Article 25fa pilot End User Agreement

This publication is distributed under the terms of Article 25fa of the Dutch Copyright Act (Auteurswet) with explicit consent by the author. Dutch law entitles the maker of a short scientific work funded either wholly or partially by Dutch public funds to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) ‘Article 25fa implementation’ pilot project. In this pilot research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and/or copyrights owner(s) of this work. Any use of the publication other than authorised under this licence or copyright law is prohibited.

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please contact the Library through email: copyright@ubn.ru.nl, or send a letter to:

University Library
Radboud University
Copyright Information Point
PO Box 9100
6500 HA Nijmegen

You will be contacted as soon as possible.
Automatic computer-aided detection of prostate cancer based on multiparametric magnetic resonance image analysis

P C Vos, J O Barentsz, N Karssemeijer and H J Huisman

Radboud University Nijmegen Medical Centre, Department of Radiology, 6500 HB, Nijmegen
E-mail: p.vos@rad.umcn.nl

Received 11 August 2011, in final form 4 December 2011
Published 6 March 2012
Online at stacks.iop.org/PMB/57/1527

Abstract

In this paper, a fully automatic computer-aided detection (CAD) method is proposed for the detection of prostate cancer. The CAD method consists of multiple sequential steps in order to detect locations that are suspicious for prostate cancer. In the initial stage, a voxel classification is performed using a Hessian-based blob detection algorithm at multiple scales on an apparent diffusion coefficient map. Next, a parametric multi-object segmentation method is applied and the resulting segmentation is used as a mask to restrict the candidate detection to the prostate. The remaining candidates are characterized by performing histogram analysis on multiparametric MR images. The resulting feature set is summarized into a malignancy likelihood by a supervised classifier in a two-stage classification approach. The detection performance for prostate cancer was tested on a screening population of 200 consecutive patients and evaluated using the free response operating characteristic methodology. The results show that the CAD method obtained sensitivities of 0.41, 0.65 and 0.74 at false positive (FP) levels of 1, 3 and 5 per patient, respectively. In conclusion, this study showed that it is feasible to automatically detect prostate cancer at a FP rate lower than systematic biopsy. The CAD method may assist the radiologist to detect prostate cancer locations and could potentially guide biopsy towards the most aggressive part of the tumour.

(Some figures may appear in colour only in the online journal)

1. Introduction

Prostatic adenocarcinoma (PCa) is the second leading cause of cancer-related deaths among males in the United States, with an estimated number of 217 730 new cases in 2010 (Jemal et al 2010). Early detection of PCa can be life saving. Recently, it was demonstrated in a large randomized European study that prostate specific antigen (PSA)-based screening reduces the...
rate of death from PCa by 31%. However, the benefit was associated with a high risk of overdiagnosis and overtreatment (Schröder et al. 2009, Roobol et al. 2009). Moreover, the PSA test is not able to predict the aggressiveness of the cancer. As a result, slow-growing and non-aggressive prostate cancer is frequently diagnosed in older patients but is not the main cause of death in these patients.

PSA is a nonspecific marker for prostate cancer. As a result, urologists are often faced with the dilemma of how to manage a patient with a high PSA level and an initial set of negative prostate biopsies. Hence, the possibility remains that these patients may still have tumour, as prostate cancer is often multifocal and heterogeneous in nature. Systematic prostate transrectal ultrasound (TRUS)-guided biopsy is the standard procedure for prostate histological sampling. The technique involves systematic sampling of multiple areas in the prostate during TRUS-guided biopsy regardless of the presence of hypoechoic lesions. In recent years, many reports have shown that systematic biopsies do not detect all clinically significant cancers and efforts have been made to improve the protocol by increasing the number of biopsies and/or changing biopsy positions. Roehl et al. (2002) showed in a large study that by using routine 6-sector TRUS-guided biopsy, nearly a quarter (23%) of detectable cancers were missed. In other words, TRUS-guided biopsy has a sensitivity of 77% with a false positive (FP) rate of 6. Nevertheless, the volume of prostatic tissue sampled is relatively small which makes it difficult to detect tumour. More importantly, the technique fails to sample the most representative part of the tumour (Hodge et al. 1989, Djavan et al. 2001). To prevent patient anxiety, more accurate methods need to be found to detect or rule out significant disease (Carlsson et al. 2007).

