

Computer-extracted Features Can Distinguish Noncancerous Confounding Disease from Prostatic Adenocarcinoma at Multiparametric MR Imaging¹

Geert J. S. Litjens, MSc
Robin Elliott, MD, PhD
Natalie NC Shih, MD, PhD
Michael D. Feldman, MD, PhD
Thiele Kobus, PhD
Christina Hulsbergen-van de Kaa, MD, PhD
Jelle O. Barentsz, MD, PhD
Henkjan J. Huisman, PhD
Anant Madabhushi, PhD

¹ From the Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Geert Grooteplein 10, 6525 GA, Nijmegen, the Netherlands (G.J.S.L., T.K., C.H.v.d.K., J.O.B., H.J.H.); Departments of Pathology (R.E.) and Biomedical Engineering (A.M.), Case Western University Hospitals, Cleveland, Ohio; and Department of Pathology, University of Pennsylvania, Philadelphia, Pa (N.N.C.S., M.D.F.). Received December 12, 2014; revision requested February 4, 2015; revision received March 20; accepted April 16; final version accepted April 24. Supported by the National Institutes of Health; the Dutch Cancer Society; the National Institute of Diabetes and Digestive and Kidney Diseases (R01DK098503-02); the Department of Defense Prostate Cancer Synergistic Idea Development Award (PC120857), Lung Cancer Idea Development New Investigator Award (LC130463), and Prostate Cancer Idea Development Award; the Ohio Third Frontier Technology Development grant; the CTSC Coulter Annual Pilot Grant; the Case Comprehensive Cancer Center Pilot Grant; the VelaSano grant from the Cleveland Clinic; and the Wallace H. Coulter Foundation Program, Department of Biomedical Engineering, Case Western Reserve University. **Address correspondence to** G.J.S.L. (e-mail: geert.litjens@radboudumc.nl).

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

© RSNA, 2015

Purpose:

To determine the best features to discriminate prostate cancer from benign disease and its relationship to benign disease class and cancer grade.

Materials and Methods:

The institutional review board approved this study and waived the need for informed consent. A retrospective cohort of 70 patients (age range, 48–70 years; median, 62 years), all of whom were scheduled to undergo radical prostatectomy and underwent preoperative 3-T multiparametric magnetic resonance (MR) imaging, including T2-weighted, diffusion-weighted, and dynamic contrast material-enhanced imaging, were included. The digitized prostatectomy slides were annotated for cancer and noncancerous disease and coregistered to MR imaging with an interactive deformable coregistration scheme. Computer-identified features for each of the noncancerous disease categories (eg, benign prostatic hyperplasia [BPH], prostatic intraepithelial neoplasia [PIN], inflammation, and atrophy) and prostate cancer were extracted. Feature selection was performed to identify the features with the highest discriminatory power. The performance of these five features was evaluated by using the area under the receiver operating characteristic curve (AUC).

Results:

High-*b*-value diffusion-weighted images were more discriminative in distinguishing BPH from prostate cancer than apparent diffusion coefficient, which was most suitable for distinguishing PIN from prostate cancer. The focal appearance of lesions on dynamic contrast-enhanced images may help discriminate atrophy and inflammation from cancer. Which imaging features are discriminative for different benign lesions is influenced by cancer grade. The apparent diffusion coefficient appeared to be the most discriminative feature in identifying high-grade cancer. Classification results showed increased performance by taking into account specific benign types (AUC = 0.70) compared with grouping all noncancerous findings together (AUC = 0.62).

Conclusion:

The best features with which to discriminate prostate cancer from noncancerous benign disease depend on the type of benign disease and cancer grade. Use of the best features may result in better diagnostic performance.

© RSNA, 2015

Online supplemental material is available for this article.

Magnetic resonance (MR) imaging is becoming an increasingly used tool for diagnosing prostate cancer, which has led to the development of standardized guidelines for acquiring, reading, and reporting prostate MR images and findings by the European Society of Urogenital Radiologists: the Prostate Imaging and Reporting Data Standard (PIRADS) (1–3). Initial results that used the European Society of Urogenital Radiologists PIRADS guidelines have been promising, both with respect to overall performance and interreader agreement (4–6). However, these initial studies have also shown a large trade-off between sensitivity and specificity, depending on the PIRADS score that was used as a threshold for biopsy.

In a recent publication by Rosenkrantz et al (7), four out of the 10 named pitfalls in prostate MR imaging are related to noncancerous disease mimicking the appearance of cancer. Some typical examples are prostatic intraepithelial neoplasia (PIN), atrophy,

inflammation, and benign prostatic hyperplasia (BPH).

Some previous research has focused on identifying discriminatory features to separate cancer from specific benign disease (8–10). Oto et al (8) investigated the use of apparent diffusion coefficient (ADC) values to differentiate between central gland (transition plus central zone) tumors and glandular and stromal hyperplasia by visually registering pathologic slides to MR images. They were able to achieve an area under the receiver operating characteristics curve (AUC) of 0.78 and 0.99 for differentiating stromal and glandular hyperplasia and prostate cancer, respectively. Liu et al (9) designed a biexponential diffusion model with 10 *b* values to characterize central gland lesions as prostate cancer and BPH. They found that the biexponential model substantially improved the discriminative performance of diffusion-weighted imaging (AUC = 0.92) compared with a monoexponential model (AUC = 0.80). Ginsberg et al (10) investigated the relative importance of MR imaging sequences in depicting cancer. They showed that T2-weighted imaging is most important to identify cancer in the transition zone, whereas diffusion-weighted imaging was most important to identify cancer in the peripheral zone.

Recently, several groups started investigating the potential of computer-extracted features to improve the diagnosis of cancer at MR imaging, an application that was successful for both breast and prostate MR imaging (10,11). The general concept uses image analysis algorithms to extract subvisual image features that are not readily apparent to the human visual system. A good example of this is texture features, which can enhance edges or, conversely, areas of intensity homogeneity.

Implication for Patient Care

- Increased understanding of the imaging characteristics of benign disease can improve the ability of radiologists to distinguish noncancerous disease from prostate cancer.

Our hypothesis is that, by investigating the use of computer-extracted features in the context of discriminating different benign disease classes from prostate cancer, we can provide initial input for more granular diagnostic guidelines as an addendum to the established PIRADS reporting strategy for prostate MR imaging. The purpose of this work was to determine the best features to discriminate prostate cancer from benign disease and their relationship to the class of benign disease and cancer grade.

Materials and Methods

Patients

The institutional review board approved this retrospective study and waived the requirement of informed consent. In the period from January 1, 2009, to June 1, 2013, a total of 271 patients underwent

Advances in Knowledge

- Benign prostatic hyperplasia is best discriminated from cancer with high-*b*-value diffusion-weighted images.
- Areas of inflammation typically have a more focal appearance than prostate cancer on dynamic contrast-enhanced and T2-weighted images.
- The most useful computer-extracted features from T2-weighted, diffusion-weighted, and dynamic contrast-enhanced MR images that help distinguish prostate cancer from benign disease vary and depend on cancer grade.
- Use of only the most useful features to discriminate each class of benign disease from cancer results in a higher overall area under the receiver-operating characteristic curve than when all features are used (0.70 vs 0.62).

