The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/91387

Please be advised that this information was generated on 2017-09-01 and may be subject to change.
Computer Aided Diagnosis of Prostate Cancer with Magnetic Resonance Imaging

Pieter Vos
COMPUTER AIDED DIAGNOSIS OF PROSTATE CANCER WITH MAGNETIC RESONANCE IMAGING

EEN WETENSCHAPPELIJKE PROEVE OP HET GEBIED VAN DE MEDISCHE WETENSCHAPPEN

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S. C. J. J. Kortmann,
volgens besluit van het college van decanen
in het openbaar te verdedigen op donderdag 8 december 2011
om 13:00 uur precies

door

PIETER CHRISTIAAN VOS
geboren op 30 september 2011
te Nijmegen
Promotoren: 
Prof. dr. ir. N. Karssemeijer
Prof. dr. J.O. Barentsz

Copromotor: 
Dr. ir. H.J. Huisman

Manuscriptcommissie: 
Prof. dr. J.A. Witjes
Prof. dr. L.A.L.M. Kiemene
Dr. J.P.W. Pluim (University Medical Center Utrecht)

The research described in this thesis was carried out at the Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Centre (The Netherlands).

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

Financial support for publication of this thesis was kindly provided by the Radboud Universiteit Nijmegen.
Chapter 1

Introduction and Outline

1.1 Motivation

Prostate cancer is the most commonly diagnosed cancer among men and remains the second leading cause of cancer death in men. In 2010, more than 217,000 men in the United States (US) were diagnosed with prostate cancer [1]. The American Cancer Society estimated that approximately 32,000 men died from the disease in the US in 2010. In Europe, more than 338,000 males were diagnosed with prostate cancer in 2008 and almost 71,000 men died because of prostate cancer [2]. The growth of the population and, more importantly, the aging population is a major cause of the high number of prostate cancer cases and will contribute to an increase in cancer burden. For that reason, there is an ongoing debate whether screening for prostate cancer should be performed.

Screening can help find cancers in an early stage when they are more easily cured. An important trial to determine the effect of screening for breast cancer was performed between 1977 and 1984 in Sweden [3]. The trial showed that after seven years of follow up a reduction of 31% in breast cancer mortality was achieved when screening was applied. This led to the introduction of breast cancer screening in most western countries. Recently, several studies have been performed that looked at whether prostate cancer screening with the prostate-specific antigen (PSA) blood test saves lives [4, 5, 6, 7]. For example, the European Randomized Study of Screening for Prostate Cancer (ERSPC) has shown significant reductions in PCA mortality in an intention-to-screen analysis [8, 9]. The reduction in mortality comes, however, at the price of over-diagnosis and over-treatment. In the study of Schröder et al. the authors specifically warn that, in order to prevent one death from prostate cancer, 1410 men would need to be screened and 48 additional cases of prostate cancer would need to be treated. Hence, controversy still exists regarding the effectiveness of prostate cancer screening. The ongoing debate is essentially a public demand for a more reliable, non-invasive method that has a sufficiently high specificity in detecting prostate cancer [10].

Magnetic resonance imaging (MRI) has evolved this decade to a competitive imaging modality for the localization of prostate cancer. The non-invasive nature and ability to provide structural, functional and metabolic information in a single examination makes the technique suitable to improve specificity when screening for prostate cancer. Many studies showed that multiparametric MRI, consisting of high resolution 3D T2-weighted sequences, 3D dynamic contrast enhanced MRI, 3D diffusion weighted imaging or spectroscopic imaging, leads to a sufficiently high accuracy for prostate cancer detection [11, 12, 13, 14, 15]. Unfortunately, multiparametric MRI analysis requires a high level of expertise, suffers from observer variability and is a labor intensive procedure [16]. For that reason the technique is considered cost inefficient and, as a result, has not been implemented in a screening environment [17].

Computer aided diagnosis (CAD) can be of benefit to improve the consistency and accuracy of interpreting radiographic images by the radiologist. Additionally, it can speed up the reading time considerably. CAD research has been successfully pursued in other diagnostic areas such as mammography [18, 19], CT chest [20, 21, 22], CT colonography [23, 24] as well as retinal imaging [25]. However, published literature on prostate CAD research is still relatively immature.

The motivation of this thesis was therefore to research state of the art CAD methods that can assist in a better diagnosis of prostate cancer, reduce the observer variability and be of benefit to a more efficient workflow for the radiologist.
1.2 Prostate Anatomy

The prostate is a relatively small organ in the pelvis and is part of the male reproductive system. It is located between the pelvic bones, in front of the rectum and below the bladder. In healthy males, it is about the size of a chestnut and is somewhat conical in shape. Urine that is collected in the bladder goes from the bladder neck into the prostate through the urethra. The amount of urine that goes through the urethra is regulated by the urethral sphincter. The prostate has a role in the normal sexual functioning. That is, seminal fluid is made by and stored in the prostate and is mixed with sperm, which is carried out of the body during ejaculation. Development of the prostate is induced by testosterone, a male hormone made in the testicles. Fig. 1.1 shows a schematic drawing of the pelvis and its different structures.

![Figure 1.1: A schematic drawing of the male pelvis in which several anatomical structures are indicated](source: US government agency National Cancer Institute).

The prostate is divided into an apex, mid and base part, where the apex refers to the lower part and the base to the upper part of the prostate. Different functional parts of the prostate are addressed as: the peripheral zone, transition zone and (compressed) central gland. The three parts are schematically shown in the standardised magnetic resonance imaging (MRI) prostate reporting scheme Fig. 1.2 (source: the European Consensus Meeting [26]). The central gland is a cone shaped region that surrounds the ejaculatory ducts, extends from bladder base to the verumontanum and comprises 25% of glandular tissue in young adults. The transition zone surrounds the urethra and typically, when aging, pushes the central gland away as a result of benign prostatic hyperplasia (BPH). The peripheral zone is located posterolateral and comprises the majority of prostatic glandular tissue.

1.3 Prostate Cancer Diagnosis in the Clinic

Prostate cancer is generally detected through PSA testing or digital rectal examination (DRE). In a DRE, the physician inserts a lubricated, gloved finger into the rectum to feel the surface of the prostate. Healthy prostate tissue is soft, whereas malignant tissue is firm, hard, and often asymmetrical or stony. Of all prostate tumors, 65% to 74% are located in the peripheral zone [11]. The remaining tumors that are more ventrally located cannot be reached by DRE and are therefore missed. Additionally, small tumors are difficult to detect by DRE [27]. For that reason, the sensitivity of detecting prostate cancer by DRE is rather low. Although it has been suggested that DRE can be replaced by PSA testing only [28], urologists tend to be less controversial and state that DRE should only be replaced by a better regimen if it detects a larger proportion of cancers with fewer biopsy examinations [29].
Figure 1.2: Schematic drawing of the prostate with three zones. On the left, three axial slices are shown of the prostate. On the right, a sagittal projection of the prostate is shown. The white sections represent the peripheral zone and light grey the transition zone and the dark grey sections the fibromuscular stroma. Most patients have an enlarged transition zone (BPH) such that the central gland is hardly visible. The central gland is therefore not added to the scheme [26].

PSA testing usually detects prostate cancer earlier than it would be detected by a DRE or the development of symptoms [28]. In clinical practice though, PSA testing and DRE are performed alongside. A PSA test measures the amount of antigen in the blood in nanograms per milliliter (ng/mL) that is specific for the prostate. Elevated levels of PSA may indicate presence of cancer in the prostate. The American Cancer Society guideline for the early detection of PCa recommends that patients with a PSA level of 4.0 ng/mL or greater should be referred for further evaluation or biopsy. To increase the sensitivity of the test, many other countries lowered their threshold to 3.0 ng/mL taking a significant increase of benign cases into account [30]. The poor specificity of PSA testing is caused by the fact that elevated levels are also measured at benign conditions such as prostatitis or benign prostatic hyperplasia. Furthermore, some men with PCa do not have elevated PSA levels. As a result, PSA testing may produce false-positive or false-negative findings such that men without prostate cancer receive unnecessary additional testing or clinically significant cancers are missed.

Definitive diagnoses of PCa is most often established through transrectal ultrasound (TRUS)-guided systematic biopsy, as it is the standard procedure for histological sampling. The painful procedure, which is somewhat prevented through local anesthesia, involves systematic targeting of core biopsies under ultrasound guidance. The European Guideline on Prostate Cancer advises that at least eight cores should be sampled, though recommends to increase to higher sampling rates to improve the sensitivity of the technique. The prostate tissue samples are evaluated by the pathologist to determine whether cancer is present and, if so, the Gleason score of the cancer. The Gleason score is based on the degree of abnormality of the cells and ranges from 2 to 10. Knowledge of the tumor grade is essential for the choice of therapy, which will be discussed below. Although TRUS-guided biopsy is sensitive to PCa located in the peripheral zone, it can hardly detect PCa in the central gland and transition zone [31]. More importantly, Noguchi et al. [32] demonstrated that grade assessment with needle biopsy underestimated the tumor grade in 46% cases and overestimated it in 39 (18%). Therefore, the technique cannot guarantee that the most descriptive part of the tumor has been sampled, or whether it has spread beyond the prostate boundaries.

The choice of therapy for men diagnosed with PCa depends on the Gleason score and the stage of the
tumor, where the stage of the tumor categorizes the risk of cancer having spread beyond the prostate. Mostly, the options of treatment are determined using nomograms of which the Partin tables is most commonly used [33]. The Partin tables estimate the chance of organ-confined disease, capsular penetration, seminal vesicle invasion and lymph node metastasis, based on the result of DRE, biopsy Gleason score and serum PSA level [34]. Higher Gleason scores indicate larger differences from normal tissue and more aggressive disease. Gleason scores of 2 to 4 resemble well differentiated or low grade tumors. Cancers with Gleason scores of 5 to 7 are called moderately differentiated or intermediate grade. Cancers with Gleason scores of 8 to 10 are called poorly differentiated or high grade.

