Computer-Aided Detection of Prostate Cancer in MRI

Geert Litjens*, Oscar Debats, Jelle Barentsz, Nico Karssemeijer, and Henkjan Huisman

Abstract—**Prostate cancer is one of the major causes of cancer death for men in the western world. Magnetic resonance imaging (MRI) is being increasingly used as a modality to detect prostate cancer. Therefore, computer-aided detection of prostate cancer in MRI images has become an active area of research. In this paper we investigate a fully automated computer-aided detection system which consists of two stages. In the first stage, we detect initial candidates using multi-atlas-based prostate segmentation, voxel feature extraction, classification and local maxima detection. The second stage segments the candidate regions and using classification we obtain cancer likelihoods for each candidate. Features represent pharmacokinetic behavior, symmetry and appearance, among others. The system is evaluated on a large consecutive cohort of 347 patients with MR-guided biopsy as the reference standard. This set contained 165 patients with cancer and 182 patients without prostate cancer. Performance evaluation is based on lesion-based free-response receiver operating characteristic curve and patient-based receiver operating characteristic analysis. The system is also compared to the prospective clinical performance of radiologists. Results show a sensitivity of 0.42, 0.75, and 0.89 at 0.1, 1, and 10 false positives per normal case. In clinical workflow the system could potentially be used to improve the sensitivity of the radiologist. At the high specificity reading setting, which is typical in screening situations, the system does not perform significantly different from the radiologist and could be used as an independent second reader instead of a second radiologist. Furthermore, the system has potential in a first-reader setting.**

Index Terms—**Computer-aided detection, image analysis, machine learning, magnetic resonance imaging, prostate cancer.**

I. INTRODUCTION

PROSTATE cancer is one of the major causes of cancer death for men in the Western world [1]. Due to the increased ageing of the general population the incidence of prostate cancer is steadily rising [1]. Current clinical practice for the diagnosis of prostate cancer is to perform a transrectal ultrasound (TRUS) biopsy, which usually is performed due to a positive prostate specific antigen (PSA) blood test. A large screening trial using PSA and TRUS has shown that is possible to reduce prostate cancer mortality by 20%–30% [2]. However, these studies have also shown that PSA testing in combination with TRUS biopsies has a relatively low specificity. Additionally, cancers are often undergraded in TRUS biopsies [3]. These problems lead to overdiagnosis and overtreatment of patients and are prohibiting screening for prostate cancer.

MRI is increasingly used to diagnose prostate cancer as it has improved sensitivity and specificity over PSA and TRUS [4]. Currently, magnetic resonance imaging (MRI) is most often used as a second-line modality after repeat negative TRUS biopsies. One of the reasons MRI has not yet progressed to a first line modality for prostate cancer diagnosis is that it requires substantial expertise from the radiologist to read prostate MRI and such expertise is not widely available. Additionally, due to the large number of 3-D images, reading prostate MR is quite time consuming.

Automated computer-aided detection and diagnosis (CAD) of prostate cancer could help reduce both of these problems and open the door to prostate cancer screening using MRI. In the past several other areas have seen successful CAD applications, such as mammography [5] and CT colonography [6]. In the last decade several researchers have investigated the use of these techniques for prostate MRI. Therefore, computer-aided detection and diagnosis of prostate cancer is becoming an active field of research [7]–[10].

Chan *et al.* [7] were the first to implement a multi-parametric CAD system for the diagnosis of prostate cancer. In their approach they used line-scan diffusion, T2 and T2-weighted images in combination with an SVM classifier to identify predefined areas of the peripheral zone of the prostate for the presence of prostate cancer. Langer *et al.* [8] included dynamic-contrast enhanced images and pharmacokinetic parameter maps as extra features to a CAD system for prostate peripheral zone cancer. They evaluated their system in predefined regions of interest, but on a per voxel basis. Tiwari *et al.* [9] investigated the use of magnetic resonance spectroscopy in combination with T2-weighted imaging to identify the voxels that are affected by prostate cancer. They also introduced the use of wavelet embedding to map MRS and T2-W texture features into a common space. This work was further expanded and evaluated in [11]. Niaf *et al.* [12] presented the use of computer-aided diagnosis in the peripheral zone of the prostate (similar to Vos *et al.* [13]. They confirmed the results in discriminating prostate cancer from normal regions (area under the ROC curve (AUC) $=0.89$) and discriminating prostate cancer from suspicious benign regions (AUC of 0.82). This is a limited CAD method as it is constrained to predefined regions of interest in only the peripheral zone. Firjani *et al.* [14] investigated the use of computer-aided detection in single-parameter MRI using diffusion weighted imaging (DWI) imaging with transrectal ultrasound guided biopsies as ground truth. They included registration of different b-values to obtain a less
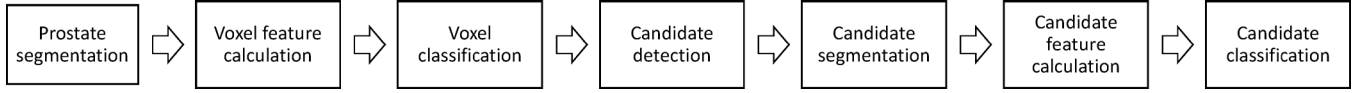
Fig. 1. Flowchart showing the different steps of the computer-aided detection system.

motion sensitive apparent diffusion coefficient map. Lastly, Vos *et al.* [10] recently implemented a two-stage computer-aided detection system for prostate cancer using an initial blob detection approach combined with a candidate segmentation and classification using statistical region features.

In this paper, we investigate a fully automated computer-aided detection system including a novel combination of segmentation, voxel classification, candidate extraction and candidate classification, which expands on the work published in [15]. Other novel aspects include a voxel classification stage in combination with a candidate classification stage and inclusion of symmetry, local contrast and anatomical features like peripheral zone likelihood. Feature design was based on the standardized guidelines for reading prostate MR, PI-RADS, and include texture, pharmacokinetic, shape and anatomy, among others [16]. Furthermore, to the best of the authors' knowledge this is the first prostate MRI CAD system that is evaluated on a per-patient basis and compared to the prospective performance of radiologists. The system was validated on a large cohort of 347 patients using per-region FROC and per-patient ROC to show the value of a two stage approach incorporating both voxel and candidate classification.

II. MATERIALS AND METHODS

A. MRI Data

In our hospital, we collected a total of 165 consecutive studies with prostate cancer (187 lesions) and 183 cases without prostate cancer on which to evaluate our CAD-system for a total of 348 studies of 347 patients.

Each MR study was read and reported by or under the supervision of an expert radiologist (Barentsz), with more than 20 years of experience in prostate MR. The radiologist indicated areas of suspicion with a score per modality using a point marker. If an area was considered likely for cancer a biopsy was performed. All biopsies were performed under MR-guidance and confirmation scans of the biopsy needle *in situ* were made to confirm accurate localization. Biopsy specimen were subsequently graded by a pathologist and these results were used as ground truth.