Published online before print

10.1148/radiol.2015142856 Content codes: **GU** **MR** **01**

Radiology 2015; 000:1–11

Abbreviations:

ADC = apparent diffusion coefficient
 AUC = area under the receiver operating characteristics curve
 BPH = benign prostatic hyperplasia
 PIN = prostatic intraepithelial neoplasia
 PIRADS = Prostate Imaging and Reporting Data Standard
 ROI = region of interest
 SFFS = sequential forward floating feature selection

Author contributions:

Guarantors of integrity of entire study, G.J.S.L., J.O.B., H.J.H., A.M.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, G.J.S.L., H.J.H., A.M.; clinical studies, M.D.F., C.H.v.d.K., J.O.B.; experimental studies, G.J.S.L., M.D.F., T.K., H.J.H., A.M.; statistical analysis, G.J.S.L., H.J.H., A.M.; and manuscript editing, G.J.S.L., N.N.C.S., M.D.F., T.K., J.O.B., H.J.H., A.M.

Funding:

This research supported by the National Institutes of Health (grants R01CA136535-01, R01CA140772-01, R21CA167811-01, R21CA179327-01, and R21CA195152-01).

Conflicts of interest are listed at the end of this article.

prostatectomy at our institution, 190 of whom were excluded because of equipment availability and time constraints, allowing us to digitize histopathologic slides for only 81 of the 271 patients. Of these 81 patients, all 41 patients from 2012–2013 with whole-mount prostatectomy specimens (ie, specimens that were not cut into parts to fit on regular glass slides) and who underwent multiparametric MR imaging were included. The remaining 40 patients were randomly selected from the remaining patients who underwent multiparametric MR imaging prior to prostatectomy between 2009 and 2011. For these 40 patients, the prostatectomy specimens were cut into parts to fit onto regular glass slides.

For these 81 patients, preoperative multiparametric MR imaging and radical prostatectomy specimens were retrospectively reviewed for inclusion in this study. Five patients were excluded because of previous treatment, three patients were excluded because the quality of MR images was too poor for analysis (eg, due to the presence of hip implants), and three patients were excluded because multiparametric MR imaging did not include all sequences. Thus, a total of 70 patients were included (age range, 48–70 years; median, 62 years). The time between MR imaging and prostatectomy ranged from 2 weeks to 6 months (mean, 10 weeks).

MR Image Acquisition

MR imaging was performed with a 3-T MR imager (TrioTim or Skyra; Siemens, Erlangen, Germany). In some cases, images were acquired with an endorectal coil ($n = 34$), and some were acquired without an endorectal coil ($n = 36$). A pelvic phased-array coil was always used. The multiparametric protocol consisted of three T2-weighted images obtained in orthogonal directions, diffusion-weighted images obtained with three b values (50, 400–500, and 800 sec/mm²) and averaged over three orthogonal directions to obtain a direction-insensitive measurement, and dynamic contrast material-enhanced images obtained

Table 1

MR Imaging Sequence Details for the Different Types of Acquisitions

Parameter	T2 Weighted	DW	DCE
Sequence name	Turbo spin-echo	Echoplanar	FLASH spoiled GRE
Spatial resolution (mm)	0.28–0.6	1.6–2	1.5–1.8
Acquisition matrix	320 × 320–384 × 384	128 × 128	128 × 128
Field of view (mm)	108 × 108–192 × 192	256 × 256	192 × 192–230 × 230
Number of averages	1–2	6–10	1
Section thickness (mm)	3.0–4.0	3.0–4.0	3.2–5.0
Number of sections	13–19	15–20	12–15
Echo time (msec)	101–104	61–81	1.41–1.47
Repetition time (msec)	3540–6840	2300–3600	36
Flip angle (degree)	120–160	90	10–14

Note.—T2-weighted images are acquired in three orthogonal directions (transverse, sagittal, and coronal). Diffusion-weighted images are obtained with three b values (50, 400–500, and 800 sec/mm²) averaged over three directions, and ADC maps are calculated with the imaging software. Some images also include a b of 0 sec/mm² and generalized autocalibrating partially parallel acquisition (GRAPPA) of 2. Dynamic contrast-enhanced images have temporal resolution of 3.38–4.65 sec and 36–50 time points, and 15 mL of contrast material is used. DCE = dynamic contrast-enhanced, DW = diffusion weighted, FLASH = fast low angle shot, GRE = gradient recalled echo.

with 3.4–4.7-second temporal resolution and 15 mL of Dotarem (Guerbet, Paris, France). Transverse T2-weighted images were acquired perpendicular to the rectal wall; diffusion-weighted and dynamic contrast-enhanced images were acquired in the same orientation. Further acquisition details are detailed in Table 1.

Prostatectomy Slide Selection and Annotation

After radical prostatectomy, prostate specimens were uniformly processed and submitted for histologic investigation in their entirety. Immediately after surgical resection, specimens were fixed in 10% neutral buffered formalin by using fine needle formalin injections and were stored overnight. Subsequently, the entire surface was marked with ink by using three different colors, then the entire prostate specimen was cut into serial transverse 4.0-mm-thick sections. Sections were cut perpendicular to the dorsal-rectal surface to ensure the same section orientation as in the MR imaging sequences. All sections were macroscopically photographed with a charge-coupled device camera. Thin portions of the sections were stained with hematoxylin-eosin stain and evaluated by one expert urologic pathologist (C.H.-v.d.K., with 17 years of experience).

The hematoxylin-eosin-stained section that contained the tumor with the highest Gleason score was selected to be digitized with a digital slide scanner (VS120-S5; Olympus, Tokyo, Japan) and ×10 or ×20 magnification, which corresponds to a resolution of 0.6 and 0.3 μm, respectively. If multiple sections contained tumor with the same Gleason grade, the section with the largest tumor volume was digitized.

Approximately one-half ($n = 41$) of the specimens were whole-mount sections, and the other one-half ($n = 40$) consisted of parts of sections (usually two or four). When the specimen consisted of parts, all parts belonging to one slide were digitized.