Several options of treatment are available: active surveillance, radical prostatectomy, radiotherapy, and focal therapy. Active surveillance is provided to men that have a small, localised, well-differentiated PCa. It involves a conservative monitoring of the tumor and not to treat the patient immediately, though the urologist can intervene when the cancer progresses above pre-defined threshold, such as short PSA doubling time or deteriorating histopathological factors on repeat biopsy. With a radical prostatectomy, the prostate is removed surgically. This can also include removal of lymph nodes in case of metastases. The neurovascular bundle should be free of tumor tissue to enable a nerve-sparing surgery. In this way, the normal sexual function and ability to urinate can be preserved. However, an accurate localization of the tumor is of high importance to be able to perform nerve-sparing surgery. Another option would be to perform radiation therapy, either with external beam radiation or internal radiation (brachytherapy). In external beam radiation, the patient receives radiation treatment from an external source, usually over an 8- to 9-week period. Brachytherapy involves placing small radioactive pellets, sometimes referred to as seeds, into the prostate tissue and is recommended for low-risk cancers. Recently, there is a shift towards minimally invasive focal therapies such as delivering a boost dose to the dominant intra-prostatic lesion, cryotherapy (ablation of prostate tissue by local induction of extremely cold temperatures) or lasertheraphy [35]. It stands to reason that an accurate localization, grading and staging of the PCa is of paramount importance before these therapies can be performed.

1.4 Prostate Magnetic Resonance Imaging

With Magnetic Resonance Imaging (MRI), structural, functional and metabolic information can be non-invasively obtained from the patient. A typical prostate multiparametric MR examination consists of three-directional T2-weighted (T2-w) MR imaging, diffusion weighted imaging (DWI), dynamic contrast enhanced (DCE) MR imaging and, in case of staging, spectroscopic imaging. The examination is performed within 30 to 60 minutes depending on the amount of sequences used. Fig. 1.3 shows a screen capture of the in-house developed system (MRCAD) that is used in the clinic to process and visualize the multiparametric MR data for diagnosis of PCa patients. Appendix A provides a detailed explanation of the MRCAD system.

T2-weighted MRI has been used to diagnose PCa for quite some years. T2-weighted MRI is most often performed in multiple views, i.e., transversal, coronal and sagittal view, with a high in-plane resolution and a relatively thick slice distance. In a T2-weighted image, PCa often appears as an area of low signal intensity in a bright normal peripheral zone. However, benign conditions such as biopsy haemorrhage, prostatitis, BPH and effect of treatment can mimic the presence of cancer. As a result, the accuracy of tumor localization using T2-weighted MRI is rather low, ranging from 67% to 72% [11]. Furthermore, a correct interpretation of a T2-weighted image requires a high level of expertise [36].

DCE-MRI is a minimal invasive technique which can be used to study tissue perfusion and vascular permeability. Due to the high vascularity, increased capillary permeability as well as interstitial hypertension in tumors, DCE-MRI shows better distinction between malignant lesions and normal tissue compared to T2-weighted MRI alone [37, 38, 39, 40, 41, 42, 43, 44]. The principles of DCE-MRI lie in the analysis of signal-time or kinetic curves at a specific location in T1-weighted images. A sequential set of T1-weighted images is acquired before and during an intravenous bolus injection of paramagnetic gadolinium chelate, preferably by using a power injector. The contrast agent will induce an increased signal intensity on a T1-weighted image at vessel lumen and interstitial space. The kinetic curves are summarized into a set of kinetic parameters. The derived kinetic parameter maps are often displayed as color-coded transparent overlays on top of anatomical images. This prevents the need to manually analyze each curve individually. An example signal-time curve with several descriptive parameters is demonstrated in figure 1.4. The kinetic parameters are used to characterize lesions. A typical malignant lesion shows a faster initial rise,
Figure 1.3: Screen capture of the in-house developed system (MRCAD) displaying multiparametric MR data using a prostate detection hanging protocol. Cross-sectional transversal views of a T1-weighted (top-left) volume, T2-weighted (right) volume, ADC map (top-middle) as well as sagittal (bottom-left) and a coronal (bottom-middle) view of T2-weighted volumes. Additionally, a perfusion map is displayed as a color-coded overlay on top of the T1-weighted volume. The MRCAD system is explained in detail in Appendix A.

Figure 1.4: Example kinetic curve and derived map of DCE-MRI

has a higher peak enhancement and shows a more negative LateWash or Wash-out when compared to a benign tissue. As the kinetic parameters do not directly correlate to physiological parameters, they cannot be compared among patients. Differences in scan parameter settings, such as repetition time and flip angle, unknown native T1 relaxation time of the tissue, presence of coil profile and differences in the patient’s systemic blood circulation, make the technique difficult to reproduce among clinical centers [45].

Pharmacokinetic (PK) modeling aims at a more quantitative analysis of DCE-MRI. PK modeling removes the above mentioned dependencies such that the derived PK parameters only reflect local tissue properties. The PK parameters have the advantage of being biologically meaningful and help to establish objective criteria for classifying lesions [46]. Fig. 1.5 demonstrates the 2-compartment model that is used to
Chapter 1

represent the PK parameters. Multiple studies have shown the benefit of using PK parameters as additional information to the traditional T2-w images to diagnose PCa [11, 12, 13, 14, 15]. However, the specificity of PK analysis is tempered by post-biopsy hemorrhage, prostatitis and BPH as they mimic PCa enhancement patterns. Furthermore, PK analysis requires dedicated software which is not available in every imaging center.

![Diagram of the 2-compartment model used to represent the pharmacokinetic parameters $K^{\text{trans}}$ and interstitial volume $V_e$. The red area is the blood volume with blood plasma and blood cells (dark blobs). The yellow area represents the tissue volume containing interstitial space and tissue cells (dark blobs). The contrast cannot enter the cells but flows through the blood plasma around the cells to pass through interendothelial fenestrations and junctions into the interstitial compartment (which relates to permeability of the vessels).](image)

Figure 1.5: Schematic drawing of the 2-compartment model used to represent the pharmacokinetic parameters $K^{\text{trans}}$ and interstitial volume $V_e$. The red area is the blood volume with blood plasma and blood cells (dark blobs). The yellow area represents the tissue volume containing interstitial space and tissue cells (dark blobs). The contrast cannot enter the cells but flows through the blood plasma around the cells to pass through interendothelial fenestrations and junctions into the interstitial compartment (which relates to permeability of the vessels).

Diffusion-weighted MRI (DWI) is a technique that lately attracted much attention. It is a non-invasive functional imaging technique that quantifies random Brownian motion properties of water molecules (diffusion) in tissue. The degree of restriction to water diffusion in biologic tissue is inversely correlated to tissue cellularity and the integrity of cell membranes [47]. The net displacement of molecules is called the apparent diffusion coefficient (ADC). In malignant prostate tissue, the cellular density is increased which results in a reduced water diffusion and is represented by a decreased ADC value.

Three-dimensional proton MR spectroscopic imaging enables detection of the metabolites choline, creatine and citrate. The choline plus creatine to citrate (Cho+Cr/Cit) ratio is suggested as marker for prostate cancers, since a decrease in citrate and an increase in choline levels is observed in malignant prostate tissue [48].

Multiparametric MRI combines the before mentioned imaging techniques such that anatomical, functional and metabolic information can be obtained within a single examination. It has been demonstrated that multiparametric MRI outperforms each independent technique when detecting PCa [11, 12, 13, 14, 15]. This concept was also demonstrated in a study by Füttnerer et al. [11], see Fig. 1.6(a). Nevertheless, Füttnerer et al. also demonstrated in a preceding study that the interpretation of multidimensional information requires a high level of experience, see Fig. 1.6(b). Furthermore, as there are no strict guidelines, the diagnosis is labour intensive and suffers from observer variability [16]. Because of the aforementioned limitations, multiparametric MRI is considered cost inefficient and, as a result, has not been implemented in a screening environment [17].
Introduction and Outline

1.5 Computer Aided Diagnosis

CAD systems can be of benefit to improve the diagnostic accuracy of the radiologist, reduce reader variability, and speed up the reading time. The aim of CAD is 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 [49].

CAD systems generally consist of multiple sequential stages, as illustrated in figure 1.7. In the initial stage, lesion candidates are selected within a likelihood map that was generated by 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 supervised classifier into a malignancy likelihood. The last stage ensures a reduction of the amount of false positives that were localized in the initial stage. The radiologist uses the calculated malignancy likelihood and the location as additional information in order to diagnose the patient.

Most prostate CAD researchers have focused on the initial voxel classification stage [50, 51, 52, 53, 54, 55]. They obtained likelihood maps by combining information from multiparametric MR images using mathematical descriptors. These studies showed on a voxel basis that the discrimination between benign and malignant tissue is feasible with good performances. However, localization of the tumor, the final diagnosis and patient management is left to the radiologist. The task of a computer aided detection system is, however, to localize suspicious lesions and to estimate a malignancy likelihood for each detected lesion. Therefore, the goal of computer aided detection is to reduce search errors, reduce interpretation errors, and reduce variation between and within observers [49]. Computer aided diagnosis systems on the contrary only have a classification task for differential diagnosis of user provided regions.

The challenge to introduce CAD in the clinical workflow with a prostate cancer screening environment is enormous. The lack of standardized sequences and objective quantitative features of PCa are important obstacles for prostate MR CAD to become widely available. Furthermore, to become successful in a clinical environment, the intended CAD system should be fully automated, robust to the large population variation...
Figure 1.7: Dataflow diagram of a typical CAD system for the automatic detection of cancer. In the initial stage, lesion candidates are selected within a likelihood map that was generated by 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 that is presented to the radiologist.

and fast enough for a typical screening production of say 30 to 40 cases a day. As of today, there is no commercial prostate CAD system available that fulfills the mentioned requirements.

The main topic of this thesis is the research and development of a CAD system for the prostate cancer screening environment. Special attention is paid to evaluation of quantitative methods, if they can assist the radiologist in localizing prostate cancer and whether they are applicable to the clinic.