All studies included T2-weighted (T2W), proton density-weighted (PD-W), dynamic contrast enhanced (DCE), and diffusion-weighted (DW) imaging. It is currently established clinical consensus that prostate cancer should be diagnosed by T2-weighted imaging with at least two functional modalities (from DWI, DCE, and spectroscopic imaging) [4], [16]. The images were acquired on two different types of Siemens 3T MR scanners, the MAGNETOM Trio and Skyra. T2-weighted images were acquired using a turbo spin echo sequence and had a resolution of around 0.5 mm in plane and a slice thickness of 3.6 mm. The DCE time series was acquired using a 3-D turbo flash gradient echo sequence with a resolution of around 1.5

mm in-plane, a slice thickness of 4 mm and a temporal resolution of 3.5 s. The proton density weighted image was acquired prior to the DCE time series using the same sequence with different echo and repetition times and a different flip angle. Finally, the DWI series were acquired with a single-shot echo planar imaging sequence with a resolution of 2 mm in-plane and 3.6 mm slice thickness and with diffusion-encoding gradients in three directions. Three b-values were acquired (50, 400, and 800), and subsequently, the ADC map was calculated by the scanner software. All images were acquired without an endorectal coil, as per the PI-RADS guidelines for acquisition of prostate MRI [16]. Although an endorectal coil would allow for further improved resolution of the images, the added value is considered negligible compared to added patient comfort when only using a pelvic phased array coil. The transversal T2W image, the PD-W image, the entire DCE time series and from the DWI the apparent diffusion coefficient map and the b800-image were used in this study. These images were chosen because they are explicitly incorporated in the PI-RADS standard, except the PD-W image, which was added purely for feature calculation purposes.

B. Overview of the CAD Pipeline

The pipeline of the CAD system is visualized schematically in Fig. 1 and follows a two stage approach.

The first (detection) stage consists of segmentation of the prostate on the transversal T2-weighted image, extraction of voxel features from the image volumes, classification of the voxels and candidate selection. The second (diagnosis) stage consists of candidate segmentation, candidate feature extraction and candidate classification. Each of the steps will be described in more detail in the corresponding subsections.

C. Segmentation

The segmentation of the prostate is required to reduce the complexity of the detection task for the classifiers in the later stages. In our system we use an atlas based segmentation approach similar to the one presented in [17], using the atlas selection mechanism presented in [18], named SIMPLE (selective and iterative method for performance level estimation). Let the image to be segmented be denoted as $I(x)$, where x is a spatial location within the image. A labeled image, $L(x)$ has to be determined. The following steps are similar in most multi-atlas based systems. A set of N manually labeled images A is nonrigidly registered to the unknown case $I(x)$. The ith atlas in this set A is denoted as $A_i(x)$, where each $A_i(x)$ consists of the atlas image and the label image: $A_i(x) = \{I_i(x), L_i(x)\}$. After registration of the atlas image $I_i(x)$ the obtained transformation $T_i(x)$ is applied to the label image of the atlas, $L_i(x)$.

An important part of the registration procedure is the similarity metric that is used. In this approach we used the localized mutual information metric [19]. After atlas registration we
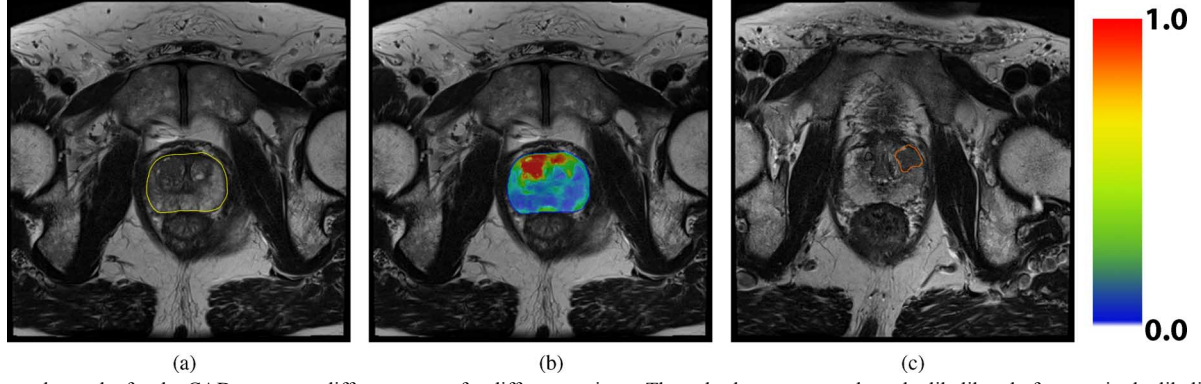
Fig. 2. Example results for the CAD system at different stages for different patients. The color bar corresponds to the likelihood of cancer in the likelihood map and the region segmentation. (a) Prostate segmentation result. (b) Likelihood map after voxel classification. (c) Region classification result.

TABLE I
OVERVIEW OF VOXEL FEATURES USED IN THE CAD SYSTEM

Name	Type	Description
T2W	Intensity	T2-weighted voxel grey value, related to voxel T2
ADC	Intensity	Apparent diffusion coefficient, measure for cellular density
b800	Intensity	High b-value image, areas with low diffusivity appear bright
T2-map	Intensity	Calculated T2-map based on proton density and transversal T2W image [20]
x-pos	Anatomical	Relative cumulative position within the prostate mask between 0 and 1 in the x-direction
y-pos	Anatomical	Relative cumulative position within the prostate mask between 0 and 1 in the y-direction
z-pos	Anatomical	Relative cumulative position within the prostate mask between 0 and 1 in the z-direction
Distance	Anatomical	Relative distance to the prostate boundary between 0 and 1
PZ likelihood	Anatomical	Likelihood of being a peripheral zone voxel between 0 and 1 [20]
K_{trans}	Pharmacokinetic	Pharmacokinetic parameter, related to vessel permeability
k^{ep}	Pharmacokinetic	Pharmacokinetic parameter, related to permeability and extracellular volume
tau	Pharmacokinetic	Dynamic parameter, related to the time-to-peak of contrast agent concentration
LateWash	Pharmacokinetic	Dynamic parameter, related to the washout of contrast agent
Gaussian texture bank	Texture	Calculate multi-scale Gaussian derivatives on the T2W image
ADC Blob	Blobness	Multi-scale blob detection using the Li blobness filter [21] on ADC map
K_{trans} Blob	Blobness	Multi-scale blob detection using the Li blobness filter [21] on K_{trans} map
LateWash Blob	Blobness	Multi-scale blob detection using the Li blobness filter [21] on LateWash map
tau Blob	Blobness	Multi-scale blob detection using the Li blobness filter [21] on tau map

have a transformed label image $L_i \circ T_i(x)$ for each atlas. We use the SIMPLE method, presented by Langerak *et al.* [18], to combine these label images into one final segmentation. Example results are shown in Fig. 2. This algorithm competed in the prostate MR image segmentation (PROMISE12) challenge (http://promise12.grand-challenge.org), where it obtained a ninth place out of 12. Overall segmentation results were still reasonably good, with a median Dice's coefficient of 0.83.

D. Voxel Features

After prostate segmentation we calculated voxel features from the image volumes. The types of features can be categorized in intensity, pharmacokinetic, texture, blobness, and anatomical features. A complete overview of the voxel features is given in Table I, implementation details are given in the corresponding subsections.

1) Intensity: One of the major issues in image analysis for MRI is the absence of a standardized signal intensity, like Hounsfield units in CT. This usually means that an algorithm will give different results as scanners, sequences or even sequence parameters are changed. To mitigate this issue we developed several algorithms to standardize signal intensity in the different MR modalities. First, for the T2-weighted imaging a T2-estimate map is generated by using the MR signal equation, the proton density image and a reference tissue.

This process was automated and is explained in more detail in [20]. This map was added as a voxel feature in addition to the original transversal T2-weighted image. Second, the MR scanner software automatically calculates the apparent diffusion coefficient map from the diffusion-weighted images, by fitting a mono-exponential function to the signal decay across the different b-values. Furthermore, studies have shown that the highest b-value image has additional diagnostic value, therefore, the b800 image was also added as a feature.