After digitization, the digital slides were annotated by using the freehand annotation tool in Aperio ImageScope (Aperio, Vista, Calif) for the presence of cancer, BPH, PIN, atrophy, or inflammation by one of two pathologists (N.S., with 8 years of experience, or R.E., with 7 years of experience). No minimum sizes were specified for the tumor annotations, and all foci of tumor were annotated. If tumor foci were close but not touching, they were annotated as separate foci. For benign lesions, the annotating pathologists were asked to annotate representative areas (larger, connected areas with no or little mixing) for each

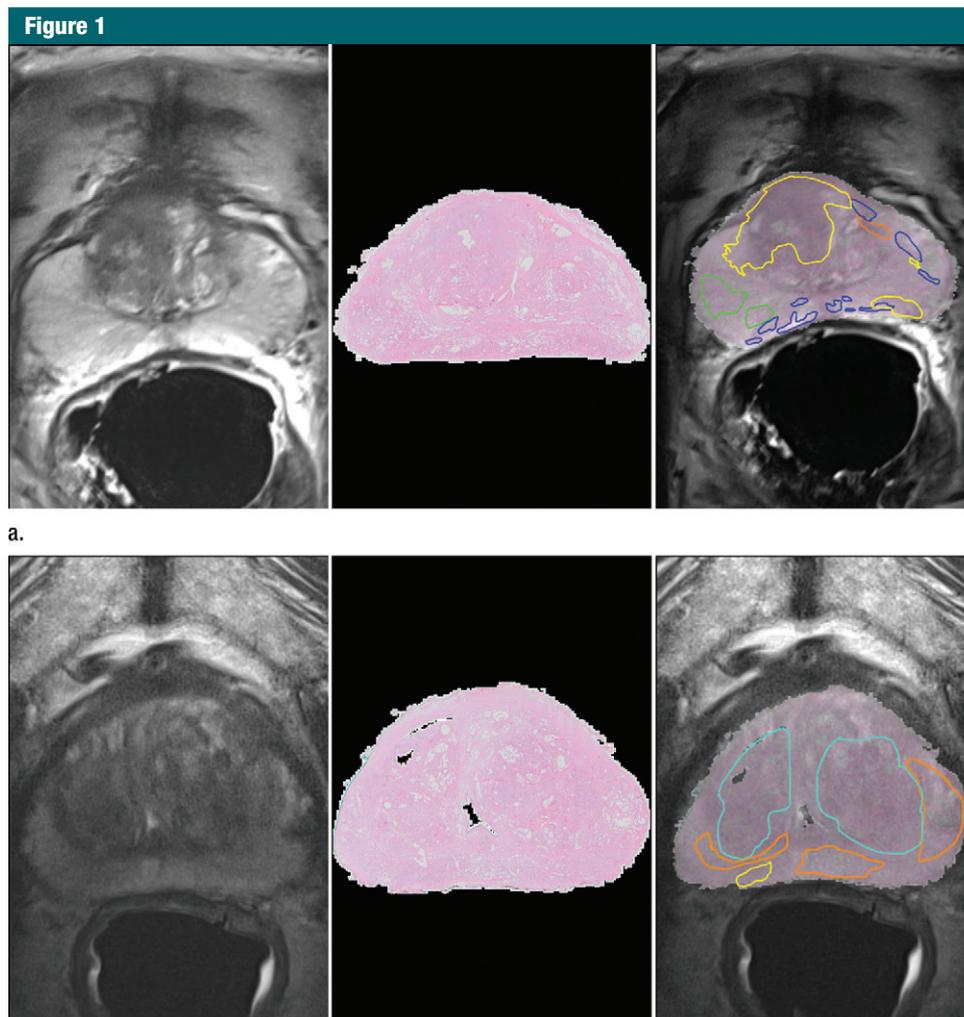


Figure 1: Pathologic-MR imaging mapping procedure in two patients. **(a)** MR image (left), pathologic image (center), and MR image overlaid with pathologic image (right) show prostate cancer (yellow line). The large lesion has a Gleason score of 3 + 4, and the other two lesions have a Gleason score of 3 + 3. Inflammation (green line), PIN (blue line), and atrophy (orange line) are also seen. **(b)** MR image (left), pathologic image (center), and MR image overlaid with pathologic image (right) show a lesion with a Gleason score of 3 + 4 (yellow line), atrophy (orange line), and BPH (blue line).

class, but no minimum size was specified. Diffusely spread out, nonmalignant lesions were not included because they would result in annotations with widths smaller than even the smallest MR imaging voxel. For example, inflammation can have multiple thin strands extending into the tissue. These strands would have a width much smaller than the MR imaging voxel size, making it impossible to correctly characterize the inflammation at MR imaging. Furthermore, no specific subdivisions were made according to benign disease types, such as chronic or

acute inflammation and stromal or glandular BPH; thus, those classes contain both types.

Coregistration of Prostatectomy Specimens and MR Imaging

To map the annotations on the histopathologic sections to the corresponding MR imaging sections, the MR image and the pathologic slide have to be registered. First, the section in the MR image that corresponds to the prostatectomy section has to be established (12). The number of MR imaging

sections that showed the prostate were counted, then the number of sections in the prostatectomy specimen were counted. By using the number associated with each prostatectomy glass slide, the most likely corresponding MR imaging section was identified with the following equation:

$$S_{MR} = \frac{T_{MR}}{T_P} S_P,$$

where S_{MR} is the section number in MR imaging, T_{MR} the total number of

Table 2

Overview of All the Features That Are Used in This Article Including References to Relevant Articles

Category and Feature	Pulse Sequence Used	Parameter
Signal intensity		
T2-weighted	Axial T2 weighted	...
ADC	Diffusion weighted	...
$b = 800 \text{ sec/mm}^2$ *	Diffusion weighted	...
Texture		
2D multiscale Gauss	T2 weighted	Up to second order, $\sigma = 2.0, 2.7, 4.1$ and 6.0 mm
2D multiangle Gabor filter	T2 weighted	$\theta = 0, \pi/4, \pi/2, 3\pi/4, \lambda = 2, 3,$ and 4 mm
2D Li multiscale Hess focality filter	T2 weighted, ADC, $b = 800 \text{ sec/mm}^2$, $K_{\text{trans}}, K_{\text{ep}}, V_e, \text{TTP, max, WOR}$	$\sigma = 2.0, 2.7, 4.1,$ and 6 mm
Pharmacokinetic		
TTP	DCE	...
Max	DCE	...
WOR	DCE	...
K_{trans}	DCE	...
V_e	DCE	...
K_{ep}	DCE	...

Source.—References 17, 19–21.

Note.—DCE = dynamic contrast-enhanced, Gauss = Gaussian derivative, Hess = Hessian filter, max = maximum enhancement, 2D = two-dimensional, TTP = time to peak, WOR = washout rate.

* Direct imager signal intensities.

prostate sections at MR imaging, T_p the total number of prostate sections at pathologic analysis, and S_p is the section number of the selected pathologic section. This approach is similar to that presented by Hambrook et al (13). Subsequently, the selected MR imaging and pathologic sections were visually assessed for correspondence by a nonclinical medical imaging researcher (G.L., with 4 years of experience with prostate MR imaging) and corrected if deemed necessary. Subsequently, a radiologist with 7 years of experience with prostate MR imaging checked the section correspondences.