1.6 Outline

The outline of the thesis is as follows. In chapter 2, the feasibility is investigated of a CAD system capable of objectively discriminating PCa from non-malignant disorders located in the peripheral zone of the prostate. DCE-MRI derived features are extracted and summarized by a supervised classifier into a malignancy likelihood. In chapter 3 the CAD scheme is extended by extracting additional features from T2-weighted images in order to improve the discriminating performance of the CAD method. Chapter 4 continues with the research on pharmacokinetic parameters. A fully automatic calibration method is presented for the calculation of pharmacokinetic parameters and it is demonstrated that the discriminating performance of the CAD method is equal to that of a manual calibration method. Chapter 5 describes an automated prostate segmentation method, which is used to reduce the number of false positive cancer candidates in a fully automated prostate cancer detection method, presented in chapter 6. In the concluding chapter 8 a summary is provided as well as a general discussion.
Chapter 2

Computerized analysis of prostate lesions in the peripheral zone using dynamic contrast enhanced MRI

This chapter is based on the manuscript “Computerized analysis of prostate lesions in the peripheral zone using dynamic contrast enhanced MRI.” by Pieter C. Vos , Thomas Hambrock, Christina A. Hulsbergen -van de Kaa, Jurgen J. Futterer, Jelle Barentsz, Henkjan Huisman Published in Medical Physics, vol. 25, no. 4, pp. 621–630, 2008.
Abstract

A novel automated computerized scheme has been developed for determining a likelihood measure of malignancy for cancer suspicious regions in the prostate based on dynamic contrast-enhanced MRI (DCE-MRI) images. Our database consisted of 34 consecutive patients with histologically proven adenocarcinoma in the peripheral zone of the prostate. Both carcinoma and non-malignant tissue were annotated in consensus on MR images by a radiologist and a researcher using whole mount step-section histopathology as standard of reference. The annotations were used as regions of interest (ROI). A feature set comprising pharmacokinetic parameters and a T1 estimate was extracted from the ROIs to train a support vector machine as classifier. The output of the classifier was used as a measure of likelihood of malignancy. Diagnostic performance of the scheme was evaluated using the area under the ROC curve. The diagnostic accuracy obtained for differentiating prostate cancer from non-malignant disorders in the peripheral zone was 0.83 (0.75-0.92). This suggests that it is feasible to develop a CAD system capable of characterizing prostate cancer in the peripheral zone based on DCE-MRI.
2.1 Introduction

It is estimated that one out of ten male cancer deaths in 2007 will be caused by prostate cancer (PCa). Furthermore, with a total of 218,890 cases, PCa is the most common non-cutaneous cancer in the United States [56]. PCa incidence rates continues to increase, although at a slower rate than those reported for the early 1990s and before. This trend may be attributable to increased screening through prostate-specific antigen (PSA) testing as well as the aging of the population. The definitive diagnosis of PCa is most often established through transrectal ultrasound (TRUS)-guided sextant biopsy.

For men diagnosed with prostate cancer, a number of treatment options exist, with differing side effects. The therapeutic options are mostly determined using nomograms of which the Partin tables is most commonly used [33]. The Partin tables estimate the chance of organ-confined disease, capsular penetration, seminal vesicle invasion and lymph node metastasis, based on the result of digital rectal examination, biopsy Gleason score and PSA value [34]. However, these clinical assessments are not accurate in determining the local stage. Elevated PSA levels can be observed in non-malignant disorders such as prostatitis or benign prostatic hyperplasia (BPH). The limitations of sextant biopsy are increasingly recognized, which has provoked interest in multimodal magnetic resonance imaging (MRI) as an alternative method of tumor evaluation [57]. Accurate staging is important for a proper disease management. Curative therapy is only effective in cases of organ confined (surgical candidate) PCa, whereas androgen therapy and/or radiotherapy is more effective in advanced disease. Accurate localization is important for evaluation of the tumor location and the distance to the neurovascular bundle and prostate capsule, to determine if a nerve sparing operation is possible, or assist the planning of intensity-modulated radiotherapy [58, 59, 60].

MRI localization can reduce the number of repeat biopsies, improve the staging performance and guide surgery or radiotherapy. T2-weighted MRI using a pelvic phased-array coil can visualize the prostate including the surrounding anatomy and depict tumor suspicious areas of low signal intensity within a high-intensity peripheral zone. An endorectal coil improves the spatial resolution, resulting in better anatomical visualization which may result in an improved diagnostic accuracy of the localization and staging of PCa [61, 62, 11, 57]. However, in addition to PCa, the differential diagnosis of a low signal intensity area includes post-biopsy hemorrhage, prostatitis, BPH, effect of hormonal or radiation treatment, fibrosis, calcifications, smooth muscle hyperplasia and fibromuscular hyperplasia [63].

Dynamic contrast-enhanced MRI (DCE-MRI) can be used as an additional tool to visualize PCa (neo-) vascularity and interstitial space. Due to the high vascularity, increased capillary permeability as well as interstitial hypertension in tumors, DCE-MRI shows better distinction between malignant lesions and normal tissue compared to conventional MRI alone [37, 38, 39, 40, 41, 42, 43, 44]. Futterer et al. [11] showed that using T2-w images in combination with DCE-MRI for localizing PCa, equal or greater than 0.5 cm$^3$, resulted in an accuracy of 81-91% whereas using T2-w MR images alone resulted in a localizing accuracy of 68%.

Post-biopsy hemorrhage, prostatitis and BPH can all mimic PCa enhancement patterns, thus comprising the specificity of the technique. Another major obstacle to the application of MRI analysis in the routine clinical practice of prostate imaging is the variability of interpretation criteria and absence of interpretation guidelines [57]. Our study aims to increase the objectivity and reproducibility of prostate MRI interpretation by developing a computer aided diagnosis (CAD) system.

The proposed method enables an objective automated quantification and classification of features to discriminate between benign and malignant lesions, and may improve the tumor localization accuracy of the radiologist. In addition to objective analysis, computerized analysis can take full advantage of information across slices in 3D multi-feature data sets which is difficult to assess visually from individual images. CAD has been successfully pursued in other diagnostic areas such as mammography [64, 65], CT chest [66] as well as breast MRI [67, 68]. In the field of the prostate, Chan et al. [50] constructed a summary statistical map of the peripheral zone based on the utility of multi-channel statistical classifiers by combining textural and anatomical features in PCa areas from T2-w images, diffusion weighted images (DWI), proton density maps and T2 maps. Madabhushi et al. [69] generated similar statistical maps based on T2-w images using histological maps as ground truth and showed the additional value of combining features. However, to our knowledge, there has been no reported studies about similar work on PCa using DCE-MRI.

The purpose of this study was to investigate the feasibility of a CAD system capable of objectively discriminating PCa from non-malignant disorders located in the peripheral zone of the prostate. Localizing
PCa in the central gland of the prostate is considered difficult because this area is often affected by BPH, which can have areas of low signal intensity on T2-w images and shows enhancement patterns in DCE-MRI similar to that of PCa. Nevertheless, 65% to 74% of the prostate tumor nodules are located in the peripheral zone and central gland tumors are often less aggressive [11]. The focus of this study is therefore on the peripheral zone of the prostate.
2.2 Methods

The proposed CAD method is based on a typical CAD setup illustrated in figure 2.1 and works as follows: a prostate MRI exam is visualized as described in section 2.2.1. While interpreting the images, the radiologist can delineate a lesion as a region of interest (ROI) in the images, using a method discussed in section 2.2.2. From here the characterization of the ROI is fully automated. The CAD system extracts a relevant feature set from the ROI as explained in section 2.2.3. The extracted set of features is presented to a trained classifier which calculates the malignancy likelihood for the lesion as described in section 2.2.4. Finally, the calculated likelihood is presented to the radiologist to assist in his or her diagnosis. The CAD system was implemented in an open source programming environment The Visualization ToolKit (VTK) using the Tool Command Language (Tcl) and C++.

Figure 2.1: Dataflow diagram of the implemented CAD system. The user defines a region of interest from which features are extracted. The features are used to calculate the likelihood of malignancy in the region of interest.

2.2.1 Volume visualization

Figure 2.2: The CAD program using a dedicated prostate hanging protocol, with in the 3 views on top axial T2-w images as background and pharmacokinetic parameter maps (latewash and $K_{\text{trans}}$) as foreground. The 3 views at the bottom show the sagittal and coronal view as well as the pre-contrast T1-w volume.

The CAD program can visualize multimodal MR volumes $I_k$, where $k = 1 \ldots K$ and $K$ is the number of image volumes. The set of $K$ volumes comprises all the volumes acquired in an MR study plus derived
volumes from the acquired volumes. Examples of acquired volumes are T2-w images and T1-w images. Additionally, descriptive parameter maps derived from DCE T1-w images by means of pharmacokinetic modeling are computed (see appendix 2.7 for a description on pharmacokinetic modeling) [11]. In each view all available volumes can be rendered either as background or as transparent color coded overlays. The cursor is positionable in one of the views with the mouse after which the CAD system will instantly update the location in all views. Although the MR data is obtained in slices, the CAD system visualizes the data as 3D volumes taking all directions into account. Figure 2.2 demonstrates the CAD system with a dedicated prostate hanging protocol as it is used in our clinic for localizing PCa.

2.2.2 Lesion segmentation

A 3D drawing tool has been implemented which allows the user to easily delineate a suspicious lesion in 3D. At the request of the user a 3D sphere shaped ROI is added at the position of the cursor and visualized in all views. It is adjustable in size to fully delineate the suspicious area. The intended use is to adjust the sphere to be large enough to fully include the lesion’s size, as to reduce inter-observer variability (see section 2.2.3).

Let an ROI $S_r$ define a set of N cartesian voxel locations $x_i$ in the MR coordinate system:

$$S_r = \{x_1, x_2, \ldots, x_N\}. \quad (2.1)$$

Let $V_{r,k}$ represent a set of scalar values in image volume $I_k$, identified by $S_r$:

$$V_{r,k} = \{I_k(x_i) | x_i \in S_r\} \quad (2.2)$$

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 segment the same lesion area in $I_{k+1}$, regardless of the image resolution or orientation.

2.2.3 Feature extraction

A reduced feature set $F_r$ is calculated from the scalars values of the available volumes ($V_{r,k}$). Each feature in the feature vector $F_r = \{f_1, f_2, \ldots, f_L\}$, with $L$ the number of features, is a first-order statistic of the scalar values of volume $I_k$. One of these statistics are the 25% or 75% percentile. These percentiles are especially suited for volumes that show an heterogeneous pattern, e.g. the derived volume $K^{trans}$ [70, 71, 72]. This heterogeneity is most common for tumor and differs from normal tissue and benign lesions [73, 74]. The 25% or 75% percentile will differ more from the average value when hotspots are present and will give an estimate of the value in that hotspot, as demonstrated in figure 2.3. This heterogeneity is also recognized by the pathologist (at macro scale). They base a histological grade on the Gleason system, in which the dominant and secondary glandular histological pattern are determined. By segmenting the whole lesion and using percentiles to extract the hotspot, variability among users is reduced. Stoutjesdijk et al. [75] showed that manual selection of the hotspots is the major source of variation in the interpretation of the DCE characteristics of breast MRI lesions. Thus, annotating the whole enhancing region instead of just the hotspot and automatically extracting the features sensitive to hotspots within the region, makes the technique more reproducible. An additional advantage of using percentiles is that it is less sensitive to extreme values.