2) Anatomical: For the anatomical features we used the relative distance to the prostate boundary

$$\text{B}(\text{x}) = \min_{y \in P_b} d(x, y) \tag{1}$$

$$\text{RD}(x) = \frac{B(x)}{\max_{x \in P} B(x)} \tag{2}$$

where x is the position of a voxel, d is the Euclidean distance operator, P is the set of prostate voxels, and P_b the set of prostate boundary voxels.

Additionally, we also use relative position features in $x, y,$ and z directions. The relative position features are defined as

$$\text{RP}(x_i) = \frac{x_i - \min_{x \in P} x_i}{\max_{x \in P} x_i - \min_{x \in P} x_i} \tag{3}$$

where x is the position of a voxel and i is the image axis, either x, y or z, and P the set of prostate voxels. Both the relative distance and the relative position features are calculated with respect to the prostate mask obtained through the multi-atlas method. Finally, we also implemented a peripheral zone probability feature, which gives a likelihood per voxel that it belongs to the peripheral zone. This feature uses a pattern recognition framework incorporating intensity, texture and anatomical features. This results in a likelihood for each voxel within a prostate mask of belonging to either the peripheral zone or the central gland. More implementation details of this filter can be found in [20]. This feature is important because we know from clinical practice that prostate cancer appearance can differ substantially between the peripheral zone and the central gland [16].

3) Pharmacokinetic: In the clinic it is common practice to use the DCE time curve to diagnose prostate cancer [16]. The approach used is described as the curve type method, were the radiologist looks at the curve and assesses two characteristics based on the first, last, and peak enhancement time points [16]. These characteristics are whether there is fast initial enhancement and if there is persistent enhancement, an enhancement plateau or wash-out. Slow initial, persistent enhancement (curve type I) or slow initial, constant enhancement (curve type II) are associated with normal and benign findings whereas fast initial enhancement combined with washout are indicative for malignancy (curve type III).

The traditional analysis is incorporated in our CAD system by using a curve fitting-technique to fit, per voxel, a bi-exponential curve to the time data, as presented in [22]. Of these curve parameters we incorporated the parameter tau (which corresponds to time-to-peak of the enhancement curve) and the parameter LateWash (which corresponds to the slope of the last part of the curve). There are two major problems with only using this type of analysis. First, the assessment of the curve is scanner and patient dependent, e.g., different protocols or patient anxiety (which increases blood flow). Second, not all information present in the curve is used. To counter these disadvantages pharmacokinetic modeling of the contrast agent concentrations has been proposed and applied in breast, brain, and prostate MRI [13], [23]. We implemented the standard Tofts pharmacokinetic model [24] including an automated reference tissue method to estimate the arterial input function, as proposed in [13], [22]. This model provides us with three parameter maps for the DCE time series. The parameters represent the permeability of the micro-vasculature, K^{trans}, the fraction of extracellular, extravascular space, V_e and the quotient of the two, k_{ep}. Due to fast and sloppy vessel construction and tightly packed cells in a cancerous region it is expected that K^{trans} and V_e will differ between cancerous and normal/benign tissue.

4) Texture: Most cancers show textural distortions in T2-weighted images [16], [25]. To capture these characteristics in features we use a Gaussian texture bank. For the Gaussian feature bank we used five different scales, from 2 to 8 mm exponentially and derivatives up to the second order. This scale range was selected to encompass the typical size ranges of lesions in prostate MRI (between 5 and 20 mm in diameter) [26]–[28]. Due to the large slice thickness the features were calculated on a slice-by-slice basis. This results in a total of 30 Gaussian texture features.

5) Blobness: Prostate cancer tends to appear as a focal, blob-like lesion in diffusion-weighted and dynamic contrast enhanced MRI. This characteristic has been previously used to detect prostate cancer [10]. There are many different blobness measures, we chose to incorporate the blobness-filter presented by Li *et al.* [21] because this filter incorporates both a shape term and a blobness strength term. The blobness feature was calculated with scales ranging from 2 to 8 mm, with five different, exponentially increasing scales. Again, this range encompasses the size of most lesions encountered in prostate MRI [26]–[28]. The maximum (bright blobs) or minimum (dark blobs) value of the blobness output across scales at each voxel was used as the final blobness measure. Blobness was calculated on the ADC, tau and LateWash images (dark blobs), and on the Ktrans and Kep images (bright blobs).

E. Voxel Classification

After feature calculation a voxel classification is performed, which results in a likelihood between 0 and 1 per voxel, 0 indicating no suspicion of prostate cancer and 1 indicating very high suspicion of prostate cancer. In this step, we experimented with three different classifiers, a linear discriminant classifier, a GentleBoost classifier [29] (with regression stumps as weak learners) and a RandomForest-classifier with regression trees [30]. Both the GentleBoost and the RandomForest classifiers are very robust to overtraining [29], [30], thus parameter optimization is usually not needed [9]. Therefore, both the RandomForest and the GentleBoost classifier were left at the default settings. For the RandomForest the default settings are that a minimum of 0.1% of all samples in the dataset is required to split a tree node, the square root of the number of features is used as the number of active variables at each node and the maximum tree depth was equal to the number of features. For both classifiers the number of weak learners has to be set, where, as explained in [29] and [30], adding more weak learners does not result in over-fitting, but produces a limiting value of the generalization error. We did a small pilot experiment using two fold cross-validation to roughly determine the amount of weak learners needed to achieve the minimal generalization error. This resulted in around 100 regression stumps for the GentleBoost classifier and 300 trees for the RandomForest classifier.

We compared the performance of the different classifiers using ROC-analysis. The output of the classifier with the highest area under the ROC curve was used for further analysis. An example of a obtained likelihood map is shown in Fig. 2.

F. Candidate Selection and Segmentation

After voxel classification a likelihood map is obtained, indicating per voxel the likelihood that it contains cancer. On this likelihood-map we perform local maxima detection using a spherical window with a diameter of 10 mm, which is about the average lesion size in prostate MRI [26]–[28]. After initial local maxima detection the local maxima which are less than 10 mm apart are merged. This merging step leaves only the local maximum with the highest probability within the 10 mm range. This is iterated until no more merging occurs.

For each of the local maxima obtained in the previous step, a region segmentation will be performed. The SmartOpening-algorithm, which has had successful applications in both nodule segmentation in CT and cancer segmentation in breast MRI, was used [31], [32]. The segmentation was performed on the likelihood-map itself instead of one of the original clinical images. The main reason for this is that lesions can show slight deviations in size and even position between the different diagnostic images. The likelihood map is essentially a combination of all original images through a classification step and thus should give a good approximation of the lesion extent across all diagnostic images. After initial segmentation, regions which overlapped for more than 50% were merged.

G. Candidate Features

After candidate segmentation and merging new candidate features can be calculated given the original feature images and the candidate segmentation. These can be categorized as: statistical (voxel feature statistics), local contrast, symmetry, and shape features.

1) Statistical: Statistical candidate features are calculated within the candidate segmentation. Statistics include mean, standard deviation and histogram percentiles. On all the initial voxel features we calculate the mean and standard deviation of the feature values within the candidate segmentation. Additionally, for the ADC and pharmacokinetic features we calculate either the 25th or 75th percentiles, depending on whether low or high values are indicative of malignancy. The percentiles are calculated because 60% of all tumors are heterogeneous [27], with a more aggressive hot spot within the tumor that for example has lower ADC values. In addition, we also calculated the mean, standard deviation and 75th percentile of the voxel likelihood.

