (ROC) curves over all subjects were compared. Our best FLD classifier achieved an average ROC area of $0.839(\pm 0.064)$, and our best SVM classifier achieved an average ROC area of 0.761 $(\pm 0.043)$. The T2W ML classifier, our best single-channel classifier, only achieved an average ROC area of $0.599(\pm 0.146)$. Compared to the best single-channel ML classifier, our best multichannel FLD and SVM classifiers have statistically superior ROC performance ($P = 0.0003$ and 0.0017, respectively) from pairwise two-sided *t*-test. By integrating the information from multiple images and capturing the textural and anatomical features in tumor areas, summary statistical maps can potentially aid in image-guided prostate biopsy and assist in guiding and controlling delivery of localized therapy under image guidance. 
[DOI: 10.1118/1.1593633]

Key words: prostate cancer detection, magnetic resonance imaging, T2 map, line scan diffusion imaging, image guided diagnosis

## INTRODUCTION

Prostate cancer is the most commonly diagnosed noncutaneous malignancy and the second-leading cause of death from cancer in American men.[1] Over 40 000 American men are diagnosed with prostate cancer each year and cancer is found at autopsy in 30% of men at the age of 50.[2] Evidence sug-

gested that 1.5 T axial T2-weighted MR prostate images with an endorectal coil enables improved visualization and localization of prostate substructure [central gland (CG) and peripheral zone (PZ)], providing valuable pathologic and anatomical information.[3–6] It is known that the majority of prostate cancers develop in the PZ.[7] While such diagnostic T2-weighted MRI images are a sensitive noninvasive imaging technique for detecting focal abnormalities in the prostate, they lack specificity distinguishing between tumor and benign prostatic hyperplasia (BPH) and other abnormalities. It is reported that T2-weighted MR only has a specificity of 43% for nonpalpable tumors but a sensitivity of 85% for nonpalpable, posteriorly located tumors.[8] Thus, there is a need to improve the image-based specificity to diagnose cancer and to integrate information from other MR methodologies or imaging modalities. As it appears that no one imaging sequence or technique will be adequate, the development of techniques which integrate multiple data sets for clinical use becomes important.

Recent advances in MR techniques have allowed us to obtain information based on water diffusion and T2 properties to improve MR specificity for prostate cancer. Quantitative T2-mapping is motivated in part by some intriguing results from Liney *et al.*,[9–11] who suggested a positive correlation between the concentration of citrate, as determined by $^1$H spectroscopy and the water T2 value obtained from T2-mapping methods. Since citrate is the ''good'' metabolite of the prostate gland, presenting the strongest metabolite signal in normal tissue and BPH,[12] a decrease in the citrate signal provides an indirect indication of potential cancer. If T2 maps indeed correlate with citrate concentrations, i.e., carry useful diagnostic data, the higher spatial resolution of T2-mapping compared to MR spectroscopy would offer potentially greater utility in tumor localization in the prostate.

Diffusion-weighted and quantitative diffusion MR imaging (ADC mapping) is used to obtain tissue contrast reflecting water molecular diffusion. Diffusion MR has become essential for assessing acute stroke in the brain.[13,14] More recently, evidence has been presented which suggests that diffusion imaging may also play a role in the early detection of tumor response to therapy.[15,16] However, diffusion studies have largely been limited to the brain for a variety of reasons including motion sensitivity and the chemical shift and susceptibility artifacts which plague single-shot echo planar imaging (EPI) techniques used to overcome motion sensitivity. A diffusion technique with low motion sensitivity and reduced susceptibility and chemical shift artifacts is the LSDI technique recently developed and shown useful for brain and spine imaging.[17–19] Its primary drawback is slower acquisition times compared to single-shot EPI methods, though full coverage of the rather small prostate gland at reasonable spatial resolutions (24 mm$^3$ voxels) is readily attained in several minutes. Thus we have used LSDI methodology to obtain quantitative diffusion data from our patients, motivated in part by preliminary results of others which suggest reduction of ADC values in tumor vs normal PZ.[20]

Early works on multichannel MRI classification frequently used a maximum-likelihood approach.[21,22] There are also many alternative statistical approaches to multichannel classification for image analysis and pattern recognition.[23] FLD[24] is a traditional approach where the $d$-dimensional feature space is projected onto a line which results in the largest variance of the data. That is,

$$y = \mathbf{w}^T \mathbf{x},$$

where $\mathbf{x}$ is the data vector and $y$ is the projected value on a line. $\mathbf{w}$ is proportional to

$$\mathbf{w} \alpha S_w^{-1} (\mathbf{m}_t - \mathbf{m}_n),$$

where $S_w = \Sigma_{\text{tumor}}(x - \mathbf{m}_t)(x - \mathbf{m}_t)^T + \Sigma_{\text{normal}}(x - \mathbf{m}_n)(x - \mathbf{m}_n)^T$ is the pooled sample variance of the two classes and $\mathbf{m}_t$ and $\mathbf{m}_n$ are the mean of the tumor and normal samples, respectively.

Introduced in 1995 by Vapnik,[25] the SVM,[26] is a classification technique that has gained popularity in recent years for medical applications.[27] The objective function that SVM maximizes is $L_D = \Sigma_i \alpha - 0.5 \Sigma_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j)$, subject to $0 \leqslant \alpha \leqslant C$ and $\Sigma_i \alpha_i y_i = 0$. $\mathbf{x}$ is the multidimensional data vector and $y$ is the corresponding class label ($+1$ or $-1$ for our 2-class problem). $i,j = 1,...,L$ where $L =$ number of training samples. $K(x,y)$ is the kernel function that maps the input to a higher dimensionality and gives the decision boundary its nonlinearity. $C$ is a user-chosen optimization parameter and describes the cost of classification error during SVM training. A large $C$ tells SVM to produce a decision boundary that avoids making any classification errors with the training data. SVM training yields $\mathbf{w}$, which is given by $\Sigma_{i=1}^{N} \alpha_i K(y_i, \mathbf{x}_i)$, where $N$ is the number of support vectors. A data point $\mathbf{x}_i$ is a support vector when the corresponding $\alpha_i > 0$. Support vectors are sample data points that lie on the decision boundary. One can assign the class label based on $y = \mathbf{x}^T \mathbf{w} + b$ where $b$ is the bias term obtained from SVM training. The more positive the value of $y$ is, the farther $\mathbf{x}$ is away from the decision boundary and the more likely it is to be a member of class $+1$. The case is similar for the $-1$ labels when $y < 0$.