Prostate magnetic resonance imaging (MRI) has the potential to improve the specificity of PSA-based screening scenarios as a non-invasive detection tool. Several studies showed that combining anatomical, functional and metabolic MRI information leads to a PCa detection accuracy of up to 92% (Füutterer et al. 2006, Haider et al. 2007, Puech et al. 2009, Tanimoto et al. 2007, Kitajima et al. 2010). Furthermore, multiparametric MRI can target biopsies towards regions determined to be suspicious of cancer (Hambrock et al. 2010). Unfortunately, prostate MRI analysis requires a high level of expertise and suffers from observer variability (Lim et al. 2009). Furthermore, the interpretation of the multiple MR images and their derived maps for a single patient diagnosis is a labour-intensive procedure. For that reason, the technique is considered cost-inefficient and, as a result, has not been implemented in a screening environment (Hoeks et al. 2009).

Computer-aided detection (CAD) systems can be of benefit to improve the diagnostic accuracy of the radiologist, reduce reader variability and speed up the reading time. CAD aims to automatically highlight cancer suspicious regions, leading to a reduction of search and interpretation errors, as well as a reduction of the variation between and within observers (Giger et al. 2008). CAD research has been successfully pursued in other diagnostic areas such as mammography (Karssemeijer et al. 2006, Singh et al. 2008), CT chest (Ge et al. 2005, Beigelman-Aubry et al. 2009, Hogeweg et al. 2010), CT colonography (Graser et al. 2007, Summers et al. 2010) as well as retinal imaging (Abramoff et al. 2008). CAD systems generally consist of multiple sequential stages. In the initial stage, lesion candidates are selected within a likelihood map that was generated by a voxel classification of one or more images. Hereafter, the lesion candidates are segmented into a region of interest from which region-based features are extracted. Finally, the extracted information is fused by a classifier into a malignancy likelihood. The last stage ensures a reduction of the amount of FPs that were localized in the initial stage.

Automatic CAD of PCa based on multiparametric MRI analysis

Figure 1. Dataflow diagram of the implemented CAD system. A Hessian-based filter initially detects possible lesion candidates at multiple scales by detecting local maxima in the generated blob likelihood map. A prostate segmentation method is applied to the pelvis and the resulting prostate segmentation is used as a mask to restrict the candidate detection to the prostate. A region of interest is defined surrounding the candidates from which features are extracted using histogram analysis. The features are summarized by a supervised classifier to calculate the likelihood of malignancy in the region of interest.

MR images using mathematical descriptors. These studies showed on a voxel basis that the discrimination between benign and malignant tissues is feasible with good performances. Recently, we presented a work that focused on the regional classification stage (Vos et al 2010). In the proposed CAD method, the radiologist was instructed to localize a lesion candidate in the peripheral zone of the prostate and delineate a region of interest. Hereafter, relevant features from multiparametric MRI were extracted on demand and summarized by a supervised classifier into a malignancy likelihood. Experience with the system, however, showed that the semi-automatic approach is subject to observer variability due the differences in lesion segmentation or incorrect segmentation. Furthermore, tumours located in the transition zone were not included. To the best of our knowledge, a fully automated prostate CAD method based on multiparametric MRI analysis has not been described in the literature.

The purpose of this study was to investigate the feasibility of a CAD method that fully automatically detects cancer suspicious regions in the prostate. The study focused on a population that included patients with elevated PSA levels with one negative biopsy. The ultimate goal was to detect more tumour regions at a lower FP rate than systematic biopsy.

2. Method

2.1. Overview

The proposed CAD method is schematically outlined in figure 1. It comprises of multiple sequential steps in order to detect locations that are suspicious for prostate cancer. In the initial part of the CAD scheme, the method detects lesion candidates in apparent diffusion coefficient (ADC) maps that are acquired during the MR examination. Firstly, a voxel classification is performed using a Hessian-based blob detection algorithm at multiple scales. Next, a parametric multi-object segmentation method is applied to the pelvis to segment the prostate automatically. The prostate segmentation is used as a mask to restrict the candidate detection to the prostate. Hereafter, candidate lesions are determined by detecting local maxima in the generated blob likelihood map and are characterized by performing histogram analysis within a region of interest on multiple MR images. Finally, the extracted features are summarized by
a linear discriminant analysis (LDA) classifier into a malignancy likelihood. The individual steps of the scheme will be explained in more detail in the remainder of this section.