2) Local Contrast: In previous work we have shown that relating tumor feature values to those of surrounding normal tissue can lead to improved characterization of tumor aggressiveness [33]. We incorporate this knowledge into our CAD system by using local contrast candidate features. The local contrast feature is calculated by dilating the original segmentation and then subtracting the original to obtain a rim of tissue outside the candidate. The local contrast is then obtained by taking the quotient of the average candidate and the average rim intensities. We use a 2-D kernel with a size of 3 mm for dilation. The local contrast feature is calculated on the b800, ADC, Ktrans, Kep, tau, and LateWash voxel feature maps. Additionally, it was also calculated on the voxel likelihood map.

3) Symmetry: A normal prostate has a distinct symmetric appearance in the transversal plane. Radiologist have reported that symmetry in prostate MRI can be important to detect prostate cancer [16]. The CAD system incorporates this knowledge by including a symmetry feature. We take the relative position of a candidate along the x-axis in the transversal plane and mirror it to the other side of the prostate (e.g., if the relative position is 0.25 we map the mirrored candidate segmentation to a relative position of 0.75). Then we calculate the mean intensity value for both the mirrored and the original candidate segmentation and take the quotient. The result is used as the symmetry feature. We calculate this symmetry feature on the b800, ADC, Ktrans, Kep, tau, LateWash voxel feature maps and the voxel likelihood map.

4) Shape: The last candidate feature set are the shape features. Prostate lesions tend to be somewhat spherical and compact. During initial stages of development, most false positives we encountered were due to small segmentation errors, large nonspherical areas of low ADC due to extensive benign prostatic hyperplasia and small artifacts caused by the scanner. By incorporating shape features like volume, sphericity and compactness the classifier can easily remove these false positives from the data. The sphericity is calculated as the ratio of the volume of a sphere having the same diameter as the maximum bounding box length of the candidate segmentation and the total volume of the candidate segmentation. The compactness is calculated as the candidate segmentation volume divided by the volume of the bounding box of the candidate segmentation.

H. Candidate Classification

After candidate feature extraction the final classification is performed. Three different classifiers were tried to obtain the best possible performance, a linear discriminant classifier, a GentleBoost classifier (with regression stumps as weak learners) and a RandomForest classifier with regression trees. The settings we used at this stage were the same as in the voxel classification stage. After classification we obtain a likelihood between 0 and 1 per candidate, 0 indicating no prostate cancer and 1 indicating definite prostate cancer. Examples of a final candidate result can be seen in Fig. 2

I. Relative Feature and MR Sequence Importance in Voxel and Candidate Classification

To establish the importance of individual features and MR sequences to the overall classification results, we performed two experiments at both the voxel and candidate levels. First, using the selected classifiers, we established the classification performance of each feature individually based on area under the ROC curve (using leave-one-patient-out crossvalidation). Subsequently, we repeated this experiment on a per-sequence basis, i.e., only include features calculated using one MR sequence, for example only using T2-texture features or only using DWI features.

J. Validation

1) Training Data: For the voxel classification stage voxels in a 10 mm area around the radiologist annotation were extracted as prostate cancer samples. This area was truncated by the prostate mask, to ensure no voxels outside the prostate were included in the training set. Furthermore, we only selected voxels which had specific feature characteristics: the ADC value had to be below the median of the area and the pharmacokinetic features had to be above the median of the area. We know from clinical experience and literature that these are usually good characteristics of prostate cancer and reduces the chance of sampling normal voxels into the malignant class. For the normal class we randomly sample within the prostate mask of normal patients. The resultant voxel dataset is used to train the voxel classifiers.

In the candidate classification stage we extract candidate features from the initially detected true positives and the false positives in normal patients after initial classification. The definition of true and false positives is given in the next section.

2) FROC Analysis: The detection performance of the CAD system is evaluated using free-response receiver operating characteristic curve (FROC) analysis. FROC analysis provides the number of false positives per normal patient for a given sensitivity (i.e., the percentage of cancer detected). The occurrence of false positives in normal patients is one of the most relevant problems in prostate cancer diagnosis on MRI as each false positive in a normal patient has the potential to lead to an unnecessary biopsy, and thus patient morbidity and healthcare cost. As such, the number of false positives should be as low as possible at reasonable sensitivity. FROC analysis can be used both after the initial and final stage, which also allows us to assess the performance gained by the second stage of the CAD system. For evaluation of the first stage the criterion for a true positive is that a local maximum should be within 10 mm of the marker annotated by the radiologist. 10 mm corresponds to the average lesions size [26]–[28]. For the final classification a true positive is defined as a candidate segmentation which has a center of gravity within 10 mm of the marker. Each candidate segmentation is only allowed to correspond to one annotation. This rule is chosen to make sure the system does not have a bias toward large segmentations, i.e., a candidate segmentation covering the entire prostate would cover all lesions, but would generally not result in an accurate localization. We evaluated the system both for the detection of all tumors and the detection of high-grade tumors (first or secondary Gleason component >3). In the second setting a hit on a low-grade tumor is not considered a false positive, the reasoning for this is that in principle low-grade prostate cancer will not require treatment, but it is not detrimental for the patient to detect it.

3) ROC Analysis and Comparison to the Radiologists: In addition to FROC analysis we also performed patient based receiver operating characteristic (ROC) analysis both after the initial voxel classification and after the candidate classification. This is relevant evaluation in a screening setting, where the first thing a clinician wants to know is whether a patient has cancer or not (i.e., the localization aspect captured by the FROC analysis in the previous section is of secondary importance). A CAD system could play a role here by separating out the easy from the difficult to diagnose patients, which could improve the efficiency of the radiologist. In each patient, the voxel (voxel stage) or candidate (candidate stage) with the highest likelihood is used as the patient score, both for patients with prostate cancer and normal patients. In this setup the CAD system can stratify patient as requiring a biopsy or not requiring a biopsy.

Additionally, we compare the system to the overall radiologist performance on this data set. In total 10 radiologists read cases in our patient cohort, each case was read prospectively by one radiologist. Therefore, we can compare the system performance to the actual prospective clinical performance in our hospital.

III. RESULTS

1) Classifier Comparison and Selection: Bootstrapping and ROC analysis were used to compare classifiers for both CAD stages: the voxel classification stage and the candidate classification stage. For both stages we performed a leave-one-patient-out cross-validation on the training data. Results are shown
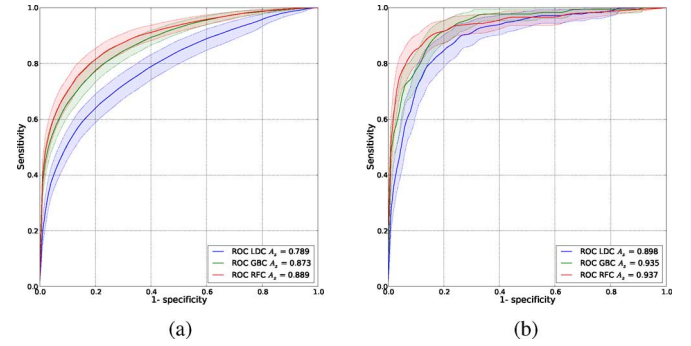