To enhance the feature space, many medical image classification applications have used machine vision techniques including CM and various spatial-frequency filters that can enhance image features. Haralick and Shanmugan[28] pioneered the use of CM to capture image texture. A CM is a probability density distribution of two pixel intensities conditioned on distance and angle between the two pixels. Recent applications of CM in medical image analysis include MR T2-weighted breast cancer,[29] breast mammograms,[30] MR imaging for tracking Alzheimer's disease,[31] and coloscopic images of cervix lesions.[32] The textural features generated by CM were reported to enhance classification power in these studies.

We conducted a set of experiments to assess our hypothesis that the SVM classifier will perform better than a FLD or a ML classifier. Also, we hypothesize that classifiers that integrate anatomical and textural information can achieve a higher tumor detection performance than a classifier based on image intensities alone. Our objective is to construct a

summary statistical map based on multiparametric MR images that can improve tumor detection and localization. Improved tumor localization may be valuable for a variety of treatment strategies for prostate cancer including brachytherapy and newer more precise therapies such as focused ultrasound.

## MATERIALS AND METHODS

### Patient selection and imaging protocols

Fifteen patients were enrolled in the study. The eligibility criteria were men with

(1) abnormal prostate specific antigen (PSA) levels (>4 ng/ml) and
(2) either have had at least two prior negative/normal prostate biopsies performed by *trans*-rectal ultrasound (TRUS) or cannot undergo TRUS biopsies because of prior rectal surgery.

For our analysis, these patients were further stratified into two groups based on whether they had undergone a brachytherapy procedure or not. This is because post-brachytherapy patients tended to have altered T2W intensities due to the induction of radiation fibrosis. We restricted our multichannel classifiers to the nonbrachytherapy group because the majority of patients belonged to that group. The nonbrachytherapy group had no confirmed cancer, whereas the brachytherapy group had biopsy-confirmed cancer. Sextant biopsy (samples collected from apex left, apex right, mid-gland left, mid-gland right, base left, and base right of the gland) was performed on each patient.

The 1.5 T MR (Signa LX, GE Medical Systems, Milwaukee WI) imaging was performed in a clinical system using an endorectal coil with an integrated pelvic-phased multicoil array. The endorectal coil is a receive-only coil mounted inside a latex balloon, and assumes a diameter of 4–6 cm once inflated in the patient's rectum. The patient was placed supine in the closed-bore magnet for the examination. The axial T2W images were fast spin echo (FSE) images using 12 echoes and a repetition time (TR)/effective echo time (ETE) = 6050 ms/105 ms, field of view of 12 cm, section thickness of 3 mm, section gap of 0 mm, matrix of 256 $\times$256, three signal averages). Typical acquisition times are 5–6 min.

Diffusion-weighted images were acquired with LSDI in an oblique coronal orientation to provide maximal gland coverage with a minimal number of slices. LSDI involved the use of a pair of slice selective RF pulses to elicit an appropriately diffusion weighted spin–echo signal from a single column (loosely referred to as ''line'' in the LSDI acronym) of tissue. The LSDI sequence employed 64 columns per slice with each column being 16 $mm^2$ in cross section and with an along-column resolution of 1.5 mm resulting in approximately 24 $mm^3$ voxel sizes. Two *b*-factors, 5 and 750 $s/mm^2$ were employed along with three separate diffusion sensitization directions (1,−1,−1/2), (1/2,1,−1), and (1,1/2,1). Though individual columns were sampled every 0.12 s, the effective TR was greater than 4 s so that minimal T1-weighting occurred. A TE of 70 ms accommodated the high *b*-factor of 750 $s/mm^2$ using our available gradient strengths (maxima of 2.3 Gauss/cm). Using two point monoexponential analyses along each diffusion sensitization direction, high quality trace ADC maps are obtained, as discussed in detail in the original LSDI article by Gudbjartsson *et al.* A total of 5 to 12 slices at 4 mm slice thickness and no gap were used to cover the entire gland with $4\times4\times1.5$ $mm^3$ voxel dimensions in total scan times between 5 and 10 minutes (50 s/slice).

An FSE sequence utilizing eight echoes with an echo spacing of 13.5 ms was employed to make maps of the spin–spin relaxation time (T2) throughout the entire prostate gland at a $3\times0.7\times0.8$ $mm^3$ spatial resolution with slices acquired in oblique coronal planes, as with LSDI. The signal intensity ($S$) has the approximate form $S=\rho(1-\exp(-TR/T1))\times\exp(-T2/TE)$, where $\rho$ is the PD intensity, T1 the spin-lattice relaxation time. A 256×192 (frequency by phase) in-plane matrix was used with a 2.5 s TR to gather 5 to 12 3-mm-thick contiguous images of the gland in 60 s. The sequence is repeated four times with four different values of the effective echo time ranging from 27 ms to 108 ms at 27 ms intervals. Thus in approximately 4 minutes a complete data set is collected allowing for mapping of the T2 value throughout the gland by performing monoexponential fits of signal intensity vs echo time for each voxel. We interpret the *y* intercept of the fits as the effective proton density (PD) intensity and the negative inverse of the slope as the effective T2 value used for the T2 Map intensities. The accuracy and precision of this T2 Map methodology was tested with five small (10 mm OD) vials of water doped with the gadolinium (Gd) based MR contrast agent ProHance (Bracco Diagnostics Inc., Princeton, NJ) to levels of 1, 1.5, 2, 2.5, and 3 mM Gd in the form of 10-(2-hydroxy-propyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triacetic acid. The five vials were bundled together in a 3 cm ring roughly the size of the prostate and were scanned with the same coil configuration used for prostate imaging. The $\mu\pm\sigma$ T2 values (in ms) calculated from a 3×3 region of interest (ROI) within each vial of the T2 Map were, from lowest to highest Gd concentrations, $255\pm18$, $155\pm9$, $111\pm4$, $84\pm7$, and $68\pm3$. The mean coefficient of variation for T2 over this range of T2 values encountered in the prostate (*vide infra*) is thus 5.8% which we assessed as our T2 Map precision. The relaxivity curve (1/T2 vs mM Gd) generated from this data was highly linear with a correlation coefficient above 0.999 and a slope of 5.40 $s^{-1}$/mM. This value for the relaxivity is typical of T2 relaxivities reported for common Gd chelates contrast agents. Reichenbach *et al.*,[33] for example, reported a relaxivity of 5.2 $s^{-1}$/mM Gd for a similar contrast agent at 1.5 T. Thus, our T2 Map methodology has an inherent precision on the order of 6% and is reasonably accurate over the range of T2 values encountered in the prostate.