To do so, $V_{r,k}$ is summarized into a single scalar value $f_{r,k,p}$ by calculating its percentile $p$:

$$H_{r,k}(f_{r,k,p}) = p, \quad (2.3)$$

where $H_{r,k}$ is the cumulative density histogram of the scalar values in $V_{r,k}$.

2.2.4 Classification

The final step of the CAD program is to combine the computed features and to estimate the likelihood of malignancy of the region of interest. The malignancy likelihood $l_r$ is calculated using a trained classifier $\tau$:

$$l_r = \tau(F_r; T), \quad (2.4)$$
Figure 2.3: A prostate cancer case to demonstrate the rationale for using percentiles. Although both regions do not differ much in their median, the tumor shows a wide spectrum of heterogeneity in their dynamic enhancement patterns and has more high values. Subfigure 2.3(a) shows a transverse T2-w view of the prostate with a transparent $K^{\text{trans}}$ color-coded overlay. The bi-lateral enhancement in the peripheral zone are both suggestive of cancer. In subfigure 2.3(c) a histogram distribution of the malignant lesion is shown, located in the right peripheral zone (solid box) and correlated to tumor 1 with Gleason 4+4 in the corresponding histopathology slice shown in subfigure 2.3(b). The 75% percentile is 0.36. Note the wide spectrum. Subfigure 2.3(d) shows a histogram distribution of a suspicious enhancing normal peripheral zone region (dotted box). The 75% percentile is 0.23. Note the narrow spectrum.
where \( T \) is a training set of feature vectors and truth states. Classification was performed using support vector machine (SVM) analysis on the feature set (provided by the statistical package R [76]) [77, 78]. SVMs are currently widely used in similar problems as they can act as a general purpose non-linear classifier. SVMs have been shown to perform well on various datasets of limited size. SVMs map input vectors to a higher dimensional space where a maximal separating hyperplane is constructed by means of a kernel function. For this study the radial basis function kernel \( K(u, v) = \exp(-\gamma \cdot |u - v|^2) \) with parameter \( \gamma = 1/5 \) (5 equals the number of features used) was chosen and the cost of constraints violation (or ‘C’-constant of the regularization term in the Lagrange formulation) was set to 1 [79, 80]. When the classifier has calculated \( l_r \), the user is prompted with the estimate of the likelihood of malignancy as shown in the example in figure 2.7(c).

2.3 Feature description

The following features were extracted from \( S_r \):

50% T1Static: The T1Static parameter is the pre-contrast static value of the T1 estimate of the longitudinal relaxation rate in ms. T1-weighted signals are not ideally suitable for use in quantitative assessment of contrast media concentration. We therefore use dynamic T1 mapping with snapshot FLASH sequences as a direct approach to quantification, as described in Hittmair et al [45]. If a post-biopsy hemorrhage is present, it is clearly visible as a high-intensity area on a T1-w image. The biopsy hemorrhage is often visible as a large homogeneous area, hence the median is used to capture this.

75% Ve: In the extravascular, extracellular space (EES) of normal tissue, pressure is near atmospheric (25mmHg) values, whereas in tumors it may reach 50mmHg 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 therefore considered a very descriptive parameter defined as percentage per unit volume of tissue [81]. Interstitial leakage space at tumor hotspots can be three to five times larger than normal tissue, hence the upper quartile is used to capture these hotspots.

75% kep & K trans: The transfer constant (K trans) and rate constant (kep) both have units 1/min, where K trans relates to permeability surface area. The permeability (or leakiness) surface area 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. In case of a tumor, both K trans and kep often show focal enhancement [73]. The upper quartile captures the presence of hotspots.

25% late wash: The late wash parameter quantifies the slope of the curve after the first wash-in phase. Although it does not directly correlate to physiological parameters, the presence of washout is highly indicative of PCa [39], and therefore used in our clinic as a diagnostic parameter. When capillary permeability is very high, the backflow of contrast medium is also rapid, resulting in a negative late wash following the shape of the plasma concentrations. Because late wash enhancement is often heterogenous, the 25th percentile is used to capture this.

The described pharmacokinetic features were extracted because quantification of kinetic parameters has the advantage of being biologically meaningful and help to establish objective criteria for classifying lesions [46], see appendix 2.7 for a description of how the kinetic features are derived from the raw T1-w images. The feature selection is based on clinical experience, previous work [11] has shown that these features are the most descriptive and are therefore preferred in our clinic. Furthermore, preselecting only five features prevents the classifier from being distracted by either poor performing or irrelevant features (peaking phenomenon) [78].
2.4 Training and Evaluation

2.4.1 Dataset

The study set consisted of 34 consecutive patients that were selected in a previous study of Fütterer et al. [11]. These patients had biopsy-proven PCa and underwent DCE-MR imaging at 1.5-T, complementary to the routine staging MR imaging examination of the prostate. Patients were included (between April 1, 2002, and June 1, 2004) in the study only if they were candidates for radical retropubic prostatectomy within 6 weeks after MR imaging. The study of Fütterer et al. was approved by the institutional review board, and informed consent was obtained from all patients prior to MR imaging. After imaging, all patients underwent radical retropubic prostatectomy. Exclusion criteria were: previous hormonal therapy, lymph nodes positive for metastases at frozen section analysis, contra-indications to MR imaging (e.g., cardiac pacemakers, intracranial clips), contra-indications to endorectal coil insertion (e.g., anorectal surgery, inflammatory bowel disease). The mean prostate specific antigen level was 8 ng/mL (range, 3.2-23.6 ng/mL), mean Gleason score was 6.1 (range, 5-8). MRI was performed on average 3 weeks after transrectal ultrasonographically guided sextant biopsy of the prostate.

2.4.2 MR Acquisition

Images were acquired with a 1.5T whole body MR scanner (Sonata, Siemens Medical Solutions, Erlangen, Germany). A pelvic phased-array coil as well as a balloon-mounted disposable endorectal surface coil (MedRad®, Pittsburgh, PA, USA) was inserted and inflated with approximately 80 cm$^3$ of air, were used for signal receiving. The machine body coil was used for RF transmitting. 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.

The protocol for acquisition consisted first of a localizer and two fast gradient spin-echo measurements for patient and coil positioning. Thereafter high-spatial-resolution T2-weighted fast spin-echo imaging in the axial, sagittal and coronal planes, covering the prostate and seminal vesicles, was performed. The frequency encoding direction was anteroposterior to increase the acquisition speed.

Thirdly, 3D T1-weighted spoiled gradient echo images were acquired before and during an intravenous bolus injection of paramagnetic gadolinium chelate (0.1 mmol/kg, gadopentetate, Magnevist®; Schering, Berlin, Germany) using a power injector (Spectris, Medrad®, Pittsburgh, PA, US) with an injection rate of 2.5 ml/second followed by a 15 ml saline flush. At these settings a 3D volume with ten partitions, covering the whole prostate, was acquired every 2 seconds for 120 seconds. Before contrast injection the same axial 3D T1-weighted gradient echo sequence was used to obtain proton density images and identical positioning to allow calculation of gadolinium chelate concentration curves [45]. See Table 2.1 for the precise specification of the acquisitions. Within 3 weeks of biopsy, there can be postbiopsy artifacts on MRI. This cannot be avoided as we feel it is unethical to unnecessarily delay a scheduled prostatectomy. The optimal timing of post-biopsy MR Imaging of the prostate has been researched by Ikonen et al. [82] and White et al [83]. They advise deferring MR imaging for at least 3 weeks after biopsy.

<table>
<thead>
<tr>
<th>Modality</th>
<th>Imaging order*</th>
<th>TR (msec)</th>
<th>TE (msec)</th>
<th>No. of echoes</th>
<th>No. of signals acquired</th>
<th>Flip angle (deg)</th>
<th>Section thickness (mm)</th>
<th>Matrix</th>
<th>No. of sections</th>
<th>Field of view (mm)</th>
<th>Phase-encoding direction</th>
<th>Dyn volume sampling time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2-w spin-echo</td>
<td>1</td>
<td>3510</td>
<td>132</td>
<td>15</td>
<td>1</td>
<td>180</td>
<td>4</td>
<td>240x512</td>
<td>11-22</td>
<td>280</td>
<td>Row</td>
<td>NA</td>
</tr>
<tr>
<td>Intermediate-w fast 3D</td>
<td>2</td>
<td>800</td>
<td>1.6</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>4</td>
<td>256x77x10</td>
<td>NA</td>
<td>280</td>
<td>Column</td>
<td>NA</td>
</tr>
<tr>
<td>gradient-echo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamic T1-w fast 3D</td>
<td>3</td>
<td>34</td>
<td>1.6</td>
<td>1</td>
<td>1</td>
<td>14</td>
<td>4</td>
<td>256x77x10</td>
<td>NA</td>
<td>280</td>
<td>Column</td>
<td>2</td>
</tr>
<tr>
<td>gradient-echo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*One of each sequence was performed before contrast agent administration. After contrast agent administration, 74 dynamic T1-weighted fast 3D gradient-echo and five dynamic T1-weighted high-resolution 3D gradient-echo MR imaging sequences were performed.
2.4.3 ROI annotation

Histopathological analysis

All patients underwent radical retropubic prostatectomy. The prostatectomy specimens were fixed overnight (10% neutral-buffered formaldehyde) and coated with Indian ink. Axial whole mount step-sections were made at 4-mm intervals in a plane parallel to the axial T2-w images and routinely embedded in paraffin. Tissue sections of 5 μm were prepared and stained with haematoxylin and eosin. An experienced pathologist (C.A.H.K) who was blinded to the imaging results, established malignancy from microscopy. Regions of malignancy were outlined on digital macroscopic whole-mount images from a CCD camera. Figure 2.7(d) shows an example of an histopathological map.