After establishing section correspondence, histopathologic and MR imaging sections were coregistered by using an interactive B-spline elastic registration method. This approach has been successfully applied in a number of previous studies in which correspondence between prostate MR images and histopathologic specimens needed to be established (14,15). The registration corresponding points on the boundary

of the prostate at MR imaging and the pathologic specimen were selected by a medical imaging researcher (G.L., with 4 years of experience with prostate MR imaging). The coregistration methodology is shown in Figure 1.

Computer-extracted Features

After coregistration, a number of MR imaging and computer-extracted features were obtained from within the regions that corresponded to the cancer, BPH, PIN, atrophy, and inflammation. To obtain a single feature vector per region of interest (ROI) mapped onto the MR image, the median value of each feature across the voxels within the ROI is calculated. All features are calculated in two dimensions because we register a single prostatectomy slide to the MR image, which results in two-dimensional annotations. A listing of these features and their associated descriptions can be found in Table 2. More detailed descriptions of feature calculation can be found in Appendix E1 (online).

Determining the Best Features to Discriminate Specific Noncancerous Tissue Categories from Cancer

We used sequential forward floating feature selection (SFFS) analysis to identify the best combination of MR imaging parameters to discriminate among benign classes (22). SFFS was used in combination with a linear discriminant classifier, and the performance metric was the AUC. In our setup, we force the SFFS to find the five most discriminative features for each pair-wise classification task (cancer vs BPH, atrophy, inflammation, and PIN). In addition to the pair-wise classification tasks, we also performed a monolithic classification, in which all benign diseases were grouped together as a single noncancerous class.

To analyze the effect of cancer grade, we repeated the SFFS procedure for the intermediate- and high-grade cancer subsets. Intermediate-grade cancer was defined as cancer with a Gleason grade of 3 + 4, and high-grade cancer was defined as any cancer with a major grade 4 or any grade 5 component (13).

A two-fold patient-stratified cross-validation scheme was used in the SFFS analysis, and the procedure was repeated 100 times. Because of the patient stratification, no lesions within a single patient could simultaneously occur in the training and test folds, which helped mitigate bias in our results. We identified the five features that most frequently appeared in the top five selected features.

Last, because of the differences in protocol (both with and without an endorectal coil), we investigated whether the selected features were affected by the presence of a coil. To this end, we performed the monolithic classification twice: once including all data and once excluding the cases in which an endorectal coil.

Classification

To evaluate whether the selected features resulted in improved classification between benign disease and prostate cancer, we compared the histograms of feature values for the top selected

Table 3

Characteristics of Patients and Identified Lesions

Characteristic	Datum
PSA level (mg/mL)*	9.2 (1–76)
Age (y)*	62 (48–70)
Gleason score	
Normal/benign	349
3 + 3	33
3 + 4	55
4 + 3	23
4 + 4	8
4 + 5	9

Note.—Unless otherwise indicated, data are numbers of lesions. *n* = 70 patients. PSA = prostate-specific antigen. PSA ranges were determined in 49 patients; for 21 patients, PSA levels before MR imaging were unknown.

* Data are median, and data in parentheses are range.

feature for each benign class with MatLab R2013a (MathWorks, Natick, Mass). A tenfold patient-stratified cross-validation scheme was used to train a random forest classifier by using the top ranked features identified for each of the pair-wise classification tasks (BPH, PIN, inflammation, and atrophy vs cancer) (23). This experiment was also specifically performed for the subsets of intermediate- and high-grade cancers to assess whether any differences in classification performance could be observed. The performance of the classifiers was evaluated by using the AUC; bootstrapping was used to obtain 95% confidence intervals for AUCs. Classifier training and evaluation were performed with software developed in-house on the basis of OpenCV (24). Finally, we also performed classification experiments with the monolithic setting.

Results

Patient Data

After annotation and coregistration of the prostatectomy slides and multiparametric MR images for all patients, ROIs with 92 PINs, 64 areas of atrophy, 120 areas of inflammation, and 73 BPH were identified, as well as 128

Table 4

Characteristics of Identified Lesions

Disease Category	Both Zones (<i>n</i> = 477)	Peripheral Zone	Central Gland Zone	Diameter (cm)*
Atrophy	64	35	29	0.82 (0.15–1.89)
Inflammation	120	29	91	0.42 (0.17–1.70)
BPH	73	0	73	0.82 (0.25–2.7)
PIN	92	67	25	0.43 (0.10–1.14)
Cancer	128	80	48	0.72 (0.13–1.88)

Note.—Unless otherwise indicated, data are numbers of lesions.

* Data are average diameter, and data in parentheses are the range.

cancer ROIs that comprised 33 cancers with Gleason grade 3 + 3, 55 with Gleason grade 3 + 4, 23 with Gleason grade 4 + 3, eight with Gleason grade 4 + 4, and nine with Gleason grade 4 + 5. Of the cancerous lesions, 48 were located in the central gland (central plus transition zone), and 80 were located in the peripheral zone. The average diameter of the analyzed ROIs was 0.72 cm, with a minimum of 0.13 cm and a maximum of 1.88 cm. These results are also summarized in Tables 3 and 4. Two example results for the MR imaging/pathologic coregistration are illustrated in Figure 1.

Determining the Best Features

The best features for each nonmalignant lesion type are shown in Table 5 and Figure 2. For each class of benign disease, a unique feature was identified as the most important (eg, ADC was identified as the most important feature with which to discriminate PIN from prostate cancer). For atrophy, inflammation, and BPH, *V_c* maps, T2 maps, and images obtained with *b* value of 800 sec/mm² provided the best separation between cancer and benign disease. Specifically, the focality of the appearance of lesions on these images seemed to provide discriminatory value. These most important features are qualitatively presented in Figure 2.

When distinguishing atrophy from cancer, features based on diffusion-weighted and dynamic contrast-enhanced images have higher discriminative power than do those that are

based on T2-weighted images, which indicates that, for this specific subtask, T2-weighted images seem to be less important. For the other types of noncancerous lesions, features from all three modalities are included, indicating that each parameter in multiparametric MR imaging provides additional information to the diagnostic process.

The influence of cancer grade on the selected features is shown in the middle and lower parts of Table 5. When separating only high-grade cancer from the benign classes, ADC was the most important feature in all two-class classification settings (eg, BPH vs high-grade cancer). We can also see that, out of the entire pool of features, a smaller subset of features is used when we only consider high-grade cancer, with a focus mostly on diffusion-weighted imaging and T2-weighted features (ie, texture).