### Tumor "ground-truth" labels

A radiologist manually outlined suspected tumor (TU), PZ, and total gland (TG) on axial T2W images. We com-

bined the tumor label along with the biopsy pathology reports to form biopsy validated tumor labels for classifier training. In a typical biopsy, the PZ is divided into six regions: left/right×base/mid/apex. Only regions marked positive for cancer in both the biopsy report and the radiologist's label are included in the validated ''ground truth'' tumor labels.

### Textural features

We eliminated the angle dependence in the CM based on the observation that tumor textures in the prostate possess radial symmetry. Our CM covered a $9\times9$ pixel window and had 14 distinct distances. We scaled image intensities to fit a range between 0–255 by mapping the 256 levels linearly onto the image intensity range of $\max\{0,(\mu-3\cdot\sigma)\}$ to $(\mu+3\cdot\sigma)$. For each center pixel, we considered its $n$ neighbors that were equidistant from the center pixel. To model the ''texture'' differences of tumor vs normal tissues, we constructed log likelihood ratios $[\log(P(\text{tumor})/P(\text{normal})]$ for each of these $n$ pairs and took the median of the $n$ log likelihoods as the feature statistic. $P(\text{tumor})$ and $P(\text{normal})$ are the conditional probabilities of tumor and normal pixel, respectively, obtained from the CM entries that correspond to the pair of pixel intensities and distance between pixels. Because radiologists often considered the slice above or below the current image slice for confirmation of a lesion/tumor, we extended the CM method to 3-dimension by constructing CM one slice above and below the current slice. Thus, 14 features for the CM from the same slice and 15 features from the CM one slice above or below the center pixel resulted in 29 CM features total for each of the four basic MR images.

To capture the frequency characteristics of tumor, we computed 2D–DCT[34] of a 7 pixel×7 pixel window centered at each pixel in the images. DCT is used in JPEG image compression[35] which retains the lower frequencies of an image. In general, these lower frequencies are able to capture the general visual features of images. The 49 DCT coefficients formed our feature statistics for each of the four basic MR images.

### Anatomical features

We used the cylindrical coordinate system $(r,\theta,z)$ to describe each anatomical location and set the origin at the centroid of the gland. The coordinates $r$ and $z$ are rescaled to fit the range of $-1$ to 1 for each gland. $\theta$ is in the range of 0 to $\pi$ with anterior set to 0 and we assume left/right symmetry of the gland. The $z$ coordinate could help distinguish the apex, mid-gland, and base, which had observably different image and anatomical features. Empirical observations pointed to common occurrences of prostate tumors in the axial 5:00 and 7:00 o'clock positions of the PZ and the $\theta$ coordinate could potentially distinguish these anatomical areas.

### Feature sets

Five feature sets were chosen to test our classifiers. *basic4* includes only the four signal intensities: T2W, ADC, PD, and T2 Map. *basic4+anatomy* includes the four signal intensities and the three cylindrical coordinates from anatomical features. Besides the seven features in *basic4+anatomy*, *all CM* also includes 29 CM entries for each MR sequence, generating an additional $4\times29$ features. *all DCT* includes all the *basic4+anatomy* features plus the DCT features from all four types of MR images. All *CM+DCT* is the union of all the features in *all CM* and *all DCT*.

### Statistical classifiers

For FLD training, we randomly sampled 10% of the PZ data and retained all the tumor data because we had about 20 times more PZ data than tumor data. All features were standardized to $\mu=0$ and $\sigma=1$ prior to training. After obtaining the FLD vector $\mathbf{w}$, we constructed a ML classifier using the training dataset.

For SVM training, we randomly sampled 10% of the PZ data and retained all the tumor data to confine the training dataset to a manageable size for SVM training convergence. All features are standardized to $\mu=0$ and $\sigma=1$ prior to training. We chose the radial basis function kernel $K(x,y)=\exp\{-|x-y|^2/\sigma\}$ with parameters $\sigma=2$ and $C=100$. In this study, we used the MATLAB support vector toolbox[36] developed by Crawley, which used Platt's sequential minimization algorithm for optimization.[37]

### Software

All feature generation and classifier training were computed using the MATLAB[38] software package. The software for manual segmentation by the radiologist, volume calculation, and image interpolation was developed in-house for this project with the Java[39] language.

### Classifier accuracy

To compare the accuracy of classifiers, we performed standard ROC[40] analysis on each of the 11 nonbrachytherapy patients. We chose the area under the ROC curve as our benchmark for classifier performance. We adopted a jackknife strategy where we trained the classifier with $n-1$ cases and applied the trained classifier on the remaining case. The $\mu\pm\sigma$ of the $n$ leave-one-out ROC areas were reported. To determine the significance in the difference in mean ROC areas of two classifiers, we utilized pairwise 2-sided students' $t$-test at $\alpha=0.05$.

## RESULTS

### Pathology summary

We divided the patients into two groups.

(1) Post-brachytherapy group: had confirmed cancer and prior brachytherapy treatment.
(2) Nonbrachytherapy group: had no prior treatment.