2.2. Initial voxel classification

In the first stage of the CAD system, dark blob-like regions are localized in the ADC map. This approach was inspired by the clinical practise of the radiologist at our hospital as they localize lesions by the more structured property of PCa in an ADC map. A common approach to automatically determine the blob likelihood of a voxel \( x \) is to use the eigenvalues \( \lambda_{\sigma k} \) of the Hessian matrix \( H_\sigma \) at scale \( \sigma \) of the ADC map, with \( k = 1, 2, 3 \) (Frangi et al 1998). The likelihood of a voxel at scale \( \sigma \) to belong to a blob is defined by Qiang et al (2003):

\[
P(x, \sigma) = \begin{cases} \frac{\lambda_{\sigma 1}(x) \lambda_{\sigma 1}(x)}{|\lambda_{\sigma 3}(x)|} & \lambda_{\sigma k}(x) < 0, \ k = 1, 2, 3 \\ 0, & \end{cases}
\]

(1)

Note that the three eigenvalues are sorted as \( |\lambda_{\sigma 1}(x)| < |\lambda_{\sigma 2}(x)| < |\lambda_{\sigma 3}(x)| \). The approach is applied using a recursive implementation of the Gaussian filter at multiple scales, namely three scales are used: 8, 10 and 12 mm in diameter. The blob detector is normalized for each scale. The likelihood \( L(x) \) of a voxel \( x \) to belong to a blob is finally given as

\[
L(x) = \max_{\sigma_{\text{min}} \leq \sigma \leq \sigma_{\text{max}}} L(x, \sigma).
\]

(2)

2.3. Prostate segmentation

The initial voxel classification is performed on the whole ADC map to prevent the need for boundary conditions. As a result, the automatic prostate segmentation is crucial to avoid detection of local maxima that lie outside the prostate. We used a parametric multi-object method that was developed in previous work (Litjens et al 2011). The method consists of two steps: model fitting and voxel segmentation with prior model constraints. In the first stage, a model is constructed based on multiple parametric objects that define the shape and appearance of each organ in multiparametric MR images. For this paper, the ADC and T2 maps were used to automatically segment the prostate. The model is fitted to the different MR images simultaneously. A realistic organ representation is obtained by constraining the parameters within a population model. In the second stage, a Bayesian framework is used to obtain the final segmentation as previously described in Litjens et al (2011). Here, the fitted model is used as a prior to the Naive Bayes classifier such that prior information about spatial and multivariate appearance relations between anatomical structures is efficiently taken into account.

The final label image is denoted by \( B(x) \) where the image voxels are labelled with the corresponding organ. An example of the segmentation method is shown in figure 2.

2.4. Lesion candidate detection

This section describes how lesion candidates are selected from the obtained likelihood map \( L(x) \). Obviously, a lesion candidate \( i \) should lie within the prostate segmentation \( B(x) \) to be selected. Additionally, the candidates are selected based on using the following peak detection criteria: the peak value \( L(x_i) \) should exceed \( \tau \); \( L(x) \) should exceed the mean value of its sphere-shaped neighbourhood \( \Omega \) with diameter \( d \) and the difference between \( L(x_i) \) and the
Figure 2. Cross-sectional transversal views of (a) ADC and (b) T2-weighted image, with a transparent (colour) overlay of the segmentation: bladder (red); prostate (blue) and rectum (green).

Figure 3. Cross-sectional transversal views of an example ADC map of a patient with a tumour in the transition zone. (a) Multiple dark blob-like regions are visible in and outside the prostate. (b) The ADC map is shown with the blob likelihood map displayed as a colour-coded overlay. Lesion candidates are detected in the pelvis showing FPs outside the prostate. These are ignored using the prostate segmentation. (c) The remaining lesion candidates are displayed within the prostate segmentation. The (red) arrow indicates a tumour with Gleason grade 3+4 which was detected by the peak detector with the highest likelihood of all detected blobs.

mean neighbourhood value should be more than the squelch threshold $\epsilon$. Let $\phi$ be the group of final selected candidates, then $x_i \in \phi$ when

$$B(x_i) = 1 \land L(x_i) > \tau \land (L(x_i) - \text{mean}(\Omega_1(x_i), r)) > \epsilon.$$  

For this paper, the diameter $d$ was set to 5 mm which represents the minimal size of a significant prostate tumour (Wolters et al 2011), and $\epsilon$ and $\tau$ were empirically set to 0.1 and 1, respectively. An example of the candidate detection procedure is shown in figure 3.

Candidates were ignored when the corresponding MR values were outside empirically determined thresholds. The following thresholds were obtained from the literature: the ADC value was below 200 or above normal prostate diffusion of $1600 \text{ mm s}^{-2}$ (Matsuki et al 2007); the T2 relaxation time was above normal prostate 100 ms or below muscle 36 ms (Gibbs et al 2001) and the interstitial volume $V_e$ was below normal prostate of 20 mmHg (Delorme and Knopp 1998).
Figure 4. Cross-sectional transversal views of an example ADC map, candidate detection and pharmacokinetic map to illustrate the idea that hotspots are visible at different locations among multiparametric MR images. (a) The arrows indicate the extent of the whole tumour with Gleason score 4+3. The (red) arrow indicates the detected location by the candidate detector, where the ADC values are most low and the Gleason 4 component is present. (b) The pharmacokinetic parameter showed multiple enhancing hotspots within the tumour. Note that the location found by the candidate detector (red cursor) and the nearby hotspot (blue region) do not overlap. The difference in location is approximately 2.5 mm.