Annotation in the MRI data

The whole-mount step-section histology tumor maps were used as ground truth for training and evaluating the performance of the CAD system. The morphology of the central gland, peripheral zone, cysts, calcifications, and urethra were used as landmarks to find the corresponding MRI slice.

Aligning MR slices to whole-mount step-sections is considered difficult [84], it is subjective and the section thickness used in the MR imaging sequences can be different. To overcome these problems a method was developed that semi-automatically matches MR slices to the step-sections of histopathology. The method has the following setup: one of the views is set to a 3D rendering mode for volumes. In this mode the volume is rendered in 3 planes in all directions. The planes can be manipulated to move through the volume slices. In this 3D view a default 3D ellipsoid is rendered as a transparent surface. The goal is to fit the prostate roughly by interactively resizing and translating the ellipsoid. The cross-sections of the ellipsoid are simultaneously displayed in the 2D views for a more accurate result. The final ellipsoid is then divided in the same number of slices as the prostatectomy specimen was cut. By doing this, the specimen images are aligned to the T2-w images. See figure 2.4 for a demonstration.

Figure 2.4: Example of a prostate segmentation to obtain an objective and more accurate correspondence with histopathology. The left view shows a cross-section at transverse view of the prostate, the ellipse indicates the surface bounds in the T2-w image. The middle (sagittal) view represents the number of transverse slices in which the prostatectomy specimen was cut (9 slices). The right view shows the 3D deformable surface which can be positioned, scaled and stretched manually to fit the prostate roughly.

The anatomy of the prostate is best imaged on T2-w images and were therefore used for correlating the histopathological map. The features used for this experiment, however, were extracted from T1-w images. Because the patient may have moved and no registration is applied to correct for patient movement, the pre-contrast T1-w images were semi-transparently overlaid on the T2-w images, to allow for visual inspection and comparison for anatomic mismatch due to patient related movements. If a mismatch was evident, it was compensated for by correcting the annotation on the pre-contrast T1-w images, thereby avoiding the annotation of periprostatic vasculature and urethra.
Table 2.2: The three classification types of $Q$ that were assigned to the annotated regions.

<table>
<thead>
<tr>
<th>Classification Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$ (normal)</td>
<td>Region of normal enhancement and histopathological analysis showed no evidence for tumor.</td>
</tr>
<tr>
<td>$M$ (malignant)</td>
<td>Histopathology confirmed tumor with a clinical relevant diameter of at least 5mm. Regions with prostatic intraepithelial neoplasia (PIN) were excluded because they are considered to be a precursor of PCa [85].</td>
</tr>
<tr>
<td>$NS$ (non-malignant suspicious enhancing)</td>
<td>Region of non-malignant suspicious irregular heterogenous enhancing areas where the underlying histopathological analysis showed no evidence for tumor. Region with histopathological confirmed prostatitis. Region of post-biopsy hemorrhage without any histopathological evidence for tumor.</td>
</tr>
</tbody>
</table>

An ROI was placed to cover the whole lesion volume based on histopathology. After a thorough inspection of the segmentation, the ROI was saved to disk along with a classification label $N$, $NS$ or $M$. The definition of the labels are given in table 2.2.

For all saved ROIs $S_r$ with one of the assigned labels $N$, $NS$ or $M$, information was summarized by collecting the features $f_{r,Ktrans,75}$, $f_{r,Kep,75}$, $f_{r,Vc,75}$, $f_{r,Washout,25}$ and $f_{r,T1Static,50}$, as described in section 2.3, into the feature vector $F_r$:

$$F_r = \{f_{r,Ktrans,75}, f_{r,Kep,75}, f_{r,Vc,75}, f_{r,Washout,25}, f_{r,T1Static,50}\} \quad (2.5)$$

2.4.4 ROC analysis

The discriminating performance of the CAD system was estimated by means of the area under the receiver operator characteristics (ROC) curve (AUC). Let $\xi = (l_1, l_2, \ldots, l_m)$ be the vector of calculated malignancy likelihoods for $m$ ROIs with the trained classifier $\tau$. The ROIs are split into two groups $\alpha$ and $\beta$. Let $\gamma_{\alpha} = \{j | q_j \in Q_{\alpha}\}$ and $\gamma_{\beta} = \{j | q_j \in Q_{\beta}\}$ be the corresponding vectors of indices, where $Q_{\alpha}$ and $Q_{\beta}$ are disjoint and subsets of $Q$ (see table 2.2 for a definition of the labels). The AUC for the classification performance between two subsets of ROIs identified by $\gamma_{\alpha}$ and $\gamma_{\beta}$ is given by [86]:

$$AUC_{\gamma_{\alpha},\gamma_{\beta}} = \frac{\sum_{j \in \gamma_{\alpha}} \sum_{j' \in \gamma_{\beta}} \psi(l_j, l_{j'})}{n_{\gamma_{\alpha}} n_{\gamma_{\beta}}} \quad (2.6)$$

with kernel function

$$\psi(l_j, l_{j'}) = \begin{cases} 
1 & \text{if } l_j > l_{j'} \\
\frac{1}{2} & \text{if } l_j = l_{j'} \\
0 & \text{if } l_j < l_{j'} 
\end{cases} \quad (2.7)$$

and $n_{\gamma_{\alpha}}$ and $n_{\gamma_{\beta}}$ the number of ROIs in $\gamma_{\alpha}$ and $\gamma_{\beta}$, respectively.

For this experiment two separate classifiers were trained and evaluated for its discriminating performance. The first classifier $\tau_{loc}$ was trained to discriminate regions of type $\{N, NS\}$ from $\{M\}$. This reflects localization, hence the subscript $loc$. The discriminating performance of $\tau_{loc}$ is denoted as $AUC_{loc}$ and is computed using Eq. 2.6 by setting $Q_{\alpha}$ to $\{N, NS\}$ and $Q_{\beta}$ to $\{M\}$. The second classifier $\tau_{dif}$ was
evaluated in a more clinical perspective, were the radiologist typically is only interested in the differentiation between abnormal enhancing areas \{NS\} and PCa \{M\}. The classification performance is denoted as \(AUC_{\text{dif}}\) where \(Q_a\) to \{NS\} and \(Q_p\) to \{M\}.

Prospective performance of the lesion analysis was estimated by means of leave-one-patient-out (LOPO) cross validation. LOPO avoids training and testing on the same data, estimating the likelihoods of ROIs in that left-out case, and repeating the procedure until each case has been tested individually. Our study was a diagnostic assessment with patient-clustered data, and, thus, the bootstrap resampling approach with 10 000 iterations was used for estimating the bootstrap mean AUCs and 95% confidence intervals proposed by Rutter [86]. When a patient case is drawn, the entire set of \(S_r\) for that case enters that bootstrap sample. In doing so, bootstrapping mimics the underlying probability mechanism that gave rise to the observed data. Statistical analyses were performed with the package R [76].

### 2.5 Results

Of the 34 patient studies, 4 were excluded because of insufficient dynamic data caused by patient movement or coil artifacts. In total 39 \(M\) regions were annotated in the peripheral zone. The number of \(NS\) regions annotated in the peripheral zone was 21. The number annotated \(N\) regions was 30.

When looking at the scatterplots of figure 2.5a noticeable clustering of features is seen. The scatterplots demonstrate that the feature values are usable to characterize lesions as \(M\), \(NS\) or \(N\). It can be observed that the \(N\) regions are compact and well clustered. Although the regions of type \(NS\) and \(M\) show a larger spread, they are still clustered and can thus be differentiated. Furthermore, the \(NS\) regions appear to be more clustered than the \(M\) regions.

The localization performance of the discrimination between \{\(N,NS\)\} and \{\(M\)\} is demonstrated in the ROC curve shown in figure 2.6(a). The figure shows that the diagnostic accuracy (\(AUC_{\text{loc}}\)) was 0.92 (95% confidence intervals = 0.87-0.97)). In figure 2.6(b) the discriminating performance between \{\(NS\)\} \{\(M\)\} regions is demonstrated. The diagnostic accuracy (\(AUC_{\text{dif}}\)) in this case was 0.83 (95% confidence intervals = 0.75-0.92)). The ROC curves show that the performances are statistically better than chance.

Figure 2.7 presents a true positive case as well as a true-negative case: in both the transverse and coronal views of the prostate, a bi-lateral enhancement is seen in the peripheral zone when overlaying several parametric maps on the T2-w images. Because of the enhancement, both sides are suspicious for cancer. The CAD system however, calculated a likelihood of malignancy of 80% for the annotated region that was identified as PCa by histophathology. In the other region, the CAD system calculated a likelihood of 20% of being malignant. Additionally, histopathology confirmed that there was no evidence for tumor at the specific location.

### 2.6 Discussion

This study showed that it is feasible to develop a CAD system capable of discriminating PCa from the normal peripheral zone and non-malignant disorders with a diagnostic accuracy of 0.92 (0.87-0.97). It was also shown that it is possible to develop a more clinically relevant CAD system, where the radiologist typically is only interested in abnormal enhancing areas. For the discrimination of solely non-malignant suspicious enhancing (\(NS\)) areas from PCa in the peripheral zone, a diagnostic accuracy of 0.83 (0.75-0.92) was obtained. This CAD system thus has the potential of being a valuable, additional diagnostic aid.

The proposed CAD method has some similarity with the study of Fütterer et al. [11]. In their study, it was shown that when using T2-w images and DCE-MRI in localizing PCa, radiologists achieved an overall accuracy of 0.92, when discriminating PCa pre-assigned regions from normal peripheral zone and non-malignant disorder pre-assigned regions. Although the focus of this study was the normal peripheral zone of the prostate, similar regions were used for the characterization by the CAD system. Furthermore, the same patient database was used. Our CAD method on the contrary, was trained with primarily pharmacokinetic features, whereas the radiologist used the T2-w images as an additional feature of region characterization.

The results of this study demonstrate for the first time in an objective manner that including DCE-MRI can discriminate PCa from \(NS\) areas in the peripheral zone. This is supported by former studies where
Computerized analysis of prostate lesions in the peripheral zone using DCE-MRI

Figure 2.5: Pairwise scatterplots of 4 kinetic parameters and T1 parameter for the whole database with triangles representing N regions, spheres as M regions and squares as NS regions. The ellipses summarize the three clusters by fitting a bivariate normal distribution and displaying the outline at 2 times standard deviation radius. A noticeable clustering of features is seen.
human observers concluded the same [44, 42, 43, 40, 41, 37, 39, 38, 87, 88].