Fig. 3. Classifier comparison using leave-one-patient-out ROC analysis at the voxel and candidate levels. 95% confidence intervals estimated using bootstrapping are shown as transparent areas around the mean curves. Input for the comparison at the voxel level are voxel in a cancerous areas and voxels in the prostate of normal patients. Input for the comparisson at the candidate level is true and false positives after initial voxel classification. (a) Voxel level. (b) Candidate level.
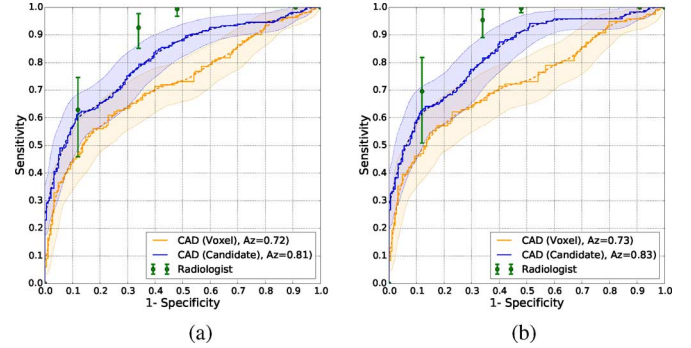

Fig. 4. ROC analysis on a per-patient level, comparison of the CAD system after the voxel stage and after the candidate stage to the prospective radiologist performance. The raw ROC curve is shown as the solid line and the mean bootstrapped curve as a dashed line. The 95% confidence intervals obtained using bootstrapping are shown as transparent areas around the mean bootstrapped curve. The radiologist ROC curve and confidence intervals are only plotted for the four PIRADS thresholds. (a) Cancer versus normal/benign. (b) HG cancer versus normal/benign.

in Fig. 3. Statistical significance testing was performed using the area under the ROC curve. Both the RandomForest and the Gentleboost classifier performed significantly better than the linear discriminant classifier in both stages ($p < 0.001$). For the voxel classification stage the random forest classifier also performed significantly better than the Gentleboost classifier ($p < 0.01$). Further analysis of the system was performed using the RandomForest classifier for the voxel stage. Although the mean area under curve was higher for the RandomForest classifier than the Gentleboost classifier in the candidate stage, this was not significantly different. Because the mean area under the curve was slightly higher we chose to use the RandomForest classifier for the candidate stage.

2) Relative Feature and MR Sequence Importance: The rankings for each feature and each MR sequence based on there individual classification performance using the RFC classifer are shown in Tables II(a) and II(b). Although the ADC intensity is the single most important feature in the voxel stage, overall the features calculated using the T2W MR sequence are the most important in voxel classification. Additionally, we can see from the performance of individual features in the voxel

TABLE II
RESULTS FOR RELATIVE FEATURE AND MODALITY IMPORTANCE EXPERIMENT BASED ON LEAVE-ONE-PATIENT-OUT CROSS VALIDATION.
(A) RELATIVE IMPORTANCE BASED ON AUC ON A PER-VOXEL BASIS. (B) RELATIVE IMPORTANCE ON AUC ON A PER-CANDIDATE BASIS

Feature	AUC	Modality	AUC
ADC	0.666	T2W	0.760
Gauss. Texture (Order=-, Scale=2.0)	0.633	DWI	0.746
tau Blob	0.632	DCE	0.714
K_{Trans}	0.629		
T2-map	0.620		
k^{ep}	0.604		
K_{Trans} Blob	0.600		
Gauss. Texture (Order=YY, Scale=5.1)	0.599		
Gauss. Texture (Order=-, Scale=3.17)	0.597		
b800	0.588		

(a)

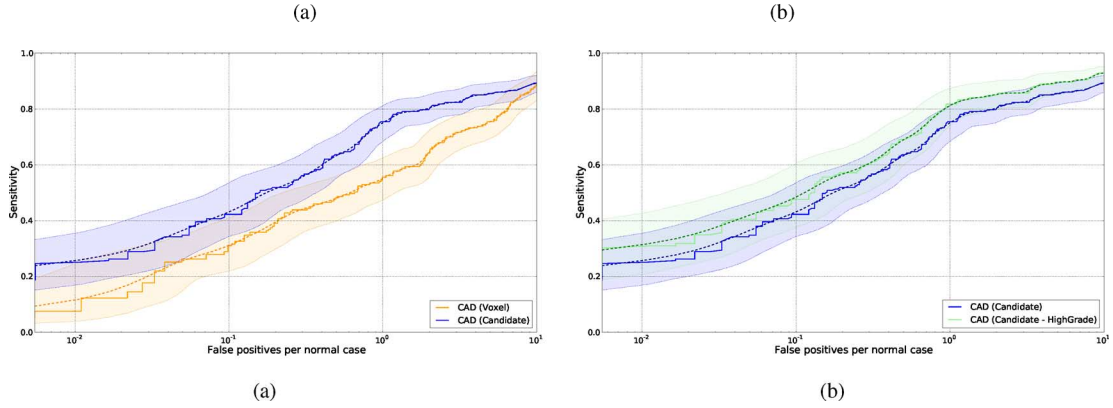| Feature	AUC	Modality	AUC
b800 (standard deviation)	0.805	DWI	0.910
ADC (standard deviation)	0.790	DCE	0.814
Voxel likelihood (mean)	0.765	T2W	0.719
LateWash Blob (standard deviation)	0.750		
ADC Blob (standard deviation)	0.741		
LateWash (standard deviation)	0.731		
Voxel likelihood (75th percentile)	0.727		
Voxel likelihood (standard deviation)	0.723		
k^{ep} (standard deviation)	0.723		
Volume	0.716		

(b)

Fig. 5. FROC analysis of the results of the CAD system. Number of false positives per normal case is shown on a logarithmic scale. The raw FROC curve is shown as the solid line and the mean bootstrapped curve as a dashed line. The 95% confidence intervals obtained using bootstrapping are shown as transparent areas around the mean bootstrapped curve. (a) Results after voxel classification and results after candidate classification compared. Adding the candidate classification step shows a marked improvement over just voxel classification. (b) Results of the candidate classification step for cancer versus normal/benign and high-grade cancer versus normal/benign.

stage that features from each of the MR sequences are selected, showing the importance of using multi-parametric MRI over single-parameter MRI. Finally, the performance per modality is much lower than the overall per-voxel performance (0.76 area under the ROC curve for just T2W and 0.89 when combining all MR sequences).

Inspecting Table II(b), it is interesting to see that especially heterogeneity of feature values within the candidates have high individual performance. Additionally, in the candidate stage, DWI is by far the best performing individual sequence. Furthermore, the initial voxel likelihood plays an important part in classification in the candidate stage. Finally, in this stage features from the DWI imaging are almost as good as using features from the combination of the three MR sequences.

3) FROC Analysis: The FROC curves for detection of prostate cancer and the detection of high-grade prostate cancer are shown in Fig. 5. The results show that adding a candidate classification step reduced the number of false positives at constant sensitivity, e.g., a reduction from approximately 7 to 1.5 false positives per normal case at a sensitivity of 80%. At similar false positive levels, sensitivities were significantly higher after the candidate classification step, e.g., after voxel classification, at 1 false positive per normal case, a sensitivity of 55% is reached, whereas the sensitivity is 75% after the candidate stage ($p < 0.001$). Additionally, the partial area under the curve between 0.1 to 10 false positives per normal case is also significantly higher (7.11 versus 8.74, $p < 0.01$). Furthermore, in Fig. 5(b) the FROC curves are shown for the detection of high-grade cancer versus normal/benign in

addition to the detection of all cancer versus normal/benign. Here, at one false positive per normal case the sensitivity for detecting high-grade cancer is 0.82 and 0.75 for all cancer. This difference is not significant. Additionally, the partial area under the curve between 0.1 and 10 false positives per normal case is not significantly different (8.74 versus 9.06). The maximum sensitivity reached by the system in cancer versus normal/benign case is 0.94 and 0.97 for the high-grade cancer versus normal/benign case. This is caused by 11 and 5 false negatives in those cases respectively. Examples of a true positive and a false positive are shown in Fig. 7. An example of two false negatives is shown in Fig. 6.