Of the 11 nonbrachytherapy patients in this study, nine had confirmed adenocarcinoma and one had prostatic intraepithelial neoplasm (PIN), a precursor to cancer, from the

TABLE I. Signal intensity summary for 11 nonbrachytherapy patients. PZ signals are standardized to $\mu=0$ and $\sigma=1$. Tumor signals are normalized with the mean and standard deviation of the PZ for comparison. $\mu$ and $\sigma$ of nonstandardized ADC and T2 Map values are also presented.

| Image type | PZ ($\mu\pm\sigma$) | Tumor ($\mu\pm\sigma$) |
|---|---|---|
| T2W | $0\pm1$ | $-0.58\pm0.48$ |
| ADC | $0\pm1$ | $-0.60\pm0.96$ |
| PD | $0\pm1$ | $-0.32\pm0.83$ |
| T2 Map | $0\pm1$ | $-0.55\pm0.79$ |
| ADC (in $\mu^2$/ms) | $1.611\pm0.366$ | $1.432\pm0.349$ |
| T2 Map (in ms) | $128.3\pm42.9$ | $102.7\pm27.5$ |

biopsy reports. The average recurrent tumor volume was $1.03\pm0.56$ cm$^3$ and the average total gland volume was $43.03\pm19.26$ cm$^3$. The mean percentage of TU volume/PZ volume was $6.80\pm3.99$ %. Patients in this group had small to medium size tumors, with minimal seminal vesicle and extra-capsule invasion. The average Gleason score for those with confirmed cancer was 6.2 and had a range of 6–7, indicating that most of the nonbrachytherapy patients had medium grade adenocarcinoma. Of the four post-brachytherapy patients in this study, the average recurrent tumor volume was $1.72\pm1.50$ cm$^3$ and the average total gland volume was $33.36\pm7.26$ cm$^3$. The mean percentage of TU volume/PZ volume was $17.62\pm18.16$ %. The average Gleason score for this group was 6.5 and had a range of 6–8, indicating that the patients in this study had more aggressive tumors with medium to high grade adenocarcinoma and larger volume tumors. This is not surprising as their tumors had failed to respond to the initial therapy and had recurred.

### Signal intensity statistics

From the 11 nonbrachytherapy patients, 9325 PZ data points were collected and out of this, 507 were TU data points. From the four post-brachytherapy patients, 2490 PZ data points were collected and out of this, 308 were TU data points. The $\mu\pm\sigma$ of the signal intensities for each MR image parameter in this study is shown in Tables I and II. For each patient, PZ intensities were standardized to $\mu=0$ and $\sigma=1$ and TU intensities were standardized using the mean and standard deviation of the PZ for comparison. (See Fig. 1.)

TABLE II. Signal intensity summary for four post-brachytherapy patients. PZ signals are standardized to $\mu=0$ and $\sigma=1$. Tumor signals are normalized with the mean and standard deviation of the PZ for comparison. $\mu$ and $\sigma$ of nonstandardized ADC and T2 Map values are also presented.

| Image type | PZ ($\mu\pm\sigma$) | Tumor ($\mu\pm\sigma$) |
|---|---|---|
| T2W | $0\pm1$ | $-0.04\pm0.89$ |
| ADC | $0\pm1$ | $-0.60\pm0.83$ |
| PD | $0\pm1$ | $0.30\pm1.19$ |
| T2 Map | $0\pm1$ | $-0.49\pm0.71$ |
| ADC (in $\mu^2$/ms) | $1.524\pm0.306$ | $1.250\pm0.314$ |
| T2 Map (in ms) | $88.0\pm20.7$ | $76.6\pm15.9$ |

### Classifier accuracy

Figure 2 shows two sample summary statistical maps generated by FLD and SVM classifier and Fig. 3 shows several ROC curves from our best single-channel ML classifier and our best multichannel classifier. The results for the single-channel ML classifiers, multichannel FLD classifiers, and SVM classifiers are shown in Table III. Pairwise two-sided *t*-tests with $\alpha=0.05$ among the four single-channel classifiers supported the hypothesis that all single-channel classifiers based on intensity alone had statistically equivalent ROC performance ($P>0.05$).

The comparison between the multichannel FLD classifiers with T2W axial ML classifier, the best of the four single-channel classifiers indicated the following: pairwise two-sided *t*-tests with $\alpha=0.05$ supported that FLD with *basic4 +anatomy* ($P$-value$=0.0024$), *all CM* ($P$-value$=0.0004$), *all DCT* ($P$-value$=0.002$), and *all CM+DCT* ($P$-value $=0.0003$) offered greater classification power than ML classifier based on T2W axial intensity alone. However, pairwise two-sided *t*-tests with $\alpha=0.05$ did not support that FLD with *basic4* classifier performed better than single-channel T2W axial ML classifier ($P$ value$=0.355$). Similarly, we compared the multichannel SVM classifiers with T2W axial ML classifier. Pairwise two-sided *t*-tests with $\alpha=0.05$ supported that SVM with *basic4+anatomy* ($P$ value$=0.0017$) offered greater classification power than ML classifier based on T2W axial intensity alone. However, pairwise two-sided *t*-tests with $\alpha=0.05$ did not support that SVM with *basic4* classifier performed better than single-channel T2W axial ML classifier ($P$ value$=0.483$).

The training time for FLD classifiers is between 1 to 6 hours and for SVM classifiers that converged numerically, between 2 to 60 hours. We failed to get SVM convergence for the feature sets *all CM*, *all DCT*, and *all CM+DCT* after 72 hours of simulation. For the other three feature sets *basic4* and *basic4+anatomy* that we managed to get convergence, we compared the average area under ROC of SVM classifier with FLD classifier using two-sided pairwise *t*-test at $\alpha=0.05$. SVM performs better than FLD for *basic4 +anatomy* ($P$ value$=0.0003$) while SVM and FLD have equal performance statistically for *basic4* ($P$ value$=0.29$).

## DISCUSSION

The study demonstrates that it is possible to create summary statistical maps that combine multiparametric data. Furthermore, the information from the data sets in combination can be more powerful than any one alone, supporting the need to increase the number of tools and methods to test this approach further.