The developed CAD system is capable of displaying multimodal MR images including DWI, T2*-w images, derived spectral maps from spectroscopic data, etc. Although the CAD program is developed in such a manner that it can include features from all available images as relevant information to train the classifier, only the pharmacokinetic and T1 estimate data was used. To further include features from the additional modalities, registration techniques are essential to compensate for patient movements. It can be expected that by extracting the additional features, the discriminating performance of the CAD system will further improve. Several studies indicated that combining multimodal MR images increased the localization accuracy [11, 57].

Histological correlation with MR images is recognized to be an imperfect gold standard for a number of reasons. These include: errors in registering the location of the imaging sections with histological slice specimens, inaccuracies resulting from tissue shrinkage secondary to fixation and errors due to partial volume averaging effects [38, 84, 89]. In most studies the number of slices is simply counted taking the shrinkage into account and using the morphology of the central gland, peripheral zone, cysts, calcifications, and urethra as landmarks to find the corresponding MRI slice. In this study great effort was put into the histopathology and MRI correspondence for an objective annotation of the ground truth. Therefore a 3D deformable surface was created to semi-automatically segment the prostate and divide it in the same number of slice-sections of the histopathology tumor maps. The method ensures that the user is only guided by the histopathology tumor maps, precontrast T1-w and T2-w images for placement of the ROIs. No DCE-MRI parametric maps were used as guidance in ROI placement, since this could introduce bias in CAD performance estimates. To further reduce user-variability, the whole lesion was annotated instead of just the hotspot as suggested by Stoutjesdijk et. al. [75].

Kiessling et al. [90] evaluated the accuracy of descriptive and physiological parameters calculated from signal intensity-time curves using T1-weighted DCE-MRI to differentiate prostate cancers from the peripheral gland. Although they did not create a CAD system capable of calculating a malignancy likelihood, they did evaluate the discriminating performance of the kinetic parameters. Their best performing parameter, early degree of enhancement, achieved an AUC of 0.81. This result can be compared to our localizing classifier $AUC_{loc}$ of 0.92. The difference in performance can be attributed to the method that was chosen to calculate the pharmacokinetic parameters. Kiessling used the method proposed by Brix et al. [91]

![Figure 2.6: ROC curves showing the discriminating performance of the CAD system of the two separate trained classifiers $\tau_{loc}$ and $\tau_{dif}$. The dotted curves are part of the bootstrapping approach and represent the 95% confidence intervals of the solid-line ROC curve. Subfigure 2.6(a) shows the discriminating performance between regions of type N and NS versus M. Subfigure 2.6(b) shows the discriminating performance between regions of type NS versus M.](image-url)
Computerized analysis of prostate lesions in the peripheral zone using DCE-MRI

Figure 2.7: Example of two regions in the peripheral zone of the prostate that are difficult to differentiate. A bi-lateral enhancement is seen with pharmacokinetic parameters, suggesting that both sides are suspicious of prostate cancer. For both regions the CAD system calculated the likelihoods of malignancy using the classifier with a diagnostic accuracy (AUROC) of 0.92. Subfigure 2.7(a) shows a transverse T2-w image of the prostate with $V_e$ color overlay map. Note the bilateral enhancement of the peripheral zone, which makes both sides suspicious for cancer. Subfigure 2.7(b) shows a coronal T2-w image of the prostate with $K_{trans}$ color overlay map. This parameters also shows a bi-lateral enhancement, suggesting the presence of tumor at both sides. In subfigure 2.7(c) the CAD predicted likelihoods of the annotated regions in the left and right peripheral zone are shown. The green area summarizes the distribution of all the calculated likelihoods for all $N$ and $N+S$ regions from the database used for training the classifier. The red area summarizes the smoothed distribution of all the calculated likelihoods for all $M$ regions from the database used for training the classifier. The areas are smoothed using a two-dimensional nonlinear variable span smoother based on local linear fits, in which local cross-validation is used to estimate the optimal span. The black lines are the calculated likelihoods of malignancy for the regions. The dotted line corresponds with region in the left of the peripheral zone (with a predicted malignancy likelihood of 80%). The solid line corresponds with the region in the right of the peripheral zone (with a predicted malignancy likelihood of 20%). Histopathology confirmed the adenocarcinoma in the left peripheral zone (subfigure 2.7(d)). The tumor was staged as T2b and with a Gleason score of 3+3.
where a fixed arterial input function for every patient is assumed (fixed calibration), whereas in this study the reference tissue model (per patient calibration) was used (see appendix A). In a previous study [92] we showed that a per patient calibration indeed has a positive effect on the discriminating performance of PK parameters over a fixed calibration.

Chan et al. [50] describe the only in-vivo CAD system that provides an estimated malignancy likelihood by combining information from T2-weighted, T2-mapping, and line scan diffusion images. They achieved a diagnostic performance of 0.84. This can be compared to our $AUC_{loc}$ of 0.92. The lower performance is likely attributed to the lack of DCE-MRI features. Moreover, we have also researched and demonstrated the ability of our method to discriminate suspicious enhancing benign regions from malignant regions. The latter is of even greater importance in actual clinical conditions.

The current study has a number of limitations. The CAD system is not fully automated, since the user needs to identify normal peripheral zone for calibration with the reference tissue method (see appendix 2.7). As a result, the healthy tissue needs to be annotated in advance, which could result in the annotation of PCa, which makes the CAD system not clinical usable. An automated calibration technique makes the CAD system fully automated and is being researched. The effect of user-variability in annotating the ground truth on the performance has not been researched.

In conclusion, this study demonstrated the possibility to develop a CAD system capable of objectively discriminating malignant lesions from $NS$ areas located in the normal peripheral zone of the prostate with an accuracy of 0.83 (95% confidence intervals = 0.75-0.92).
2.7 Appendix: DCE-MRI postprocessing and pharmacokinetic modeling

All MRI data was transferred to an independent workstation with in-house build software. Each MR signal enhancement-time curve was first fitted to a general exponential signal enhancement model as described previously [37]. This reduces a curve to a 5 parameter model: baseline (s0); start of signal enhancement (t0), which defines the onset of the exponential curve; time-to-peak (τ), the exponential constant; peak enhancement (s_p), the signal amplitude at which the exponential curve levels off; and late wash, defined as the slope of the late part of the exponential curve. The reduced signal enhancement-time curve was converted to a reduced tracer concentration [mmol/ml] - time curve [45] effectively converting s_p to C_{p.d,p}.

We have implemented the method [45] such that in an intermediate step the T1 estimates are computed. The T1Static parameter is the baseline T1 estimate (s0) prior to contrast enhancement.

Analysis of DCE-MRI data is usually based on the indicator dilution theory and requires knowledge of the concentration of the contrast agent in the blood plasma. Without any calibration, inter-patient plasma profile variability causes fluctuations in PK estimates, which are not related to the tissue condition. When using a power injector the most likely cause of plasma curve differences is the patient itself, e.g. differences in body weight (total distributional volume), heart rate, vascular condition. Removing the plasma shape can be regarded as a form of patient calibration. Among the wide variety of techniques for estimating plasma profiles, we have chosen for the reference tissue method and experienced robust results with the technique [93]. The reference tissue method assumes that a tissue area within the patient is available with a known tissue model based on literature values [94, 93]. By doing a deconvolution the actual tissue impulse response can be determined. Deconvolution of the plasma profile and estimation of pharmacokinetic parameters conforms to the theoretical derivations [95] but is implemented in the reduced signal space as shown in the following equation:

\[
Ve = \frac{C_{gd,p,tissue}}{C_{gd,p,plasma}}
\]

\[
k_{cp} = \frac{1}{\tau_{tissue} - \tau_{plasma}}
\]

\[
K^{trans} = Ve.k_{cp}
\]

where \( Ve \) is an estimate of the extracellular volume [%], \( K^{trans} \) the volume transfer constant [1/min], and \( k_{cp} \) the rate constant [1/min] between extracellular extravascular and plasma space. The subscript ‘tissue’ stands for a measurement in the tissue under investigation and subscript ‘plasma’ for the reference tissue plasma estimates based on literature values [94]. The reference tissue was determined by selecting manually a set of voxels in the healthy (normal) peripheral zone using whole mount section histopathology as guidance.
Chapter 3

Computer assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI

This chapter is based on the manuscript “Computer assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI.” by Pieter C. Vos, Thomas Hambrock, Jelle Barentsz, Henkjan Huisman Physics in Medicine and Biology, vol. 55, no. 6, pp. 1719–1734, 2010.
Abstract

In this study, computer assisted analysis of prostate lesions was researched by combining information from two different magnetic resonance (MR) modalities: T2-weighted (T2-w) and dynamic contrast enhanced (DCE) T1-w images. Two issues arise when incorporating T2-w images in a Computer Aided Diagnosis (CADx) system: T2-w values are position as well as sequence dependent and images can be misaligned due to patient movement during the acquisition. A method was developed that computes T2 estimates from a T2-w and proton density value and a known sequence model. A mutual information registration strategy was implemented to correct for patient movement. Global motion is modelled by an affine transformation, while local motion is described by a volume preserving non-rigid deformation based on B-Splines. The additional value to the discriminating performance of a DCE T1-w based CADx system was evaluated using Bootstrapped ROC analysis.

T2 estimates were successfully computed in 29 patients. T2 values were extracted and added to the CADx system from 39 malignant, 19 benign and 29 normal annotated regions. T2 values alone achieved a diagnostic accuracy of 0.85 (0.77-0.92) and showed a significantly improved discriminating performance of 0.89 (0.81-0.95), when combined with DCE T1-w features.