When differentiating only intermediate-grade cancer from benign disease, the subset of selected features is larger (ie, it contains a wider variety of features) than that for the high-grade cancer and includes more dynamic contrast-enhanced features, such as washout. The results of the experiment to assess whether differences in MR imaging protocol (ie, the presence or absence of an endorectal coil) resulted in different selected features are presented in Table 6. ADC, *b* value of 800 sec/mm², Gabor, and maximum enhancement were selected in both situations. The main difference was that, when including cases in which endorectal and nonendorectal

Table 5
Selected Features for Each of the Different Pair-Wise Classification Tasks Considered in This Work

Cancer Grade	PIN		Atrophy		Inflammation		BPH		Monolithic	
	Feature	No.*	Feature	No.*	Feature	No.*	Feature	No.*	Feature	No.*
All grades										
1	ADC	117	Hess (V_p)	117	Hess (T2 map)	127	Hess ($b = 800 \text{ sec/mm}^2$)	105	ADC	179
2	Gauss ($XX, \sigma = 4.1$)	99	Hess ($b = 800 \text{ sec/mm}^2$)	116	ADC	83	Hess (T2 map)	94	Hess ($b = 800 \text{ sec/mm}^2$)	101
3	T2 map	81	K_{trans}	109	Hess (V_p)	69	ADC	86	Gabor ($b = 1, \lambda = 2, \theta = 0.4$)	79
4	Hess (K_{trans})	66	Washout	91	Hess (K_{trans})	63	Hess (K_{ep})	86	Hess (T2 map)	77
5	Gauss ($XX, \sigma = 6$)	65	ADC	87	Washout	52	TTP	59	Hess (max)	75
High grade										
1	ADC	117	ADC	114	ADC	83	ADC	117	ADC	149
2	Hess (TTP)	88	Hess (max)	52	Hess (ADC)	80	Hess (T2 map)	82	Hess (ADC)	106
3	Hess (ADC)	84	Gabor ($b = 1, \lambda = 2, \theta = 0$)	52	Hess (T2 map)	76	Hess ($b = 800 \text{ sec/mm}^2$)	50	Gauss ($XY, \sigma = 4.1$)	50
4	Gabor ($b = 1, \lambda = 4, \theta = 0.4$)	70	K_{trans}	52	Gauss ($XY, \sigma = 6$)	72	Gauss ($XY, \sigma = 4.1$)	50	V_p	49
5	Gauss ($Y, \sigma = 6$)	57	K_{ep}	51	Hess (K_{ep})	71	Gauss ($XY, \sigma = 6$)	49	Hess (TTP)	47
Intermediate grade										
1	Gauss ($XX, \sigma = 4.1$)	128	Gabor ($b = 1, \lambda = 2, \theta = 0$)	105	ADC	84	Hess ($b = 800 \text{ sec/mm}^2$)	145	Hess (T2 map)	121
2	ADC	98	Washout	84	K_{trans}	73	ADC	86	Hess (V_p)	120
3	Gabor ($b = 1, \lambda = 2, \theta = 0$)	86	K_{trans}	80	Hess (V_p)	57	Gabor ($b = 1, \lambda = 2, \theta = 0$)	70	ADC	109
4	Hess (TTP)	67	Hess ($b = 800 \text{ sec/mm}^2$)	77	Gabor ($b = 1, \lambda = 2, \theta = 0.8$)	53	Hess (T2 map)	66	Hess ($b = 800 \text{ sec/mm}^2$)	84
5	Gauss ($XX, \sigma = 6$)	58	TTP	76	Gabor ($b = 1, \lambda = 4, \theta = 0.8$)	51	Gabor ($b = 1, \lambda = 4, \theta = 0$)	62	Hess (K_{trans})	77

Note.— For features that were calculated with different parameters, the parameter values are given in addition to the feature (eg, Gaussian derivatives show the order and scale on which they were calculated). Gabor = Gabor filter; Gauss = Gaussian derivative; Hess = Hessian filter; max = maximum enhancement; TTP = time to peak.
* Data are the number of times a feature was selected during the feature selection phase.

coils were used, the Hessian filter (focality) feature was selected over the original area of signal intensity for the b value of 800 sec/mm^2 and maximum enhancement features.

Classification

The histogram analysis in Figure 3 illustrates that there is less overlap in feature values all benign classes as not grouped as a single category. Figure 3a shows the distribution of b of 800 sec/mm^2 lesion focality values for BPH, cancer, and noncancerous lesions. The graph shows that, if there is a highly focal lesion on the image obtained with b value of 800 sec/mm^2 , the chance of it being BPH is relatively low. Furthermore, Figure 3d shows that the mean ADC value of PIN lesions has less overlap with cancer than with other benign lesions, such as BPH. This characteristic implies that ADC might be a useful feature because it appears to have very good discriminability between PIN and cancer; however, its performance in discriminating BPH from cancer is lower.

Table 7 shows the performance of the pair-wise classifiers (one benign class vs cancer) when using only the selected five features. The average AUC was 0.70 for the pair-wise classification (BPH, PIN, inflammation, and atrophy vs cancer had AUCs of 0.69, 0.73, 0.63, and 0.75, respectively) and 0.62 for the monolithic classifier (cancer vs all noncancerous classes). In particular, discrimination of PIN and atrophy from cancer appears to improve with the use of specific features, with AUCs of 0.73 and 0.75, respectively, compared with 0.62 for the monolithic case.

The results of separately discriminating benign disease from high- and intermediate-grade cancer are also presented in Table 7. Comparing the AUCs of the pair-wise classifiers with the case in which all cancer grades were grouped together, we observed that, on average, the AUCs for discriminating high-grade cancer from the benign class were higher (average AUC = 0.74), whereas the AUCs for discriminating intermediate-grade cancers were lower (average AUC = 0.64).