2.5. Local feature analysis

Prostate cancers can be discriminated from benign abnormalities by their strong heterogeneity. That is hotspots or local enhancements are visible in multiparametric MR images at different locations within the extent of the tumour. Because lesion segmentation methods typically are applicable to monoparametric MR images only, crucial information available from the multiparametric MR images may be missed. As a result, those approaches potentially underestimate the size and, more importantly, the grade of the tumour. Figure 4 illustrates the idea that hotspots are visible at different locations among multiparametric MR images. For this paper, we therefore define a spherical region $R$ with radius $r$ that surrounds a lesion candidate $i$ at location $x_i$ such that analysis can be performed on multiple MR images simultaneously.

Analysis was performed within a region $R$ by combining a histogram analysis of T2, pharmacokinetic, T1 and ADC maps with texture-based features. Quartiles were used for further analysis as they are less sensitive to extreme values often observed in the MR image data. In total, nine features were collected from the MR data within each $R$. The selected intensity-based features reflect the clinical practise at our hospital, where prostate cancer diagnosis is performed on a daily basis (Füitterer et al 2006). Based on the preferences in our clinic, we expect the following intensity features to be the most representative for detecting aggressive cancers.

$f_1$ 25% percentile $T2$. The 25% percentile was extracted from the T2 map as it has been established that prostate cancer typically demonstrates lower T2 relaxation time than normal prostate tissue (Wang et al 2008).

$f_2$ 25% percentile $ADC$. The 25% percentile was extracted from the ADC map because lower ADC values in prostate cancer are related to tightly packed glandular elements found in
cancers that locally replace the fluid-containing peripheral and transition zone ducts (Tamada et al. 2008, Langer et al. 2010, Hambrock et al. 2011).

\( f_3 \) 75\% percentile \( K^{\text{trans}} \). The pharmacokinetic parameter \( K^{\text{trans}} \) or transfer constant (1/min) relates to the permeability surface area product. The permeability area surface product refers to the ability of tracer molecules to pass through interendothelial fenestrae and junctions into the interstitial compartment. High permeability of the vasculature is a characteristic of pathological blood vessels in inflamed tissues and tumors. An increased capillary permeability is observed in prostate cancer (Collins et al. 2004). The upper quartile captures the presence of hotspots.

\( f_4 \) 75\% percentile \( V_e \). In the extravascular extracellular space (EES) of normal tissue, pressure is near atmospheric (25 mmHg) values, whereas in tumors it may reach 50 mmHg or even more. The interstitial hypertension may be due to increased vascular permeability in combination with a lack of lymphatic drainage due to the absence of functional lymphatic vessels within the tumor itself. This results in an increase of the EES. The EES is defined as percentage per unit volume of tissue. An increased interstitial leakage space is observed in tumor; hence, the upper quartile is used to capture these hotspots (Delorme and Knopp 1998).

\( f_5 \) 25\% percentile wash-out. The kinetic parameter wash-out quantifies the slope of the curve after the first wash-in phase. Although it does not directly correlate with physiological parameters, such as pharmacokinetic parameters, the presence of wash-out is considered highly indicative of PCa. When capillary permeability is high, the backflow of contrast medium is also rapid, resulting in a negative wash-out following the shape of the plasma concentrations. The 25th percentile is used because the presence of wash-out is often heterogeneous within the extent of the tumour (Collins et al. 2004).

\( f_6 \) 50\% percentile \( T1 \) map. Post-biopsy hemorrhage mimics high tumour vascularity. Fortunately, hemorrhage is clearly visible as a high-intensity area on a T1-weighted (T1-w) image. As biopsy hemorrhage is often visible as a large homogeneous area, the 50\% percentile is extracted from the T1 map that was automatically generated from the T1-w image as described in Hittmair et al. (1994).

The following texture features were extracted.

\( f_7 \) Peak value. The peak value or blob likelihood \( L(x_i) \) that was obtained after the initial voxel classification stage for a lesion candidate \( i \) at location \( x_i \).

\( f_8 \) Mean neighbourhood value. The mean neighbourhood value of \( \Omega(\cdot) \) at location \( x_i \).

\( f_9 \) Squelch value. The squelch value is the difference between the peak value \( f_7 \) and the mean neighbourhood value \( f_8 \).