In conclusion, the study demonstrated a simple T2 estimation method that has a diagnostic performance such that it complements a DCE T1-w based CADx system in discriminating malignant lesions from normal and benign regions. Additionally, the T2 estimate is beneficial to visual inspection due to the removed coil profile and fixed window and level settings.
3.1 Introduction

Several studies have indicated that multimodal MRI increases the prostate cancer (PCa) localization accuracy of the radiologist. The accuracy is, however, dependent on the experience of the radiologist [57, 11, 50]. To help improve the diagnostic accuracy of the unexperienced radiologist, we investigated the value of a Computer Aided Diagnosis (CADx) system. In a previous study [96], the feasibility was shown of a CADx system that calculates the malignancy likelihood of a given suspicious area in the peripheral zone of the prostate using T1-w DCE-MRI. Discrimination of malignant and benign regions was performed using a support vector machine (SVM) as classifier that was trained with features extracted from quantitative pharmacokinetic (PK) maps as well as T1 estimates. The study showed that a diagnostic accuracy of 0.83 (0.75-0.92) was obtained by a stand-alone CADx. It is expected that by adding more MR modalities, the discriminating performance of the CADx system will further improve.

In this paper, the possibility of using T2-w images as an additional MR modality to discriminate PCa from benign regions in the peripheral zone (PZ) of the prostate is studied, as they are also used by the radiologist for localizing PCa. Two issues arise when including the T2-w images in a DCE based CADx system. Firstly, there can be misalignment of T2-w and DCE images as patient movement between the series is inevitable during a prostate study. Secondly, the acquired T2-w signal intensities are not linearly related to the underlying tissue T2 relaxation times and depend on the 3D spatial position relative to the receiving coil elements. To resolve the nonlinearity, a T2 estimator was used that was published previously [97, 98]. The approach requires both T2-w, proton density (PD) images and a known sequence model. Note that although the acquisition of quantitative T2 maps is possible using a multi-echo spin echo technique in a reasonable time span on contemporary MR systems, spatial resolution will be considerably lower. Patient movement can be retrospectively corrected by aligning images using image registration, which is necessary for both the T2 estimation as well as inclusion in the CADx system. In this study a method is proposed that models global shifts with affine transformation and local deformation using B-Splines. The method also estimates the coefficients to maximize the mutual information (MI) between the two images. The method is based on the work of Rueckert et al. [99] which was further developed and tested by Mattes et al. [100]. MI based registration is a common choice to register different modalities and is considered suitable for registration of images that do not have the same pixel intensity range [101, 102, 103, 104].

To our knowledge, there have been no reported studies about similar work on PCa using DCE-MRI and T2-w images. Chan et al. [50] describe the only in-vivo CADx system that provides an estimated malignancy likelihood using multimodal MRI. They constructed a summary statistical map of the peripheral zone based on the utility of multi-channel statistical classifiers combining textural and anatomical features in PCa areas from T2-w images, diffusion weighted images (DWI), PD maps and T2 maps. The achieved diagnostic performance of 0.84 is, however, of limited clinical value because the discrimination did not include benign regions such as prostatitis or hemorrhage. Madabhushi et al. [69] generated similar statistical maps based on T2-w images using whole mount sections as the “ground truth” and showed the additional value of combining numerous features. Unfortunately no discrimination performance was calculated, computation time for analysis of one complete MRI scene exceeds an hour and the method is limited to 2D ex vivo MRI. Viswanath et al. [105] extended the method of Madabhushi et al. with a non-rigid registration scheme to map PCa whole mount histological sections onto corresponding 2D DCE-MRI. Though the method potentially improves the objective annotation of PCa, the corresponding slice still needs to be selected. The unsupervised classification by k-means clustering achieved an accuracy of 77%. Their advocated methodology is, however, evaluated on a per-pixel basis, whereas the proposed method captures the heterogeneous nature of PCa by using percentiles within a given region [73, 74].

The purpose of this study was to investigate the feasibility of including T2-w MR in a multi-modal computer aided diagnostic system for prostate MR.

3.2 Materials and Methods

3.2.1 Patient Characteristics

The study set consisted of 34 consecutive patients (mean age, 60 years; range, 50–69 years) that were selected in a previous study of Futterer et al. [11]. These patients had biopsy-proven PCa and underwent
DCE-MR imaging at 1.5-T, complementary to the routine staging MR imaging examination of the prostate. Patients were included (between April 1, 2002, and June 1, 2004) in the study only if they were candidates for radical retropubic prostatectomy within 6 weeks after MR imaging. The study of Futterer et al. was approved by the institutional review board and informed consent was obtained from all patients prior to MR imaging. After imaging, all patients underwent radical retropubic prostatectomy. Exclusion criteria were: previous hormonal therapy, con-traindications to MR imaging (e.g., cardiac pacemakers, intracranial clips), contraindications to endorectal coil insertion (e.g., anorectal surgery, inflammatory bowel disease). The mean prostate specific antigen level was 8 ng/mL (range, 3.2-23.6 ng/mL), mean Gleason score was 6.1 (range, 5-8). MRI was performed on average 3 weeks after transrectal ultrasonographically guided sextant biopsy of the prostate.

3.2.2 MR Protocol

Images were acquired with a 1.5T whole body MR scanner (Sonata, Siemens Medical Solutions, Erlangen, Germany). A pelvic phased-array coil as well as a balloon-mounted disposable endorectal surface coil (MedRad®, Pittsburgh, PA, USA, inserted and inflated with approximately 80 cm$^3$ of air) were used as receiver. The integrated body coil was used for transmit. All patients received one mg of glucagon (Glucagon®, Novo Nordisk, Bagsvaerd, Denmark) was administered intramuscularly directly prior to the MRI scan, to reduce peristaltic bowel movement during the examination.

Table 3.1 shows the sequence parameters. The protocol for acquisition consisted first of a localizer and two fast gradient spin echo measurements for patient and coil positioning verification. Thereafter, high-spatial-resolution T2-w turbo spin echo imaging in the axial, sagittal and coronal planes, covering the prostate and seminal vesicles, was performed. Next, 3D T1-w spoiled gradient echo images were acquired before and during an intravenous bolus injection of paramagnetic gadolinium chelate (0.1 mmol/kg, gadopentetate, Magnevist®; Schering, Berlin, Germany) using a power injector (Spectris, Medrad®, Pittsburgh, PA, US) with an injection rate of 2.5 ml/second followed by a 15 ml saline flush. The recorded repetition time (TR) of 34 msec represents the time to acquire 12 k-lines (two of 12 slices where used for oversampling). A 6/8 partial Fourier reconstruction (FFT) was used to reduce the acquisition time. With these settings, complete 3D volumes were acquired every two seconds for a duration of 120 seconds. Before contrast injection, the same axial 3D T1-w gradient echo sequence, though with a longer TR time without partial Fourier reconstruction, was used to obtain PD-w images at identical positioning to allow calculation of the T2 estimates and the PK parameters. The gadolinium concentration curves and PK parameters were calculated as reported by Huisman et al. [37]. The PK parameter set that was used in this experiment, consisted of the pre-contrast static value of the T1 estimate of the longitudinal relaxation rate in [ms] ($T1_{Static}$), the relative size of the extracellular, extravascular space ($V_e$), the rate constant ($k_{eq}$) and the transfer constant ($K_{trans}$). They are described in detail in [96].

3.2.3 Image Registration Algorithm

This section describes an image registration method that automatically seeks for an estimate of the transformation $T$ that aligns the PD image $F$ and T2-w image $M$ by optimizing a similarity measure $S$ over the transformation $T$:

$$S(F, M(T)),$$  \hspace{2cm} (3.1)

The registration procedure consisted of several components, the most important of which were the choice of similarity measure, the transformation degrees of freedom and the cost function for similarity of measure-ment.

Deformation model

An initial alignment of $F$ and $M$ was achieved using a 12 element affine transformation matrix $T_{global}$, as reported previously [106]. In general, an affine transformation is composed of linear transformations (rotation, scaling or shear) and a translation. It ensures a global registration with low computational costs. An additional transformation $T_{local}$ modeled local deformation. In this work, deformations are modelled on cubic B-splines, because of their computational efficiency (separability in multiple dimensions, calculation...
### Table 3.1: Parameters for MR Imaging

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Seq. type</th>
<th>Duration (min)</th>
<th>TR (msec)</th>
<th>TE (msec)</th>
<th>Echo train length</th>
<th>No. of averages (NEX)</th>
<th>Flip angle (dgr)</th>
<th>Slice thickness (mm)</th>
<th>Matrix</th>
<th>No. of slices</th>
<th>Field of view (mm x mm)</th>
<th>Phase-encoding direction</th>
<th>Spatial resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2-w Axial</td>
<td>tse¹</td>
<td>3m50s</td>
<td>4400</td>
<td>132</td>
<td>15</td>
<td>2</td>
<td>180⁰</td>
<td></td>
<td>240x512</td>
<td>11-15</td>
<td>132x280</td>
<td>R-L</td>
<td>.55x.55x4</td>
</tr>
<tr>
<td>T2-w Sagittal</td>
<td>tse</td>
<td>4m40s</td>
<td>4000</td>
<td>132</td>
<td>15</td>
<td>2</td>
<td>180⁰</td>
<td></td>
<td>179x512</td>
<td>11-15</td>
<td>100x280</td>
<td>A-P³</td>
<td>.55x.55x4</td>
</tr>
<tr>
<td>T2-w Coronal Dynamic T1-w PD 3D gradient-echo</td>
<td>tfl²</td>
<td>4m</td>
<td>4000</td>
<td>132</td>
<td>15</td>
<td>2</td>
<td>180⁰</td>
<td></td>
<td>179x512</td>
<td>11-15</td>
<td>100x280</td>
<td>R-L</td>
<td>.55x.55x4</td>
</tr>
<tr>
<td>Dynamic T1-w fast 3D gradient-echo</td>
<td>tfl</td>
<td>1m56s</td>
<td>800</td>
<td>1.6</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td></td>
<td>77x256</td>
<td>10³</td>
<td>240x280</td>
<td>A-P</td>
<td>3.1x1.1x4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0m2s</td>
<td>34⁵</td>
<td>1.6</td>
<td>1</td>
<td>1</td>
<td>14</td>
<td></td>
<td>77x256</td>
<td>10³</td>
<td>240x280</td>
<td>A-P</td>
<td>3.1x1.1x4</td>
</tr>
</tbody>
</table>

¹ Turbo spin echo, ² Turbo flash, ³ Right-left, ⁴ Anterior-Posterior, ⁵ Represents the time to acquire 12 k-lines, ⁶ Two of 12 slices used for oversampling, ⁷ Refocussing flip angle
via filtering), smoothness, and local control [99]. A B-Spline transformation deforms a volume by manipulating an underlying mesh of control points ($D$). Deformation is calculated using the positions of a 4x4x4 neighborhood of control points and third-order spline polynomials, where the parameters of the B-Spline transformation are the coordinates of the control points. In the presented method, the initial $T_{global}$ was used as offset of $T_{local}$:

$$T = T_{local} - T_{global},$$

(3.2)

**Similarity measure**

In this study $MI$ was used as a similarity measure because $F$ and $M$ do not have the same appearance. When $F$ and $M$ are optimally aligned, their $MI$ is maximal:

$$\hat{T} = \arg \min_{\psi} ( -MI(\psi) ),$$

(3.3)

where $MI(\psi)$ denotes the $MI$ similarity measure as a function of the transformation parameters, $\psi$. Although several implementations for the computation of $MI$ exists, the presented method is based on the work of Mattes et al. [100]. Their registration method makes efficient use of the B-Spline basis functions that models the deformation field and converges quickly when compared to other methods, due to stochastic sampling [107].