Students' *t*-test comparisons of ROC areas suggested that SVM produced greater detection power than FLD classifiers for the feature sets *all CM* and *basic4+anatomy* and statistically equivalent detection power for the *basic4* feature set. The merits of the nonlinear decision boundary of the SVM classifier become noticeable when the number of features increases. Although SVM achieved a better performance

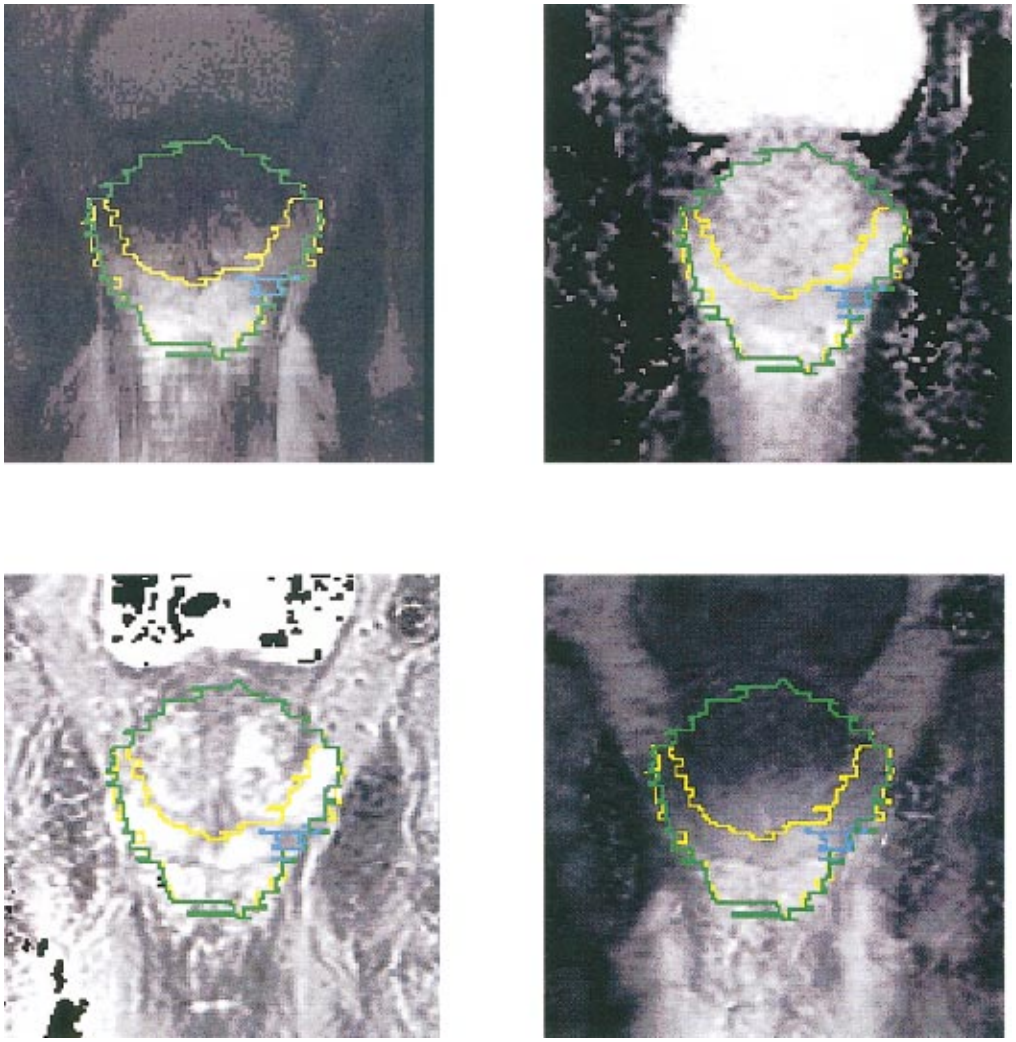

FIG. 1. A sample set of multiparametric MR images in the oblique coronal plane. Top left is a T2W image, resampled from the axial planes to the oblique coronal planes of the other images. Top right is an ADC Map from LSDI. Bottom left and bottom right are T2 Map and proton density images from T2-mapping. The green label is total gland, the yellow label is PZ, and the blue label is biopsy validated tumor label.

than FLD, classifier training convergence and simulation time are two issues to consider.

The training time for FLD classifiers is between 1 to 6 hours and for the converged SVM classifiers between 2 to 60 hours, depending on number of classification channels and the number of samples. We included all tumor samples from the nine cases for training (507 voxels) and 10% of healthy PZ samples (933 voxels). We failed to get convergence on SVM training after 72 hours for the larger feature sets *all* $CM+DCT$ and *all DCT*. This is because both of these feature sets include over 150 channels. The FLD results suggest that feature sets *all CM* and *all* $CM+DCT$ have statistically equivalent performance and therefore, we expected similar findings for SVM had the *all* $CM+DCT$ training converged. To include both frequency and CM features, we could have randomly selected a smaller number of CM entries and DCT frequencies to reduce the dimensionality of the problem as many CM entries and DCT frequencies contain high mutual information.

Utilization of CM and DCT significantly enhanced tumor features in the images as supported by *t*-test analysis. For both FLD and SVM classifiers, we noticed that *all CM*, *all DCT* and *all* $CM+DCT$ performed better than *basic4* and *basic4*+*anatomy* feature sets according to the ROC area analysis, which proved the effectiveness of these machine vision techniques for prostate cancer detection with MRI.