4) ROC Analysis: The ROC curves for classifying patients as either having prostate cancer or not having prostate cancer are shown in Fig. 4. Again we also show the improved performance obtained by adding a candidate classification step compared to just using a voxel classification step. In these figures the CAD system is also compared to the clinical diagnosis made by the radiologist for each patient on the basis of the PIRADS system. A radiologist scores each suspicious lesion on a scale from 1 to 5, 1 meaning definitely not cancer and 5 meaning definitely cancer. The patient score is than obtained by taking the highest PIRADS score. For the radiologist only the confidence intervals for the four actual PIRADS thresholds are used for evaluation, as the ROC curve is not well defined at other positions due to the low number of thresholds.

The addition of the candidate classification shows a marked improvement when evaluating on a per-patient basis, with an increase in AUC from 0.722 to 0.81 ($p < 0.01$) and from 0.73
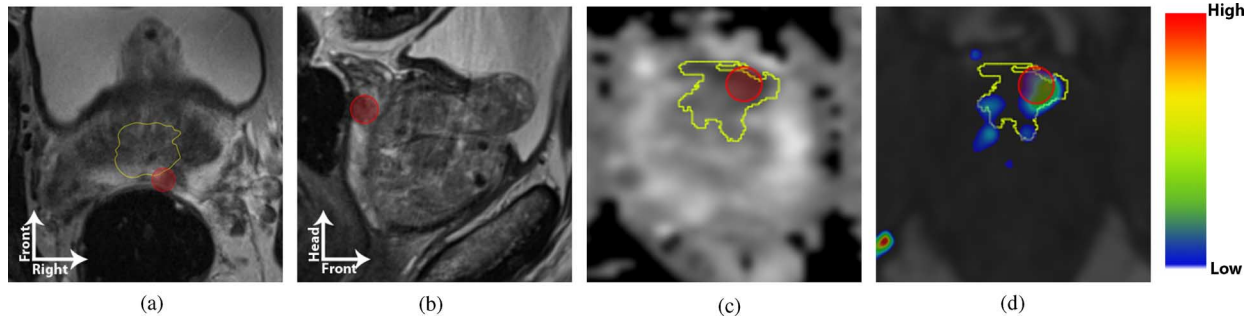
Fig. 6. Examples of false negatives (FN). The location of the false negatives is indicated with the red circle. False negative 1 is caused by segmentation errors. In figure a the prostate segmentation is indicated in yellow. The segmentation most likely fails due to the fact that the prostate is growing into the bladder, as can be seen in b. False negative 2 is caused by our evaluation criterion, the region segmentation [indicated in yellow in (c) and (d)] is quite large and therefore the mark (the red sphere) and the center of gravity of the region segmentation are more than 10 mm apart. (a) FN 1 (Transversal). (b) FN 1 (Sagittal). (c) FN 2 (ADC). (d) FN 2 (DCE + Ktrans).
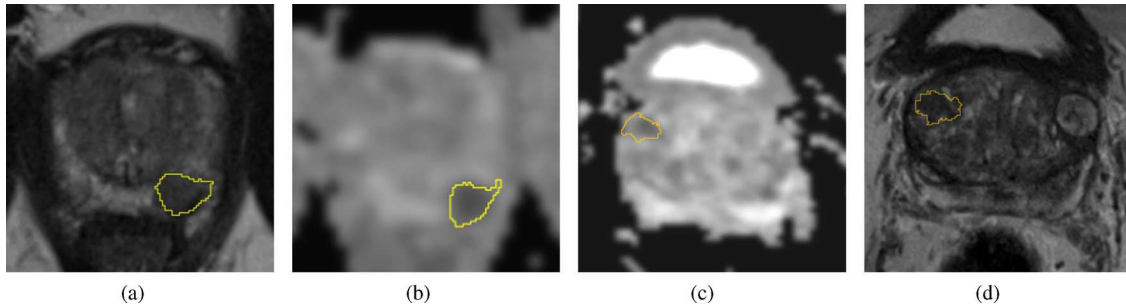

Fig. 7. Examples of a true positive (a)–(b) and a false positive (c)–(d) from the CAD system. (a) True positive (T2W). (b) True positive (ADC). (c) False positive (T2W). (d) False positive (ADC).

to 0.83 ($p < 0.01$) for high-grade cancer versus normal/benign. At a high specificity (left part of the ROC curve) of 0.88 (PIRADS score 5), there is no significant difference between the radiologist and the CAD system ($p = 0.334$ for detection of cancer, $p = 0.37$ for detection of high-grade cancer). At the other thresholds the radiologist performance is significantly better than the CAD system ($p < 0.01$). The radiologist is significantly better at every PI-RADS threshold compared to CAD system when only using the voxel stage.

IV. DISCUSSION

A CAD system which detects prostate cancer in MRI images was presented in this paper. The performance of the system was evaluated on a large consecutive set of patients, with MR-guided biopsy as a reference standard. Quantitatively, the area under the ROC curve for classifying patients was 0.81. If we investigate the performance from an FROC perspective, at 1 false positive per image we obtain a sensitivity of 0.75 for detecting any cancer lesion and 0.83 for detecting a high-grade cancer lesion. Compared to the radiologist, the system shows no significant differences in performance at high specificity (Fig. 4, left part of the curve). However, at lower specificity the radiologist performs significantly better ($p < 0.01$). Furthermore, in both Figs. 4 and 5 we show that adding a candidate stage in addition to a voxel stage significantly improves performances on both a per-lesion (sensitivity at 1 false positive per normal patient increases from 0.55 to 0.75) and per-patient level (area under the ROC curve increases from 0.72 to 0.81).

There are several potential use-cases for this system. The system could be used as an independent second observer in a screening setting, where radiologists read at high specificity as the vast majority of screened patients is healthy; a high percentage of false positives would be too costly (due to subsequent over-diagnosis and over-treatment). Double reading in screening is quite typical and the CAD system in this paper could potentially replace one of the radiologists, thus cutting work in half. We do note that the system was not yet evaluated in an observer study in such a setting.

A second potential use-case is as a triage test (first-reader setting). Here the CAD system would be used at high sensitivity, before any radiologist will look at the images. If we look at the potential of our CAD system in such a setting, at an operating point with a sensitivity of 96% (similar to the radiologists, [34]), the specificity of the CAD system is between 15%–40%. This could indicate that between 15%–40% of all studies would potentially require no human intervention, which could reduce the workload of the radiologist substantially. Additionally, the system can of course be used in both settings simultaneously.