The assumption is that all image volumes \( I_1, I_2, \ldots, I_k \) are registered to each other in the MR coordinate system and as a result, a lesion segmentation in \( I_k \) will represent the same lesion area in \( I_{k+1} \), regardless of the image resolution or orientation. A mutual information affine registration strategy was applied to correct for patient movement such that the assumption will hold (Vos et al. 2010).

Let \( \theta_i = \{f_1, f_2, \ldots, f_L\} \) represent a feature vector for a candidate \( i \), with \( L \) being the number of features, where each feature is a first-order statistic of the scalar values of volume \( I_k \).
2.6. Classification

In the classification step, candidate regions are classified into malignant or benign using a two-stage classification approach. This approach removes spurious candidates from the data in the first stage by estimating a coarse decision boundary using only a subset of features. After spurious candidate removal, the decision boundary is refined in the second stage taking into account the complete set of features. The proposed two-stage classification approach avoids that the final estimation of the classification boundary is driven by spurious and/or outlying data.

For the first stage, the two most discriminant features according to the Fishers discriminant (FD) ratio (Jobson 1992) were independently selected and a LDA classifier (Friedman 1989) was trained to remove spurious candidates from the data. Note that the FD ratio analyses the individual discrimination power of the features without taking into account the rest of the features. This provides a fast determination of a coarse classification boundary. Those candidates that lie far away from the obtained decision boundary (i.e. their posterior probability is higher than a specific threshold) are removed and the threshold is set such that no true positives (TPs) are lost in the first phase for the training set. Before performing the feature selection, the feature values were transformed using the Box–Cox transformation (Box and Cox 1964) in order to approximate the feature distribution to a normal distribution.

In the second stage, a classifier was trained using a selection of features from the complete feature set $\theta$. After pilot experiments with several classifiers, a LDA classifier yielded the best results and was therefore chosen in favour of $k$-nearest neighbour (Cover and Hart 1967) classifier, quadratic discriminant analysis classifier (Friedman 1989) and support vector machines (Chang and Lin 2001). The selection of features was carried out by sequential forward floating selection (SFFS) (Pudil et al 1994) to establish the most discriminant features. The SFFS procedure uses leave-one-out training and testing with the area under the receiver operating characteristic (ROC) curve as the criterion to be optimized. Table 2 summarizes the selected features in order of selection.

3. Data and experiments

3.1. Data

Imaging data were used from a cohort of clinical patients scheduled between January and December 2009 at the RUNMC radiology department. These patients had elevated PSA levels and one negative biopsy. Images were acquired with a 3.0T whole-body MR scanner (TrioTim, Siemens Medical Solutions, Erlangen, Germany). The machine body and a pelvic phased-array surface coil were used for RF transmitting and receiving, respectively. An amount of 1 mg of glucagon (Glucagon®, Novo Nordisk, Bagsvaerd, Denmark)) was administered directly before the MRI scan to all patients to reduce peristaltic bowel movement during the examination.

For our experiments, the MR data from each patient comprised a T2-weighted axial volume (with dimensions $256 \times 256 \times 15$ and voxel size $0.75 \text{ mm} \times 0.75 \text{ mm} \times 4 \text{ mm}$), a proton density-weighted volume (with dimensions $128 \times 128 \times 12$ and voxel size $1.8 \text{ mm} \times 1.8 \text{ mm} \times 4 \text{ mm}$), contrast enhanced 3D T1-w spoiled gradient echo images and an ADC map (with dimensions $136 \times 160 \times 10$ and voxel size $1.625 \text{ mm} \times 1.625 \text{ mm} \times 3.6 \text{ mm}$) that was generated by the MR scanner. Additionally, gadolinium chelate concentration curves were calculated and dynamic contrast enhanced derived parameter maps ($K^{trans}$, $V_e$, wash-out) were generated at a dedicated workstation (Vos et al 2008). All the MR data were automatically
normalized such that quantitative assessment was possible for the pharmacokinetic, T1 and T2 map using a previously presented method (Vos et al., 2009, 2010).

The reference standard for lesion localization and pathology was established by combining the findings of an experienced prostate radiologist with, when available, histopathology of MR-guided MR-biopsy samples. The workflow was as follows: firstly, the radiologist screened the MR examination for PCa. When no evidence of PCa could be found, the patient was considered healthy. If a repeat study was advised and performed, the result was used to establish the reference standard. Secondly, all locations that were considered malignant by the radiologist were recorded in a local database. At those locations, a biopsy was performed after which histopathology established the true nature of the finding. The pathologist was blinded to the imaging results. The annotated tumour regions were labelled as low grade when histopathology confirmed a tumour of Gleason grade less than 6. Regions were labelled as high-grade tumours when histopathology confirmed a tumour of Gleason grade if 7 or more. When the Gleason score was 6 or no histopathology was available and the radiologist indicated the presence of tumour, the region was labelled as intermediate.