We found as expected that the group of patients who had undergone prostate brachytherapy had different T2W signal intensity properties compared to the group without brachytherapy. The results in Tables I and II showed that the post-brachytherapy patients almost had no difference in mean T2W signal intensity between PZ and tumor tissues while for the nonbrachytherapy patients, the difference in standardized means of T2W signal intensities between PZ and tumor tissues is 0.58. From the T2W images of post-brachytherapy patients, one can observe that the PZ intensity is darkened as a result of brachytherapy and therefore, it is difficult to differentiate between tumor and PZ tissues based on T2W intensity level. This finding is consistent with the quantitative shortening of T2 observed in the post-brachytherapy patients

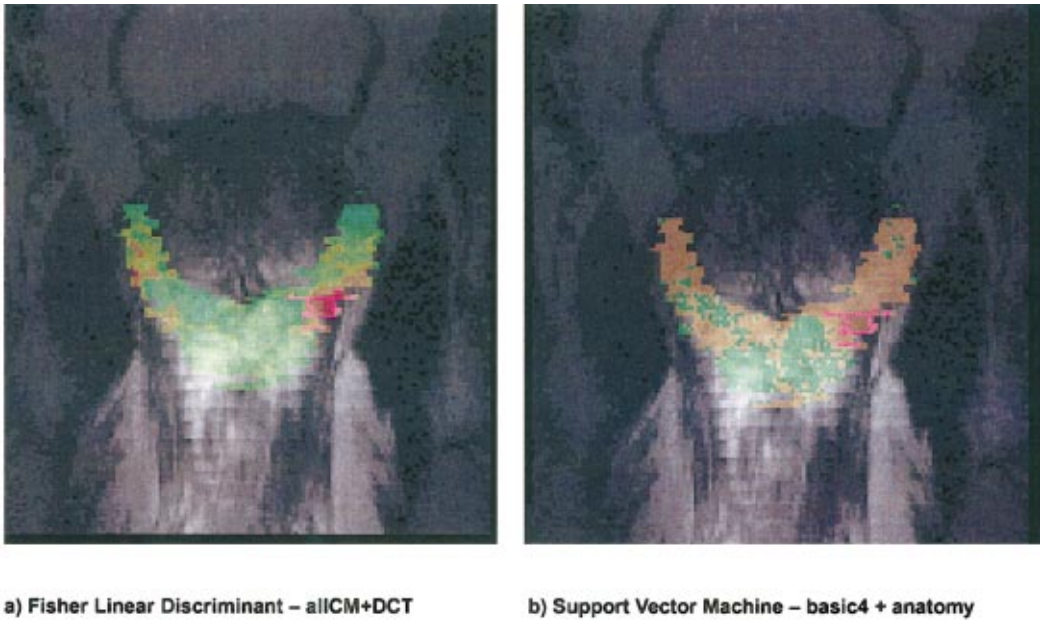


FIG. 2. Summary statistical maps of (a) Fisher linear discriminant (FLD) classifier and (b) support vector machine (SVM) classifier. The FLD classifier utilizes all co-occurrence, DCT, anatomical and signal intensity features and the SVM classifier utilizes signal intensity and anatomical features only. The statistical maps of the PZ are superimposed on the T2W axial images of the patient in Fig. 1 and the magenta label indicates the biopsy-validated tumor region identical to Fig. 1. The statistical maps use a rainbow color scheme with red indicating high tumor likelihood and green indicating low tumor likelihood. One can observe that both statistical maps correctly pick out the tumor area by shading it with red and most of the nontumor areas with green.

compared to the nonbrachytherapy groups (Tables I and II). Specifically, the radiation damage to tissue results in a roughly 27% decrease in T2 regardless of whether the tissue has been characterized as normal PZ or TU. We attribute this to radiation induced fibrosis which causes T2 shortening.

Regarding the trace ADC values obtained in our study, a decrease-between cancerous PZ and noncancerous PZ was found for both groups. The noncancerous mean PZ ADC values we found were 1.6 $\mu m^2$/ms, somewhat smaller than the 1.8 $\mu m^2$/ms mean ADC reported by Issa.[17] The mean cancerous ADC values in the PZ we found were 1.43 $\mu m^2$/ms, which were very similar to the 1.38 $\mu m^2$/ms mean ADC reported by Issa.

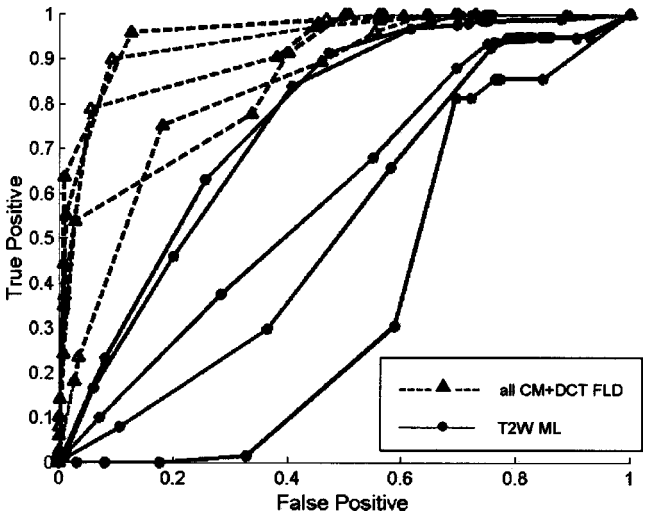


FIG. 3. ROC curves for five patients. The circle series of curves are from the single-channel T2W maximum likelihood (ML) classifier and the triangle series of curves are from the *all CM+DCT* Fisher linear discriminant (FLD) classifier. Although ROC curves can be processed to be convex and above the diagonal by randomized decisions, these sets of ROCs are derived from empirical data without processing. From the ROC curves, we notice that the *all CM+DCT* FLD classifier outperforms the T2W ML classifier in these five cases.

From the results of the single-channel ML classifiers in Table III, one may draw the conclusion that the T2W axial images are most informative about differentiating prostate tumor out of the four signal intensities, with the largest average ROC area of 0.599. Yet, we found that the four image intensities are equally informative about tumor statistically. Another explanation for the slightly larger ROC area with T2W is that we obtained ''ground truth'' tumor label with a radiologist contouring axial T2W images. This could lead to additional bias even though the ''ground truth'' label is confirmed by biopsy reports. Furthermore, the radiologist examined the axial T2W images, which have a higher spatial resolution than the LSDI and T2 Map images. Therefore, it can be self-serving that we found the T2W images to be most informative about tumor when we defined the ''ground truth'' tumor regions by expert examination of these images. However, this limitation is difficult to avoid because we had no access to the ''ground truth.'' Confirming the radiologist's ''ground truth'' label with biopsy reports only partially remedied this bias.