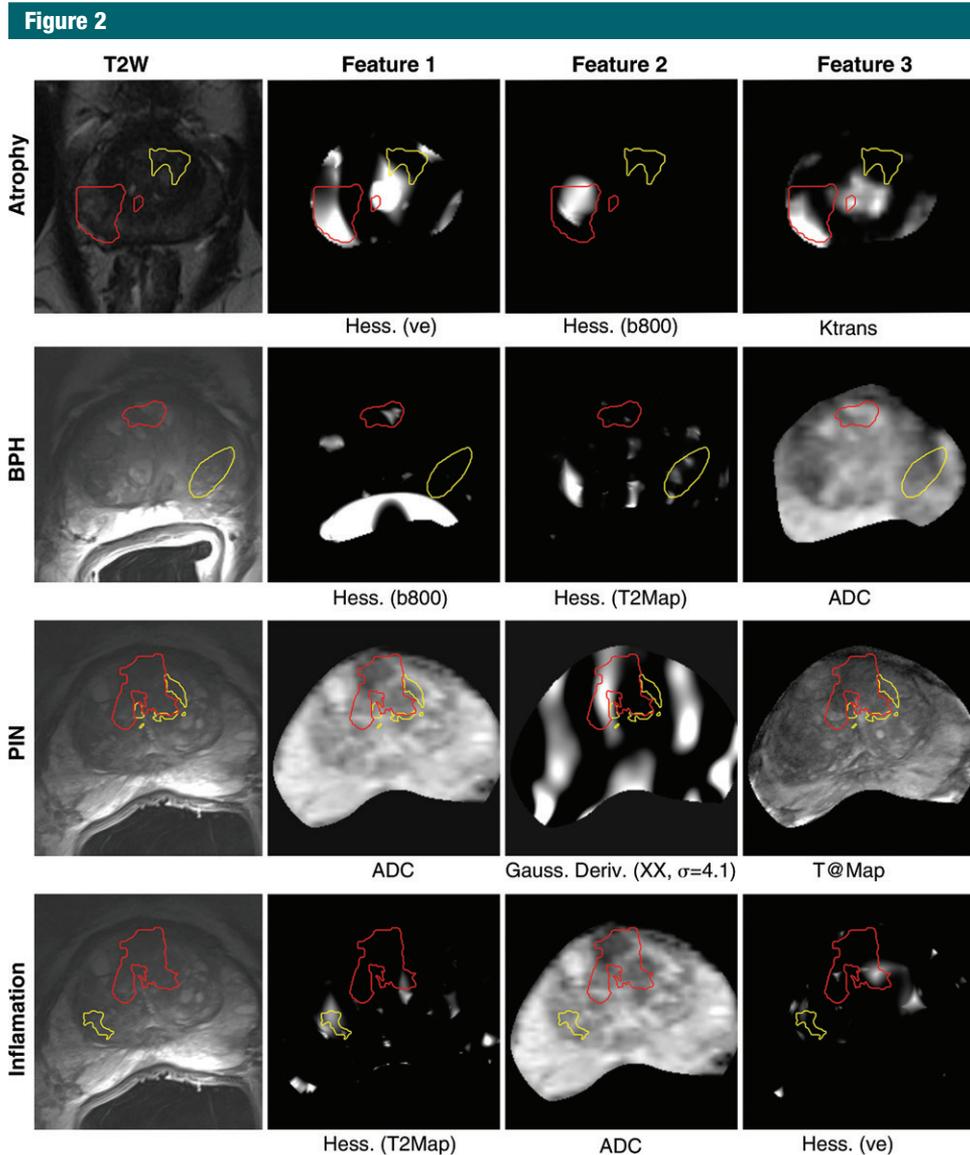


Figure 2: Feature maps of the top three selected features for atrophy, BPH, PIN, and inflammation (cf Table 6) show cancer, with low, intermediate, and high grades grouped together (red line), and the specific benign class (yellow line). The axial T2-weighted image is provided as a reference (left-most column). The selected features provide a good contrast between cancer and the specific benign class.

Discussion

The presence of nonmalignant disease (eg, BPH, inflammation, PIN, and atrophy) is a common cause of false-positive errors in diagnosing prostate cancer at multiparametric prostate MR imaging. For each class of benign disease, we identified a unique set of computer-extracted MR imaging-derived features that could help improve

the differential diagnosis of prostate cancer, such as high-*b*-value images for BPH and focal appearance on dynamic contrast-enhanced images for atrophy.

Several helpful guidelines can be extracted from our results to help radiologists improve their performance in diagnosing prostate lesions. First, this study shows that the appearance of a lesion on a high-*b*-value image may have a higher discriminatory value than ADC

when large BPH nodules are present, a finding that is consistent with previous reports in the literature (25,26). Second, looking at lesion focality (eg, roundness and well-defined edges) on both T2-weighted and dynamic contrast-enhanced images could help discriminate inflammatory processes from prostate cancer, which may be useful in patients who underwent transrectal ultrasonography-guided biopsy near the

Table 6

Selected Features for the Monolithic Classifier Including All Cases and Excluding Cases in Which an Endorectal Coil Was Used

Monolithic Feature	No.
All cases	
ADC	179
Hess ($b = 800 \text{ sec/mm}^2$)	101
Gabor ($b = 1, \lambda = 2, \theta = 0.4$)	79
Hess (T2 map)	77
Hess (max)	75
With nonendorectal coil	
ADC	125
Hess ($b = 800 \text{ sec/mm}^2$)	99
Gabor ($b = 1, \lambda = 2, \theta = 0.4$)	84
Gauss (XY, $\sigma = 6$)	83
Max	64

Note.—Data are the number of times a feature was selected during the feature selection phase. Features remained relatively similar; however, because of increased homogeneity of signal intensity when cases in which an endorectal coil was used were removed, more raw signal intensity features were selected over those that were processed (eg, $b = 800 \text{ sec/mm}^2$ was picked over $b = 800 \text{ sec/mm}^2$ with a Hessian filter). There were in total 34 cases in which an endorectal coil was used. Gabor = Gabor filter, Gauss = Gaussian derivative, Hess = Hessian filter, max = maximum enhancement.

time that MR imaging was performed because of the high risk for acute inflammation.

Furthermore, ADC is typically considered the most useful feature for discriminating prostate cancer from noncancerous lesions, although it is known that ADC values observed in cancer overlap with those of stromal BPH. This characteristic appears to be reflected in the results of the feature selection with the monolithic classifier setting; however, it is only the most important feature for one specific two-class problem, namely, cancer versus PIN. In the cases of BPH and atrophy, it is only the third and fifth most important feature, respectively. This finding appears to support our initial hypothesis that each of the nonmalignant lesions appears to have a distinct set of imaging descriptors that can be used to help characterize them. This finding is further confirmed by the classification results, in which the average and individual AUCs

of the pair-wise classification tasks were higher than the AUC for the monolithic classification. As a whole, these results suggest that a more granular diagnostic process, with an additional focus on specific features (eg, focality and spiculation, similar to the Breast Imaging Reporting and Data System), may be useful compared with the clinically important cancer versus not the clinically important cancer guidelines, which are currently established within PIRADS (3).

Our results indicate that the most diagnostically relevant features depend on cancer grade. For high-grade cancer, a smaller subset of discriminatory features were identified in our experiments compared with the experiments that included all grades of cancers, indicating that fewer features are important; the single most important feature to discriminate high-grade cancer from benign lesions is ADC. This is accompanied by an increased AUC for diagnosing high-grade cancer relative to all cancer grades grouped together. Both of these results indicate that high-grade cancer has its own distinct imaging characteristics (eg, low ADC and a distinctly different texture) compared with all lower cancer grades (even low and intermediate) and is easier to discriminate from noncancerous disease than intermediate-grade cancer, which, on average, has a lower AUC.

Diffusion-weighted imaging seems to be the most important modality to determine cancer aggressiveness, as features derived from ADC and images obtained with b value of 800 sec/mm^2 appear to have increased discriminability as the cancer grade increases. Results reported in the literature on ADC and high- b -value images have also shown that the difference between benign prostate lesions and high-grade cancer is relatively large (8,13). The overlap in ADC between intermediate-grade cancer and noncancerous disease is much larger and definitely requires more and more specific features to allow discrimination.