If we compare our system to the current state-of-the-art, two types of systems can be distinguished: systems which perform only a voxel-based analysis and systems which perform both a voxel-based analysis and a candidate evaluation step. For the first type of system, Tiwari *et al.* have shown the best voxel classification performance using a system with manual prostate segmentation and MR spectroscopy. They obtain an average area under the curve of 0.89, which is similar to our results obtained during the classifier comparison at the voxel level Fig. 6, average area under the curve of 0.889). However, we show in our study set that a voxel classification performance of 0.89 only results in a per-lesion classification performance of 0.55 sen-

sitivity at 1 false positive per normal patient and a per-patient area under the ROC curve of 0.72. Our addition of a subsequent candidate classification step increases the performance of the system by a significant amount (0.75 sensitivity at 1 false positive per normal patient ($p < 0.01$), figure 8 and 0.81 area under the ROC curve for a per-patient analysis ($p < 0.01$). However, as Tiwari *et al.* did not extend their system to a per-region and per-patient evaluation we cannot directly compare this. We do have to note that this comparison has not been made on the same data set, which is unfortunate, but there is no open availability of a significant number of multi-parametric prostate MRIs. We are currently considering organizing a prostate cancer detection challenge similar to other grand challenges in medical image analysis to allow our algorithm to fairly compete against others. For the second type of system incorporating a candidate detection and a candidate classification step we can compare our results to Vos *et al.* [10], which is the only other two-stage system. Instead of a voxel classification step, they use a blob detector to obtain the candidates. At 0.1, 1 and 10 false positives per normal case they obtained a sensitivity of 0.15, 0.48, and 0.89 where we obtain a sensitivity of 0.42, 0.75, and 0.89. Especially at the lower false positive rates we obtain a substantially better performance.

The use of multi-parametric MRI over single-parameter MRI is already part of clinical guidelines [16]. In this study we investigated the performance of individual features and MR sequences in both the voxel and candidate stages of our CAD system. Especially in the voxel stage, the combined interpretation of all three MR sequences, T2-weighted imaging, diffusion-weighted imaging and dynamic contrast-enhanced imaging shows a large improvement over using any single MR sequence (highest performing single sequences AUC is 0.76, combination is 0.89), and all play about an equal role. In the subsequent candidate stage there is a preference for DWI over DCE over T2W. This experiment also showed that the individual performance of features was relatively low compared to the combination of all features (best performing feature in the voxel stage had a AUC of 0.66, whereas the overall voxel stage AUC was 0.89). These observations confirm clinical practice. A limitation of this study is the fact that the ROC evaluation is positively biased toward the radiologist. Although the reference standard for cancerous regions is well defined by the MR-guided biopsy specimens, for most of the normal regions we have to depend on the opinion of the radiologist. While we incorporated only data with either negative biopsy results or very low PI-RADS scores (1 and 2) there is still the risk that some areas we deem normal are actually prostate cancer. Furthermore, in the evaluation and the comparison to the radiologist, it may well be that the radiologists did have some false negatives. Recently, prospective preliminary results were published by Thompson *et al.* [34] They found that the sensitivity for radiologists for detecting high-grade prostate cancer was 96%. Another limitation is the fact that although the multi-parametric MRI is implicitly registered (all sequences are acquired in one go, without the patient leaving the scanner), registration errors between the different sequences could occur due to patient movement. This was mostly circumvented in our data by 1) not using an endorectal coil, which significantly

improves patient comfort and as such reduces patient movement, and 2) by administering Buscopan prior to the MRI to reduce bowel movement, and 3) using multi-scale features (Gaussian texture, blobness) where exact voxel alignment is less important. Further improvement could be achieved by implementing a registration algorithm for prostate MRI. However, this is currently an untackled problem, which is nontrivial to solve. Registration in multi-parametric prostate MRI would require within-modality registration (different time-points in DCE-MRI, different b-values in DWI-MRI) and between-modality registration (where the registration has to cope with large resolution and appearance differences). Nonetheless, registration certainly has potential value and is an interesting topic for future work in this area.

The false negatives in our system are mostly caused by prostate segmentation errors. Of the 11 false negatives after candidate classification, six are caused by the prostate segmentation missing the lesion entirely or partly. Most of these issues can be solved by incorporating a segmentation method which is more robust to strange prostate shapes. An example is shown in Fig. 6. Although the prostate segmentation algorithm is a candidate for improvement, it is missing only six out of 183 lesions (or 3% of total sensitivity), which is still a reasonable result. For the remaining false negatives, in four cases the lesion was detected, but the candidate segmentations were so large that the center of the candidate segmentation and the lesion marker were more than 10 mm apart, thus failing our criterion for a true positive. This can happen because in big lesions the radiologist did not always put the point annotation at the center of the lesion. For one false negative the area was not identified by the voxel classification and the local maxima detection and thus lost to the second part of the system.

Summarizing, a fully automated CAD system was developed for the detection of prostate cancer in MRI images. Performance evaluation shows that it outperforms the state-of-the-art, although the comparison has its limitations due to different evaluation data sets. The system is not significantly different from radiologist performance at high specificity. Therefore, we believe it to be a potentially valuable tool to aid radiologists, in both first- and second-reader settings.

References

[1] R. Siegel, D. Naishadham, and A. Jemal, "Cancer statistics, 2013," *CA Cancer J. Clin.*, vol. 63, pp. 11–30, 2013.
[2] F. H. Schröder, J. Hugosson, M. J. Roobol, T. L. J. Tammela, S. Ciatto, V. Nelen, M. Kwiatkowski, M. Lujan, H. Lilja, M. Zappa, L. J. Denis, F. Recker, A. Berenguer, L. Määttänen, C. H. Bangma, G. Aus, A. Villers, X. Rebillard, T. van der Kwast, B. G. Blijenberg, S. M. Moss, H. J. de Koning, and A. Auvinen, E. R. S. P. C. Investigators, "Screening and prostate-cancer mortality in a randomized European study," *N. Eng. J. Med.*, vol. 360, pp. 1320–1328, 2009.
[3] T. Hambrock, C. Hoeks, C. Hulsbergen-van de Kaa, T. Scheenen, J. Fütterer, S. Bouwense, I. van Oort, F. Schröder, H. Huisman, and J. Barentsz, "Prospective assessment of prostate cancer aggressiveness using 3-T diffusion-weighted magnetic resonance imaging-guided biopsies versus a systematic 10-core transrectal ultrasound prostate biopsy cohort," *Eur. Urol.*, vol. 61, pp. 177–184, 2012.
[4] K. Kitajima, Y. Kaji, Y. Fukabori, K. Yoshida, N. Suganuma, and K. Sugimura, "Prostate cancer detection with 3 T MRI: Comparison of diffusion-weighted imaging and dynamic contrast-enhanced MRI in combination with T2-weighted imaging," *J. Magn. Reson. Imag.*, vol. 31, pp. 625–631, 2010.

[5] N. Karssemeijer, J. D. M. Otten, H. Rijken, and R. Holland, "Computer aided detection of masses in mammograms as decision support," *Br. J. Radiol.*, vol. 79, pp. S123–S126, 2006, Spec No 2.

[6] R. M. Summers, J. Liu, B. Rehani, P. Stafford, L. Brown, A. Louie, D. S. Barlow, D. W. Jensen, B. Cash, J. R. Choi, P. J. Pickhardt, and N. Petrick, "CT colonography computer-aided polyp detection: Effect on radiologist observers of polyp identification by CAD on both the supine and prone scans," *Acad. Radiol.*, vol. 17, pp. 948–959, 2010.

[7] I. Chan, W. Wells, R. V. Mulkern, S. Haker, J. Zhang, K. H. Zou, S. E. Maier, and C. M. C. Tempany, "Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel statistical classifier," *Med. Phys.*, vol. 30, pp. 2390–2398, 2003.

[8] D. L. Langer, T. H. van der Kwast, A. J. Evans, J. Trachtenberg, B. C. Wilson, and M. A. Haider, "Prostate cancer detection with multi-parametric MRI: Logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI," *J. Magn. Reson. Imag.*, vol. 30, pp. 327–334, 2009.