The average ROC results show that both SVM and FLD classifiers with *basic4* feature set did not perform better than the best single-channel ML classifier based on T2W intensities according to *t*-test analysis. However, it is not valid to conclude that LSDI and T2 Map did not add any useful in-

TABLE III. Summary of maximum likelihood (ML), Fisher linear discriminant (FLD), and support vector machine (SVM) classifiers results of 10 nonbrachytherapy patients. Mean(std dev) of area under ROC of each classifier are presented. T2W axial, ADC Map, proton density and T2 Map are the four basic image intensities and the classifiers for these single-channel cases are based on maximum likelihood. "*basic4*" consists of the four basic image intensities. "*basic4+anatomy*" consists of the four basic image intensities and the three cylindrical coordinates that describe anatomical location relative to the centroid of the prostate. "*all DCT*" consists of all four basic intensities, anatomical information, and discrete cosine transform frequency statistics for all four basic images. "*all CM*" consists of all four basic intensities, anatomical information, and co-occurrence statistics for all four basic images. "*all CM+DCT*" consists of all intensity, co-occurrence, anatomical, and frequency features.

| Classifier | Features | ROC area: $\mu$ ($\sigma$) |
|---|---|---|
| ML | T2W axial | 0.599 (0.146) |
| | ADC Map | 0.533 (0.114) |
| | Proton density | 0.521 (0.165) |
| | T2 Map | 0.562 (0.058) |
| FLD | *basic4* | 0.620 (0.089) |
| | *basic4+anatomy* | 0.729 (0.058) |
| | *all CM* | 0.825 (0.056) |
| | *all DCT* | 0.791 (0.043) |
| | *all CM+DCT* | 0.839 (0.064) |
| SVM | *basic4* | 0.635 (0.079) |
| | *basic4+anatomy* | 0.761 (0.043) |
| | *all CM* | no training convergence |
| | *all DCT* | |
| | *all CM+DCT* | |

formation. Although not statistically significant, we notice that the average ROC area for the ML T2W classifier is 0.599, which is less than 0.620 and 0.635 of the FLD *basic4* and SVM *basic4*, respectively. Further studies need to be done to compare the textural information in PD, ADC, and T2 Map images and their mutual information with the textural features in T2W images.

T2W intensities in the PZ near the rectum were corrupted by sharp near-field endorectal coil artifacts. This largely limited the tumor detection ability of the T2W images for PZ tissues near the coil. The locations of the coil artifacts often coincide with PZ where tumors often develop. Advances in intensity correction methodologies may remove the coil artifacts, which can dramatically improve the quality of T2W images and boost classifier performance.

We adopted a jack-knife strategy when performing classifier assessment because we only had 10 patients with pathologically confirmed tumor in the nonbrachytherapy group. As we collect more multiparametric cases, we could divide patients into two separate groups for classifier training and validation. A larger number of patients would provide image samples that covered a broader spectrum of focal abnormalities and cancer.

In general, we found that adding textural and anatomical information increases the accuracy of all three classifiers used in this study. We also found that SVM is the best of the three classification techniques provided that we achieved SVM training convergence. Both observations are consistent with our hypotheses.

## CONCLUSIONS

Integrating information from multiple images and enhancing prostate tumor features in these images were the two main objectives of this study. We have demonstrated the utility of two multichannel classifiers with feature enhancements using machine vision techniques for prostate cancer detection. We have also shown that our classifiers have statistically superior performance over single-channel intensity-based classifiers. The summary statistical map generated by our classifiers allows radiologists to visualize the high volume of image data and provides summarized preoperative information for intraoperative procedures. The summary statistical map has the potential of improving biopsy accuracy and enhancing tumor target identification for the delivery of localized therapies.

## ACKNOWLEDGMENTS

This work was in part supported by the following NIH grants: R01 AG19513-01, AG 5 P01CA67165-03, 1 R33 CA99015, 1RO1 NS39335-01A1, and R03HS13234-01.

[a]Author to whom correspondence should be addressed; MRI Division, Department of Radiology, Brigham and Women's Hospital, L1-050, 75 Francis St., Boston, MA 02115; electronic mail: ctempany@bwh.harvard.edu

[1] *Cancer Facts and Figures* (American Cancer Society, Atlanta, GA, 1997).
[2] L. Garfinkel and M. Mushinski, "Cancer incidence, mortality, and survival trends in four leading sites," Stat. **75**, 19–27 (1994).
[3] F. Sommer, J. McNeal, and C. Carrol, "MR depiction of zonal anatomy of the prostate at 1.5T," J. Comput. Assist. Tomogr. **10**, 983–989 (1986).
[4] M. Schnall, R. Lenkinski, H. Pollack, Y. Imai, and H. Kressel, "Prostate: MR imaging with an endorectal surface coil," Radiology **172**, 570–574 (1989).
[5] M. Schnall and H. Pollack, "Magnetic resonance imaging of the prostate gland," Urol. Radiol. **12**, 109–114 (1990).
[6] M. Chelsky, M. Schnall, E. Seidmon, and H. Pollack, "Use of endorectal surface coil magnetic resonance imaging for local staging of prostate cancer," Urology **150**, 391–395 (1993).
[7] D. Cheng and C. Tempany, "MR imaging of the prostate and bladder," Semin. Ultrasound CT MR **19**, 67–89 (1998).
[8] H. Carter, R. Brem, C. Tempany, A. Yang, J. Epstein, P. Walsh, and E. Zerhouni, "Nonpalpable prostate cancer: detection with MR imaging," Radiology **178**, 523–525 (1991).
[9] G. Liney, L. Turnbull, M. Lowry, L. Turnbull, A. Knowles, and A. Horsman, "*In vivo* quantitation of citrate concentration and water T2 relaxation time of the pathologic prostate gland using $^1$H MRS and MRI," Magn. Reson. Imaging **15**, 1177–1186 (1997).
[10] G. Liney, M. Lowry, L. Turnbull, D. Manton, A. Knowles, S. Blackband, and A. Horsman, "Proton MR T2 maps correlate with the citrate concentration in the prostate," NMR Biomed. **9**, 59–64 (1996).
[11] G. Liney, A. Knowles, D. Manton, L. Turnbull, S. Blackband, and A. Horsman, "Comparison of conventional single echo and multi-echo sequences with a fast spin-echo sequence for quantitative T2-mapping: Application to the prostate," J. Magn. Reson. Imaging **6**, 603–607 (1996).
[12] J. Garcia-Segura, M. Sanchez-Chapado, C. Ibarburen, J. Viano, J. Angulo, J. Gonzalez, and J. Rodriguez-Vallejo, "*In vivo* proton spectroscopy of diseased prostate: Spectroscopic features of malignant versus benign pathology," Magn. Reson. Imaging **17**, 755–765 (1999).
[13] P. Schaefer, P. Ellen Grant, and R. Gilberto Gonzalez, "Diffusion-weighted MR imaging of the brain," Radiology **217**, 331–345 (2000).
[14] R. Gonzalez, P. Schaefer, F. Buonanno, L. Schwamm, R. Budzik, G. Rordorf, B. Wang, A. Sorensen, and W. Koroshetz, "Diffusion-weighted