We acknowledge that our study had a number of limitations. For example, we limited our analysis to a

single prostatectomy slide because three-dimensional reconstruction of an entire prostatectomy specimen and subsequent mapping to MR images is difficult, and often impossible, in current diagnostic practice. Trivedi et al (27) presented a method with a three-dimensional printed mold that could be an avenue for future research.

In addition, to keep the data unbiased, we did not edit or remove lesions on the basis of size or visibility at MR imaging. Some annotated lesions on the prostatectomy specimen were only a couple of voxels large at MR imaging and suffered from partial volume effect, making it difficult to correctly register and characterize them.

Another limitation of the study was the fact that some patients were imaged with an endorectal coil, whereas others were not. We investigated the effect of this difference on the selected features for the monolithic classification. Although the selected features are mostly the same, when the protocol is more homogeneous (ie, only cases in which a nonendorectal coil was used), signal intensities were more often selected as a feature because they are directly comparable between studies. The advantage of the variability in protocol is that the selected features obtained when including all cases are more likely to be protocol-agnostic and, thus, more widely useable (ie, features that are robust to changes in imaging protocol are selected).

In addition, although pathologists were asked to annotate representative areas for each of the benign classes, interobserver variability was not specifically assessed. We do note that annotating the discussed noncancerous disease classes is not very difficult for pathologists. Because of limited data, we did not further subdivide the benign classes into subcategories (eg, acute vs chronic inflammation and stromal vs glandular BPH).

No test to determine statistical significance of the difference between the monolithic and pairwise classification was performed. The monolithic classification contains all noncancerous samples, whereas each pair-wise

Figure 3

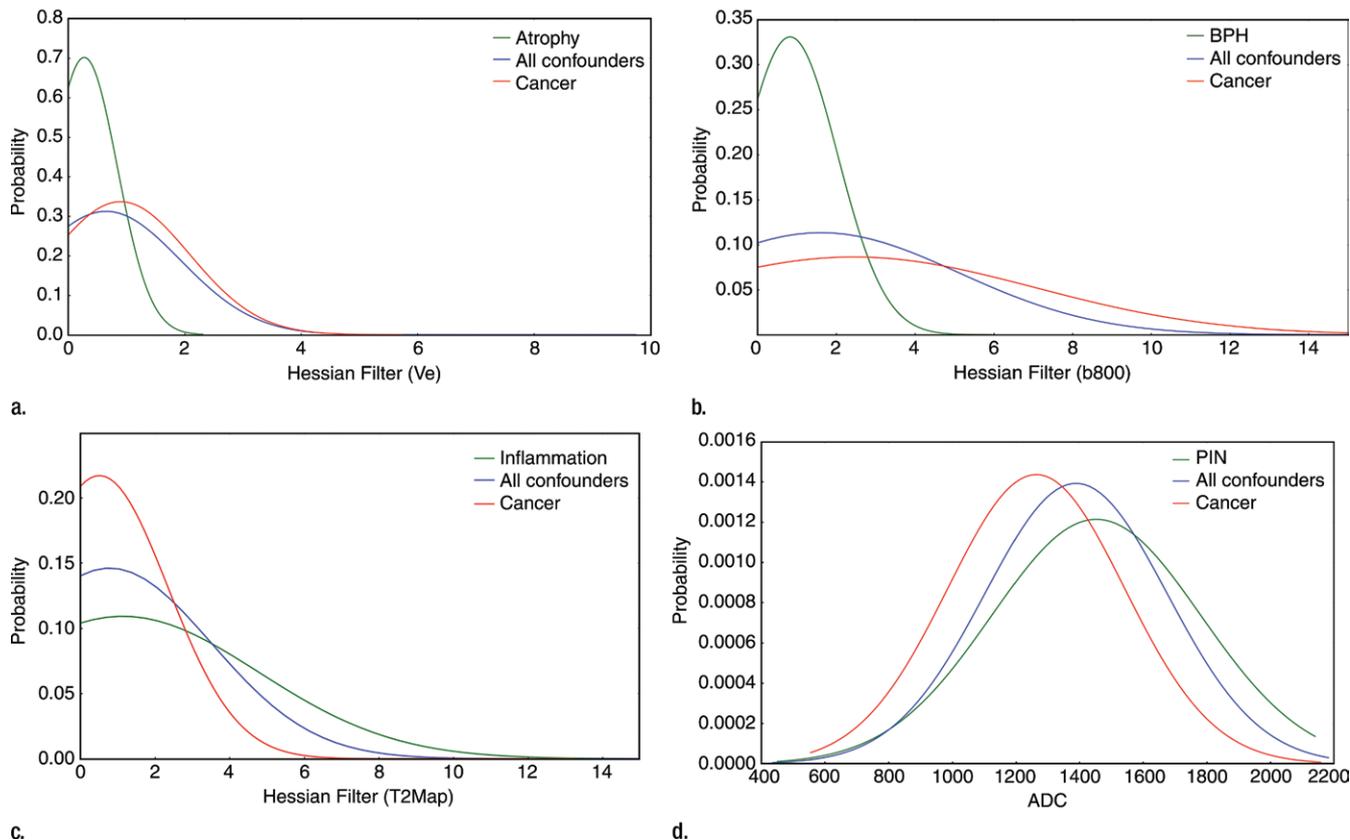


Figure 3: Fitted histograms of the feature value distribution of the top selected feature for each of the classification tasks show cancer, with low, intermediate, and high grades grouped together (red line), all benign classes (blue line), and atrophy (green line in **a**) (green line in **b**), inflammation (green line in **c**), and PIN (green line in **d**). Each specific benign class histogram has less overlap with that of cancer relative to the histogram of all benign disease grouped together, which indicates that this feature allows higher discriminability between cancer and the specific benign class.

Table 7

Classification Performance of a Random Forest Classifier for Each Pair-Wise Classification Task

Cancer Grade	PIN	Atrophy	Inflammation	BPH	Benign*
All grades	0.73 (0.67, 0.80)	0.75 (0.69, 0.82)	0.63 (0.54, 0.70)	0.69 (0.61, 0.76)	0.62 (0.53, 0.70)
High grade	0.73 (0.63, 0.81)	0.77 (0.66, 0.86)	0.77 (0.67, 0.84)	0.69 (0.55, 0.82)	0.64 (0.55, 0.74)
Intermediate grade	0.65 (0.57, 0.72)	0.70 (0.61, 0.79)	0.57 (0.47, 0.66)	0.63 (0.54, 0.73)	0.62 (0.56, 0.69)

Note.—Data are AUCs, and data in parentheses are 95% confidence intervals, which were obtained with bootstrapping.

* Monolithic classification, in which all benign classes are grouped together as a single benign class.

classification only contains the specific noncancerous class, making it impossible to perform a paired *t* test. Because the samples are not independent, a regular *t* test is also not applicable.

Further validation of the results obtained in this study is needed. By using an independent cohort, our results

could be confirmed and significance established. This is an important avenue for future work.