Our principal interest is the detection of malignant abnormalities. Benign abnormalities were not annotated, and therefore will induce a number of FP signals.

3.2. Experiments

Free response operating characteristic (FROC) methodology was performed to evaluate the detection accuracy of the CAD system. The detection performance of the CAD system was estimated by a threefold cross validation in which a fold was used to train both stages separately. Cross-validation folds were obtained by randomly drawing whole patient cases.

A tumour was considered as detected when the detection location was inside the reference standard. If multiple detections were found inside the same reference standard region, they were recorded as a single hit. Candidates outside the reference standard were counted as FPs. The specificity was computed using detected tumour locations in those patients where no presence of PCa was found based on radiology reports, follow-up reports and MR-biopsy outcome. In this way, a FP rate is obtained that is representative for a screening population, where the majority of the patients will be normal. Another reason only to use non-cancerous patients is that prostate cancer is often multifocal and may incorrectly induce FP signals when not all tumour areas are carefully annotated or if they are missed.

Three experiments were performed to evaluate the detection performance of the CAD method. Firstly, it was determined which size of the spherical region $R$ provides the optimal detection performance by varying the radius parameter $r$. In the second experiment, the results of the proposed approach were compared to the results obtained using only feature $f_7$ for classification, i.e. the peak value obtained after the initial voxel classification stage. In that way, we analyse the improvement obtained by adding additional information from multiparametric MR images. The difference in performance between the initial voxel classification stage and the local feature analysis was evaluated by jackknife FROC analysis (Chakraborty 2006).

In the third experiment, the detection performance for prostate cancer was analysed taking into account the malignancy grade. Therefore, the detection performance was evaluated for all tumours, high-grade tumours, intermediate-grade tumours and low-grade tumours. The results were compared to the performance of a method that detects malignant regions based on complete random guess (random detection based on the average prostate size and tumour size). All detection performances were evaluated at FP levels of 1, 2 and 5 per healthy patient.
Table 1. Discriminating performance $Az$ of the CAD system for different radius $r$ of region $R$.

<table>
<thead>
<tr>
<th>$r$ (mm)</th>
<th>$Az$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.786 ± 0.117</td>
</tr>
<tr>
<td>5</td>
<td>0.833 ± 0.052</td>
</tr>
<tr>
<td>7</td>
<td>0.832 ± 0.065</td>
</tr>
<tr>
<td>9</td>
<td>0.799 ± 0.120</td>
</tr>
</tbody>
</table>

Table 2. The features and their descriptions that were selected in the two-stage classification approach in order of selection.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_7$</td>
<td>Peak value</td>
</tr>
<tr>
<td>$f_5$</td>
<td>25% percentile wash-out</td>
</tr>
<tr>
<td>$f_2$</td>
<td>25% percentile ADC</td>
</tr>
<tr>
<td>$f_3$</td>
<td>75% $K_{\text{trans}}$</td>
</tr>
<tr>
<td>$f_1$</td>
<td>75% T2</td>
</tr>
</tbody>
</table>

4. Results

4.1. PSA levels and histopathological findings

The study set consisted of 200 consecutive patients with the mean age 60 (range 50–69) years. The mean PSA level was 13.6 (range 1–58) ng mL$^{-1}$ and the mean Gleason score was 7.3 (range 5–9). MRI was performed on average three weeks after the transrectal-ultrasonographically-guided sextant biopsy of the prostate. From those patients, 23 had to be excluded due to failed calculation of the ADC map (2), missing or incomplete DCE data (11) or because those patients were scanned for staging (3), post-therapy evaluation (1) or recurrence detection (6).

In the resulting 177 patients, the radiologist annotated 48 locations of prostate cancer in 41 patients. In the 41 patients, biopsy confirmed five low-grade, five intermediate and 15 high-grade tumours. Additionally, the radiologist identified 23 patients with prostate cancer that did not undergo a biopsy. Those prostate cancers were graded as intermediate. The prostate volume of all patients was measured by the radiologist and was on average 67.5cc (±33.8). The average tumour volume used as reference standard was 2.78cc (±3.9).

4.2. CAD performance

The candidate detection step generated 6227 candidates of which 44 were TPs, resulting in a sensitivity of 92%. The maximum performance in the second-stage classification is corrected to the sensitivity of 92%. The 6227 candidates were used for training and evaluation of the two-stage classification approach by a threefold cross validation.

The results of the first experiment are shown in table 1. It can be observed that using a spherical region $R$ with radius 5 results in the optimal discriminating performance of the classifier. Although the obtained performances are not statistically different, a radius of 5 was used in the remaining experiments.