[9] P. Tiwari, S. Viswanath, J. Kurhanewicz, A. Sridhar, and A. Madabhushi, "Multimodal wavelet embedding representation for data combination (maweric): Integrating magnetic resonance imaging and spectroscopy for prostate cancer detection," *NMR Biomed.*, 2011.

[10] P. C. Vos, J. O. Barentsz, N. Karssemeijer, and H. J. Huisman, "Automatic computer-aided detection of prostate cancer based on multiparametric magnetic resonance image analysis," *Phys. Med. Biol.*, vol. 57, pp. 1527–1542, 2012.

[11] P. Tiwari, J. Kurhanewicz, and A. Madabhushi, "Multi-kernel graph embedding for detection, gleason grading of prostate cancer via MRI/MRS," *Med. Image Anal.*, vol. 17, pp. 219–235, 2013.

[12] E. Niaf, O. Rouvière, F. Mège-Lechevallier, F. Bratan, and C. Lartizien, "Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI," *Phys. Med. Biol.*, vol. 57, p. 3833, 2012.

[13] P. C. Vos, T. Hambrock, J. O. Barentsz, and H. J. Huisman, "Computer-assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI," *Phys. Med. Biol.*, vol. 55, pp. 1719–1734, 2010.

[14] A. Firjani, A. Elnakib, F. Khalifa, G. Gimelfarb, M. A. El-Ghar, A. Elmaghraby, and A. El-Baz, "A diffusion-weighted imaging based diagnostic system for early detection of prostate cancer," *J. Biomed. Sci. Eng.*, vol. 6, 2013.

[15] G. Litjens, J. Barentsz, N. Karssemeijer, and H. Huisman, "Automated computer-aided detection of prostate cancer in MR images: From a whole-organ to a zone-based approach," in *Proc. SPIE Med. Imag.*, 2012, vol. 8315, pp. 83 150G–83 150G-6.

[16] J. O. Barentsz, J. Richenberg, R. Clements, P. Choyke, S. Verma, G. Villeirs, O. Rouviere, V. Logager, and J. J. Fütterer, European Society of Urogenital Radiology, "ESUR prostate MR guidelines 2012," *Eur. Radiol.*, vol. 22, pp. 746–757, 2012.

[17] S. Klein, U. A. van der Heide, I. M. Lips, M. van Vulpen, M. Staring, and J. P. W. Pluim, "Automatic segmentation of the prostate in 3-D MR images by atlas matching using localized mutual information," *Med. Phys.*, vol. 35, pp. 1407–1417, 2008.

[18] T. R. Langerak, U. A. van der Heide, A. N. T. J. Kotte, M. A. Viergever, M. van Vulpen, and J. P. W. Pluim, "Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE)," *IEEE Trans. Med. Imag.*, vol. 29, no. 12, pp. 2000–2008, Dec. 2010.

[19] C. Studholme, C. Drapaca, B. Iordanova, and V. Cardenas, "Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change," *IEEE Trans. Med. Imag.*, vol. 25, no. 5, pp. 626–639, May 2006.

[20] G. J. S. Litjens, O. A. Debats, W. J. M. van de Ven, N. Karssemeijer, and H. J. Huisman, "A pattern recognition approach to zonal segmentation of the prostate on MRI," in *Med. Image Comput. Comput. Assist. Interv.*, 2012, vol. 7511, pp. 413–420.

[21] Q. Li, S. Sone, and K. Doi, "Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans," *Med. Phys.*, vol. 30, pp. 2040–2051, 2003.

[22] H. J. Huisman, M. R. Engelbrecht, and J. O. Barentsz, "Accurate estimation of pharmacokinetic contrast-enhanced dynamic MRI parameters of the prostate," *J. Magn. Reson. Imag.*, vol. 13, pp. 607–614, 2001.

[23] H. Laue, S. Behrens, F. Giesel, N. Hoggard, C. Zechmann, I. Wilkinson, S. Krass, and H.-O. Peitgen, "A statistical comparison of tofts and brix model parameters for glioma and prostate MRI data," in *Magn. Reson. Mater. Phys., Biol. Med.*, 2006, vol. 19, p. 318.

[24] P. S. Tofts, G. Brix, D. L. Buckley, J. L. Evelhoch, E. Henderson, M. V. Knopp, H. B. Larsson, T. Y. Lee, N. A. Mayr, G. J. Parker, R. E. Port, J. Taylor, and R. M. Weisskoff, "Estimating kinetic parameters from dynamic contrast-enhanced t(1)-weighted MRI of a diffusable tracer: Standardized quantities and symbols," *J. Magn. Reson. Imag.*, vol. 10, pp. 223–232, 1999.

[25] S. E. Viswanath, N. B. Bloch, J. C. Chappelow, R. Toth, N. M. Rofsky, E. M. Genega, R. E. Lenkinski, and A. Madabhushi, "Central gland and peripheral zone prostate tumors have significantly different quantitative imaging signatures on 3 tesla endorectal, in vivo T2-weighted MR imagery," *J. Magn. Reson. Imag.*, vol. 36, pp. 213–224, 2012.

[26] T. Wolters, M. J. Roobol, P. J. van Leeuwen, R. C. N. van den Bergh, R. F. Hoedemaeker, G. J. L. H. van Leenders, F. H. Schröder, and T. H. van der Kwast, "A critical analysis of the tumor volume threshold for clinically insignificant prostate cancer using a data set of a randomized screening trial," *J. Urol.*, vol. 185, pp. 121–125, 2011.

[27] W. J. M. van de Ven, C. A. Hulsbergen-van de Kaa, T. Hambrock, J. O. Barentsz, and H. J. Huisman, "Simulated required accuracy of image registration tools for targeting high-grade cancer components with prostate biopsies," *Eur. Radiol.*, vol. 23, pp. 1401–1407, 2013.

[28] F. Bratan, E. Niaf, C. Melodelima, A. L. Chesnais, R. Souchon, F. Mège-Lechevallier, M. Colombel, and O. Rouvière, "Influence of imaging and histological factors on prostate cancer detection and localisation on multiparametric MRI: A prospective study," *Eur. Radiol.*, pp. 1–11, 2013.

[29] J. Friedman, T. Hastie, and R. Tibshirani, "Special invited paper. Additive logistic regression: A statistical view of boosting," *Ann. Stat.*, vol. 28, pp. 337–374, 2000.

[30] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.

[31] J. M. Kuhnigk, V. Dicken, L. Bornemann, A. Bakai, D. Wormanns, S. Krass, and H. O. Peitgen, "Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans," *IEEE Trans. Med. Imag.*, vol. 25, no. 4, pp. 417–434, Apr. 2006.

[32] B. Platel, H. Huisman, H. Laue, R. Mus, R. Mann, H. Hahn, and N. Karssemeijer, "Computerized characterization of breast lesions using dual-temporal resolution dynamic contrast-enhanced MR images," in *MICCAI Workshop: Breast Image Anal.*, 2011.

[33] G. Litjens, T. Hambrock, C. Hulsbergen-van de Kaa, J. Barentsz, and H. Huisman, "Interpatient variation in normal peripheral zone apparent diffusion coefficient: Effect on the prediction of prostate cancer aggressiveness," *Radiology*, vol. 265, pp. 260–266, 2012.

[34] J. Thompson, P. Stricker, P. Brenner, D. Moses, R. Shnier, W. Delprado, A. M. Haynes, and A. Hayen, "Magnetic resonance imaging detects significant prostate cancer and could be used to reduce unnecessary biopsies: Initial results from a prospective trial," *J. Urol.*, vol. 189, pp. e910–-e911, Apr. 2013.