MR imaging: Diagnostic accuracy in patients imaged within 6 hours of stroke symptom onset,'' Radiology **210**, 155–162 (1999).
[15] T. Chenevert, L. Stegman, J. Taylor, P. Robertson, H. Greenberg, A. Rehemtulla, and B. Ross, ''Diffusion magnetic resonance imaging: an early surrogate marker of therapeutic efficacy in brain tumors,'' J. Natl. Cancer Inst. **92**, 2029–2036 (2000).
[16] M. Zhao, J. Pipe, J. Bonnett, and J. Evelhoch, ''Early detection of treatment response by diffusion-weighted 1H-NMR spectroscopy in a murine tumour *in vivo*,'' Br. J. Cancer **73**, 61–64 (1996).
[17] R. Robertson, S. E. Maier, R. V. Mulkern, S. Vajapeyam, C. Robson, and P. Barnes, ''MR line-scan diffusion imaging of the spinal cord in children,'' Am. J. Neuroradiol. **21**, 1344–1348 (2000).
[18] H. Gudbjartsson, S. E. Maier, R. V. Mulkern, I. Morocz, S. Patz, and F. Jolesz, ''Line scan diffusion imaging,'' Magn. Reson. Med. **36**, 509–518 (1996).
[19] S. E. Maier, H. Gudbjartsson, S. Patz, L. Hsu, K. Lovblad, R. Edelman, S. Warach, and F. Jolesz, ''Line scan diffusion imaging: Characterization in healthy subjects and stroke patients,'' AJR, Am. J. Roentgenol. **171**, 85–93 (1998).
[20] B. Issa, ''*In vivo* measurement of the apparent diffusion coefficient in normal and malignant prostatic tissues using echo-planar imaging,'' J. Magn. Reson Imaging **16**, 196–200 (2002).
[21] M. Vannier, R. Butterfield, D. Rickman, D. Jordan, W. Murphy, and P. Biondetti, ''Multispectral magnetic resonance image analysis,'' Radiology **154**, 221–224 (1985).
[22] H. Cline, E. Lorensen, R. Kikinis, and F. Jolesz, ''Three-dimensional segmentation of MR images of the head using probability and connectivity,'' J. Comput. Assist. Tomogr. **14**, 1037–1045 (1990).
[23] R. Duda and P. Hart, *Pattern Classification and Scene Analysis* (Wiley, New York, 1973).
[24] C. Bishop, *Neural Networks for Pattern Recognition* (Clarendon, Oxford, 1995).
[25] V. Vapnik, *The Nature of Statistical Learning Theory* (Springer, New York, 1995).
[26] C. Burges, ''A tutorial on support vector machines for pattern recognition,'' Data mining knowledge discovery **2**, 121–167 (1998).
[27] P. Golland *et al.*, ''Small sample size learning for shape analysis of anatomical structures, Proc. MIOCAI'' Lect. Notes Comput. Sci. **1935**, 72–82 (2000).
[28] R. Haralick, K. Shanmugan, and I. Dinstein, ''Texture for image classification,'' IEEE Trans. Syst. Man. Cybern. **3**, 610–621 (1973).
[29] G. Torheim *et al.*, ''Feature extraction and classification of dynamic contrast enhanced T2-weighted breast image data,'' IEEE Trans. Med. Imaging **20**, 12 (2001).
[30] J. Kim and H. Park, ''Statistical texture features for detection of microcalcifications in digitized mammograms,'' IEEE Trans. Med. Imaging **18**, 3 (1999).
[31] P. Freeborough and N. Fox, ''MR image texture analysis applied to the diagnostic and tracking of Alzheimer's disease,'' IEEE Trans. Med. Imaging **17**, 3 (1998).
[32] J. Qiang and E. Craine, ''Texture analysis for classification of cervix lesions,'' IEEE Trans. Med. Imaging **19**, 11 (2000).
[33] J. R. Reichenbach, T. Hacklander, T. Harth, M. Hofer, M. Rassek, and U. Modder, ''1H T1 and T2 measurements of the MR imaging contrast agents Gd-DTPA and Gd-DTPA BMA at 1.5 T,'' Eur. Radiol. **7**, 264–274 (1997).
[34] A. Oppenheim and R. Schafer, *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1989).
[35] A. Jain, *Fundamentals of Digital Image Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1989).
[36] G. Cawley, MATLAB Support Vector Machine Toolbox v0.50, http://theoval.sys.uea.ac.uk/~gcc/svm/toolbox University of East Anglia, 2000.
[37] J. Platt, ''Fast training of support vector machines using sequential minimal optimization,'' in *Advances in Kernel Methods—Support Vector Learning*, edited by B. Scholkopf, C. Burges, and A. Smola (MIT Press, Cambridge, MA, 1999), Chap. 12, pp. 185–208.
[38] Mathworks Inc. http://www.mathworks.com
[39] Sun Microsystem Java software http://java.sun.com
[40] C. Metz, ''ROC methodology in radiologic imaging,'' Invest. Radiol. **21**, 720–733 (1986).