The first stage of the two-stage classification approach removed 37.2% of the candidates at the expense of eliminating two TPs, after selecting the features $f_9$ and $f_2$. The selected features in the second stage of the two-stage classification approach are summarized in order of selection in table 2.
Figure 5 shows the detection results for the second experiment. The results demonstrate that the additional local feature analysis leads to a significantly improved detection performance compared to voxel classification and random detection ($p < 0.05$).

The detection performances for the different tumour grades were measured at the FP levels of 1, 3 and 5 per patient. At these levels, the CAD method obtained the sensitivities of 0.48, 0.73 and 0.88, respectively, for the detection of high-grade tumours. The sensitivities for detecting all malignant regions were 0.41, 0.65 and 0.74, respectively. Detecting intermediate-grade tumour resulted in sensitivities of 0.41, 0.65 and 0.74. Low-grade tumours were detected with sensitivities of 0.27, 0.40 and 0.68, respectively. Random detection of prostate cancer based on prostate volume and tumour volume resulted in sensitivities of 0.041, 0.12 and 0.21. Figure 6 demonstrates the detection results of the different data sets that were obtained by FROC analysis.

5. Discussion

In this paper, we have presented a novel fully automatic CAD method and applied it to a cohort of prostate MRI data acquired of men with one negative biopsy. This study showed that it is feasible to automatically detect PCa using information from multiple MR images simultaneously using a two-stage classification system. The CAD system is able to detect 74% of all tumours at a FP level of 5 per patient. When focusing only on high-grade tumours, a sensitivity of 88% was obtained at a FP level of 5 per patient, which is a strong improvement over systematic biopsy being known to understage half of the tumours. Furthermore, the CAD system is able to detect tumours in all zones of the prostate.

The proposed method achieved a good performance for the detection of prostate cancer in a challenging cohort of patient MR data. The MR data were acquired during screening for prostate cancer for patients with elevated PSA levels. As a result, the patient database consisted
Figure 6. FROC curves showing the detection performances of the CAD method for different tumour grades. The horizontal axis shows the number of FP detections per healthy patient and the vertical axis shows the sensitivity that is achieved at this specificity level. The (blue) solid FROC curve corresponds to the detection performance of high-grade tumours. The (red) dashed FROC curve corresponds to the detection performance of intermediate-grade tumours. The (green) dotted FROC curve corresponds to the detection performance of all tumours. The (purple) dot-dashed FROC curve corresponds to the detection performance of low-grade tumours. The (cyan) big-dashed FROC curve corresponds to the detection performance when a random detection is performed taking into account the average tumour and prostate size.

of both healthy and prostate cancer patients. In figure 6, it can be observed that high-grade tumours are detected with a sensitivity of 0.734 at only a FP rate of 2.3 per patient. Detecting only the high-grade tumour can potentially help in preventing the inclusion of patients that have slow-growing and not aggressive prostate cancer. When detecting tumour of all grades, the sensitivity was 0.60. Although intermediate-grade tumours are considered to be more difficult to detect, the obtained performance was similar to the result of detecting all tumours. However, it is difficult to differentiate between high- and intermediate-grade tumours. Low-grade tumours are more difficult to detect. The results show a sensitivity of 40% with 2.3 FPs. Nevertheless, the achieved performance is still considerably better than randomly detecting PCa.

In figure 5, it is demonstrated that the usage of information from multiple MR images has a benefit to the detection performance of the CAD system. The results demonstrate that the multi-stage approach performs significantly better than only using feature $f_7$ for classification, i.e. the peak value obtained after the initial voxel classification stage. A significant improvement was obtained by adding additional information from multiple MR images ($p < 0.05$). Furthermore, the two-stage classification approach removed 37.2% of the number of candidates after the first stage. This showed to be an important stage to mitigate the effect on the training of the classifier.

Only four candidates remained undetected in the initial stage. Nevertheless, the undetected cases can be considered less clinically relevant. In the first undetected case, histopathology found only a 5% volume of cancer with Gleason 3+3 in the biopsy sample and the radiologist indicated a normal diffusion. In two patients, there were two regions annotated by the radiologist. In both cases, the most dominant tumour was detected, while the other location was missed due to a high diffusion. In the fourth undetected case, the radiologist identified the location of PCa in both the current and follow-up examinations. However, the lesion was
graded less significant due to a high diffusion. Furthermore, no biopsy was performed such that the reference standard could not truly be established.