We explored and showed the utility of computerized image and feature analysis in conjunction with multiparametric MR imaging to distinguish between prostate cancer and

noncancerous disease. For each pair of PIN, atrophy, BPH, and inflammation versus cancer, we identified a unique set of features that could help improve the diagnosis of prostate cancer. Our results could be used in two ways: first, they could form the basis for more granular guidelines for prostate MR

imaging interpretation, and second, the results could allow for development of improved computerized decision support systems for diagnosis and characterization of prostate cancer.

Disclosures of Conflicts of Interest: R.E. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: personal fees from Roche Diagnostics/Ventana. Other relationships: disclosed no relevant relationships. N.N.C.S. disclosed no relevant relationships. M.D.F. disclosed no relevant relationships. T.K. disclosed no relevant relationships. C.H-v.d.K. disclosed no relevant relationships. J.O.B. disclosed no relevant relationships. H.J.H. disclosed no relevant relationships. A.M. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: member of the board, consultant for, and stock in Inspirata and patent and royalties from Elucid Bioimaging and Inspirata. Other relationships: disclosed no relevant relationships.

References

1. Tanimoto A, Nakashima J, Kohno H, Shimoto H, Kuribayashi S. Prostate cancer screening: the clinical value of diffusion-weighted imaging and dynamic MR imaging in combination with T2-weighted imaging. *J Magn Reson Imaging* 2007;25(1):146–152.
2. Hoeks CM, Barentsz JO, Hambroek T, et al. Prostate cancer: multiparametric MR imaging for detection, localization, and staging. *Radiology* 2011;261(1):46–66.
3. Barentsz JO, Richenberg J, Clements R, et al. ESUR prostate MR guidelines 2012. *Eur Radiol* 2012;22(4):746–757.
4. Portalez D, Mozer P, Cornud F, et al. Validation of the European Society of Urogenital Radiology scoring system for prostate cancer diagnosis on multiparametric magnetic resonance imaging in a cohort of repeat biopsy patients. *Eur Urol* 2012;62(6):986–996.
5. Rosenkrantz AB, Kim S, Lim RP, et al. Prostate cancer localization using multiparametric MR imaging: comparison of Prostate Imaging Reporting and Data System (PI-RADS) and Likert scales. *Radiology* 2013;269(2):482–492.
6. Thompson J, Stricker P, Brenner P, et al. Magnetic resonance imaging detects significant prostate cancer and could be used to reduce unnecessary biopsies: initial results from a prospective trial. *J Urol* 2013;189(4 Suppl):e910–e911.
7. Rosenkrantz AB, Taneja SS. Radiologist, be aware: ten pitfalls that confound the interpretation of multiparametric prostate MRI. *AJR Am J Roentgenol* 2014;202(1):109–120.
8. Oto A, Kayhan A, Jiang Y, et al. Prostate cancer: differentiation of central gland cancer from benign prostatic hyperplasia by using diffusion-weighted and dynamic contrast-enhanced MR imaging. *Radiology* 2010;257(3):715–723.
9. Liu X, Zhou L, Peng W, Wang C, Wang H. Differentiation of central gland prostate cancer from benign prostatic hyperplasia using monoexponential and biexponential diffusion-weighted imaging. *Magn Reson Imaging* 2013;31(8):1318–1324.
10. Ginsburg SB, Viswanath SE, Bloch BN, et al. Novel PCA-VIP scheme for ranking MRI protocols and identifying computer-extracted MRI measurements associated with central gland and peripheral zone prostate tumors. *J Magn Reson Imaging* 2015;41(5):1383–1393.
11. Agner SC, Rosen MA, Englander S, et al. Computerized image analysis for identifying triple-negative breast cancers and differentiating them from other molecular subtypes of breast cancer on dynamic contrast-enhanced MR images: a feasibility study. *Radiology* 2014;272(1):91–99.
12. Xiao G, Bloch BN, Chappelw J, et al. Determining histology-MRI slice correspondences for defining MRI-based disease signatures of prostate cancer. *Comput Med Imaging Graph* 2011;35(7-8):568–578.
13. Hambroek T, Somford DM, Huisman HJ, et al. Relationship between apparent diffusion coefficients at 3.0-T MR imaging and Gleason grade in peripheral zone prostate cancer. *Radiology* 2011;259(2):453–461.
14. Tiwari P, Kurhanewicz J, Madabhushi A. Multi-kernel graph embedding for detection, Gleason grading of prostate cancer via MRI/MRS. *Med Image Anal* 2013;17(2):219–235.
15. Viswanath SE, Bloch NB, Chappelw JC, et al. Central gland and peripheral zone prostate tumors have significantly different quantitative imaging signatures on 3 Tesla endorectal, in vivo T2-weighted MR imaging. *J Magn Reson Imaging* 2012;36(1):213–224.
16. Madabhushi A, Udupa JK. Interplay between intensity standardization and inhomogeneity correction in MR image processing. *IEEE Trans Med Imaging* 2005;24(5):561–576.
17. Litjens G, Debats O, van de Ven W, Karssemeijer N, Huisman H. A pattern recognition approach to zonal segmentation of the prostate on MRI. *Med Image Comput Assist Interv* 2012;15(Pt 2):413–420.
18. Litjens GJ, Huisman HJ, Elliott RM, et al. Quantitative identification of magnetic resonance imaging features of prostate cancer response following laser ablation and radical prostatectomy. *J Med Imaging* 2014;1(3):035001.
19. Li Q, Sone S, Doi K. Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans. *Med Phys* 2003;30(8):2040–2051.
20. Huisman HJ, Engelbrecht MR, Barentsz JO. Accurate estimation of pharmacokinetic contrast-enhanced dynamic MRI parameters of the prostate. *J Magn Reson Imaging* 2001;13(4):607–614.
21. Vos PC, Hambroek T, Barentsz JO, Huisman HJ. Computer-assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI. *Phys Med Biol* 2010;55(6):1719–1734.
22. Somol P, Pudil P, Novovicova J, Paclik P. Adaptive floating search methods in feature selection. *Pattern Recognit Lett* 1999;20(11-13):1157–1163.
23. Breiman L. Random forests. *Mach Learn* 2001;45(1):5–32.
24. Bradski G. The Open CV library. *Doctor Dobbs J* 2000;25(11):120–126.
25. Kobus T, Vos PC, Hambroek T, et al. Prostate cancer aggressiveness: in vivo assessment of MR spectroscopy and diffusion-weighted imaging at 3 T. *Radiology* 2012;265(2):457–467.
26. Hoeks CM, Hambroek T, Yakar D, et al. Transition zone prostate cancer: detection and localization with 3-T multiparametric MR imaging. *Radiology* 2013;266(1):207–217.
27. Trivedi H, Turkbey B, Rastinehad AR, et al. Use of patient-specific MRI-based prostate mold for validation of multiparametric MRI in localization of prostate cancer. *Urology* 2012;79(1):233–239.