The two-stage classification approach removed 37.2% of the candidates in its initial stage. The approach avoided that the final estimation of the classification boundary was driven by the spurious and/or outlying data. Experience with a single classification approach showed that the classifier is indeed performing less. However, two TPs were additionally eliminated. One TP was graded intermediate, but this was not confirmed by biopsy. Regarding the second TP, biopsy confirmed a prostate cancer with Gleason score 3+4. Both locations were discarded because they appeared to have a high diffusion. Additionally, the peripheral zone appeared to have an abnormal high diffusion. This may suggest the need of a normalization step and is part of further research.

This is the first study that analysed the detection performance using FROC methodology. Most studies presented in the literature evaluated a discriminating performance of malignant and benign voxels using ROC analysis (Chan et al. 2003, Viswanath et al. 2009, Langer et al. 2009, Liu et al. 2009, Ozer et al. 2010, Lopes et al. 2011). ROC analysis, however, misses information about the number of FP candidates. Therefore, the method may have a high discriminative performance though it presents many FP candidates. This can have a negative influence on the detection performance of the radiologist. Furthermore, the studies presented in the literature generally use only cancer patients and therefore do not reflect a screening population. This study, however, was performed on a cohort of patients that had elevated PSA and an initial set of negative prostate biopsies. As a result, it better reflects a screening population as both benign and malignant patients were present in the database.

Noguchi et al. (2001) demonstrated that the grade assessment with needle biopsy underestimated the tumour grade in 46% cases and overestimated it in 39 (18%) and as a result, no single parameter in the biopsy was a predictor of tumour significance. Hence, the gold standard for detecting PCa, systematic biopsy, lacks sensitivity as well as grading accuracy. Multiparametric MRI has a potential to guide prostate biopsy towards the most aggressive and representative part of the tumour (Hambrock et al. 2010). However, its clinical application is limited due to the required high level of experience of the radiologist. Moreover, it is a difficult and time-consuming procedure to localize the most aggressive part of the tumour. The presented CAD method has the potential to assist radiologists in the detection of prostate cancer and to guide prostate biopsy towards the most aggressive and representative part of the tumour. Therefore, the CAD method could improve the sensitivity of MR-guided TRUS biopsy without introducing many additional biopsies and is part of further research.

A limitation of this study was that the reference standard for some patients could not be accurately established, as follow-up studies were or biopsy was not yet performed. The reference standard for those patients was graded as intermediate which in fact could be different. Also, the patient data used to represent a screening population are somewhat biased, as they were scheduled for an MRI examination after one negative biopsy session.

To conclude, this study demonstrated that it is feasible to fully automatically detect locations of prostate cancer at an acceptable FP rate better than random detection by e.g. systematic biopsy. The CAD method may assist the radiologist to detect prostate cancer locations and could potentially reduce the number of biopsies.

Acknowledgments

This work was funded by grant KUN 2004-3141 of the Dutch Cancer Society.
References


Beigelman-Aubry C, Hill C, Boulander X, Brun Al, Leclercq D, Golmard JJ, Grenier Pa and Lucidarme O 2009 Evaluation of a computer aided detection system for lung nodules with ground glass opacity component on multidetector-row CT J. Radiol. 90 1843–9


Djavan B et al 2001 Prospective evaluation of prostate cancer detected on biopsies 1, 2, 3 and 4: when should we stop? J. Urol. 166 1679–83


Friedman J H 1989 Regularized discriminant analysis J. Am. Stat. Assoc. 84 165–75


Hambrick T et al 2010 Magnetic resonance imaging guided prostate biopsy in men with repeat negative biopsies and increased prostate specific antigen J. Urol. 183 520–7


Hogeweg L, Mol C, de Jong P A, Ayles H, Dawson R and van Ginneken B 2010 Evaluation of a computer aided detection system for tuberculosis on chest radiographs in a high-burden setting Annual Meeting of
Automatic CAD of PCs based on multiparametric MRI analysis

1541

the Radiological Society of North America SSA19-01 (Available at http://rsna2010.rsna.org/search/search.cfm?action=add&filter=Author&value=84463)


Jobson J D 1992 Applied Multivariate Data Analysis (Berlin: Springer)

Karssemeijer N, Otten J D M, Rijken H and Holland R 2006 Computer aided detection of masses in mammograms as decision support Br. J. Radiol. 79 S123–6


Noguchi M, Stamey T A, McNeal J E and Yemoto C M 2001 Relationship between systematic biopsies and histological features of 222 radical prostatectomy specimens: lack of prediction of tumor significance for men with nonpalpable prostate cancer J. Urol. 166 104–9


Qiang L, Sone S and Doi K 2003 Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans Med. Phys. 30 2040–51


Summers R M et al 2010 CT colonography computer-aided polyp detection: effect on radiologist observers of polyp identification by CAD on both the sapine and prone scans Acad. Radiol. 17 948–59