$MI$ relates the joint entropy to the entropies of the modalities separately:

$$MI(\psi) = H(f(x)) + H(m(\psi(x))) - H(f(x), m(\psi(x))),$$

(3.4)

where $H(f(x))$ and $H(m(\psi(x)))$ are the marginal entropies, $H(f, m; \psi)$ the joint entropy, $f(x)$ and $m(x)$ denote observations of $F$ and $M$, respectively and $x$ denotes the voxel coordinate. The joint entropy $H(f, m; \psi)$ uses a set of fixed bin centres in $f$ and moving bin centres in $m$ as initialization.

The entropy of an image is computed from the probability density function (pdf) of the image intensities. Parzen-windowing was used to obtain a differentiable pdf. This reduces the effects of quantization from interpolation and discretization from binning the data, see Thevenaz et al. [108] for a detailed description. The joint entropy is computed using:

$$H(f, m; \psi) = \sum_{x \in X_F} w_f(I_F(x_F) - f)w_m(I_M(g(x_F; \psi)) - m),$$

(3.5)

where $I_F(x_F)$ and $I_M(g(x_F; \psi))$ are samples of the fixed and interpolated moving images, respectively and $w_f$ and $w_m$ are the Parzen window kernels [100]. To reduce the computational costs, the set of intensity samples ($X_F$) was set to a randomly selected 20% of the total number of voxels and the joint histogram was calculated based on 100 bins for each image. These settings gave robust results on all experiment data. To further speed up the calculation of the $MI(\psi)$, samples were only drawn from a subregion of $F$. This will be further discussed in section 3.2.3.

**Constraint on the transformation**

In addition to the $MI$ similarity measure, a regularization term $P$ was incorporated in the registration method to constrain the deformation of the coordinate space:

$$C(F, M(T)) = -S(F, M(T)) + \alpha P(T),$$

(3.6)

where $\alpha$ is a user defined weight factor for $P$, allowing more control over the influence of the regularization term on the overall cost $C$. Similar to the work of Rohlfing et al. [109], a volume preserving constraint was incorporated and implemented by penalizing deviations of the Jacobian determinant $J$ of the deformation from unity, that is, it assumes local rigidity and penalizes local tissue expansion and compression:

$$P(\psi) = \frac{1}{ND} \sum_{x \in D} |\log(J(\psi(x))|,$$

(3.7)

where $D$ represents the set of control points used for the deformation.
Calculation of the gradient of the constraint is necessary for an efficient and robust optimization (see section 3.2.3). The derivatives were calculated using the common finite-difference approximation:

$$\frac{\delta P}{\delta i} \approx \frac{P(\psi_i + \delta_i) - P(\psi_i)}{\delta_i},$$

(3.8)

where $\delta_i$ equals 1 mm and the subscript $i$ refers to the parameter index.

**Optimization of the cost function**

The L-BFGS-B [110], a limited-memory, quasi-Newton minimization package, was used to optimize the cost function $C$ until termination criteria were satisfied. The limited-memory method is well suited for optimizing the large number of parameters $\psi$ in the B-Spline transformation when a high resolution of control points is used. L-BFGS-B provides an additional advantage in that it allows bound constraints ($B$) on the independent variables. In this manner, maximum displacements of the control points can be controlled. Note that displacements in the slice direction were considered unlikely and therefore restricted with a fixed bound set to 0. The optimization terminates when the change in $C$ between consecutive iterations falls below the tolerance of $10^{-4}$.

**Centered cropped optimization strategy**

A high resolution of control points in the B-Spline transform $T$ comes with high computational costs, due to the high number of displacements of the control points $D$ that needs to be explored. For the intended application, the focus is on the non-rigid displacement of the prostate. Hence, it suffices to limit the registration method to the center area of the pelvis, where the prostate is situated. The set of control points ($D$) is defined over the entire volume ($F$) plus a finite support region of three at the borders, thereby extending the region of interest. Simply defining a smaller mesh inside the volume results in interpolation artifacts near the edges of the grid. Therefore, optimization was limited to a centered subset ($D_R \subset D$) of control points. Because the number of parameters in $D_R$ is much smaller, the registration method becomes considerably faster. The resolution of the grid control points was set to 14x14x8 (including the finite support region). A centered subregion ($D_{R_c}$) of size 4x4x3 was chosen such that the prostate area was covered in all cases.

**3.2.4 Histological Verification**

All patients underwent radical retropubic prostatectomy. From the prostatectomy specimens, whole-mount step-section histology tumor maps were created by an experienced pathologist who was blinded to the imaging results. The whole-mount step-section histology tumor maps were used as ground truth for training and evaluating the performance of the CADx system. The morphology of the central gland, peripheral zone, cysts, calcifications and urethra were used as landmarks to find the corresponding MRI slice. The anatomy of the prostate is best imaged on T2-w images and were therefore used for correlating the histopathological map. A method was implemented to prevent bias when annotating the ground truth, as described elsewhere [96]. Using this method, regions of normal PZ, PCa, benign PZ tissue (identified as tumor suspicious on T2-w and DCE-MRI but not representing tumor on pathology) and levator ani muscle were outlined manually. The regions that contained muscle were used for the reference tissue method, as explained in section 3.2.5.

**3.2.5 Quantitative T2 estimation from T2-w and PD series**

T2 relaxation times can be computed with a fast method that uses a known sequence signal model as prior knowledge and only a few echo times (TE) to fit the T2 relaxation curve. Hittmair et al. [45] presented a method that estimates the T1 relaxation rate with only two repetition times (TR) and a spin echo sequence as signal model. In this study a comparable approach is applied to estimate the T2 using a T2-w turbo spin echo sequence and PD-w gradient echo sequence. It is assumed that a T2-w sequence is used, such that the TR is much larger then the T1 tissue relaxation times. Then the effect of T1 can be neglected and the received signal $s$ at location $x$ for a T2-w turbo spin echo sequence is:

$$s_{T2w}(x) = G_{T2w} \sin(\theta_{T2w}) \rho(x) \exp(-TE/T2(x)),$$

(3.9)
where $G_{T2w}$ represents the gain setting and $\theta_{T2w}$ the excitation flip angle for the T2-w sequence and $\rho$ is a function comprising proton density fluctuations and coil profile at location $x$. The spatial dependence through $\rho(x)$ models the commonly observed spatial inhomogeneity that is caused by the receive and send coil sensitivity profiles. In case of a PD-w image, the effect of T1 and T2 should be reduced when TE is set short and TR is set long. Assuming that both sequences use the same coil setup ($\rho(x)$ is identical), the received PD signal can be approximated by:

$$s_{pd}(x) = G_{pd}\sin(\theta_{pd})\rho(x), \quad (3.10)$$

where $G_{pd}$ is the gain setting and $\theta_{pd}$ the excitation flip angle for the PD-w gradient echo sequence. The $T2$ at position $x$ is derived by rewriting Eq. 3.9 and Eq. 3.10:

$$T2(x) = \frac{-TE}{\log(s_{T2w}(x)) - \log(s_{pd}(x)) - \log(\eta_{T2w,pd})}, \quad (3.11)$$

where $\eta_{T2w,pd}$ is the gain ratio of $G_{T2w}\sin(\theta_{T2w})$ to $G_{pd}\sin(\theta_{pd})$. In theory, $\eta$ can be estimated from the MR sequences parameters, but the data was not stored in the DICOM header provided by the MR system. Furthermore, any coil profile artifacts, differences in voxel size, TR or receiver bandwidth can still result in acquisition-to-acquisition signal intensity variations, in spite of the fact that commercial MR systems are nowadays equipped with internal calibration methods. These variations are captured in $\eta$ and need to be estimated from the data. Therefore, an estimator for the gain ratio $\eta_{T2w,pd}$ was built using a reference tissue $R_{ref}$ with known T2 value from literature. In this study levator ani muscle was used (35.3±3.85 msec at 1.5T and 37°C [111]). The gain setting can now be found by performing least squares optimization using the data from all pixels within the annotated reference tissue:

$$\eta_{opt} = \arg\min_{\eta} \sum_{x \in R_{ref}} (T2(x; \eta) - 35)^2. \quad (3.12)$$

The least square procedure that was used, uses Brent’s method of parabolic interpolation, protected by golden-section subdivisions if the interpolation is not converging.

### 3.2.6 Training and Validation

The CADx system extracts a PK and T2 feature set from a region of interest (ROI) using percentiles. The extracted set of features is presented to a trained SVM which calculates the malignancy likelihood for a lesion. The calculated likelihood is presented to a radiologist to assist in his or her diagnosis. The CADx system was implemented in an open source programming environment The Visualization ToolKit (VTK) using the Tool Command Language (Tcl) and C++.

The discriminating performance of the CADx system was estimated using the area under the receiver operator characteristics (ROC) curve (AUC). Classification was performed using SVM analysis on the feature set (provided by the statistical package R [76]).

A prospective–performance estimate of the lesion analysis was made by means of leave-one-patient-out (LOPO) cross validation. LOPO avoids training and testing on the same data, estimating the likelihoods of ROIs in that left-out case and repeating the procedure until each case has been tested individually. This study was a diagnostic assessment with patient-clustered data, therefore a bootstrap resampling approach with 10 000 iterations, was used for estimating the mean AUCs and 95% confidence intervals, as proposed by Rutter [86].

The intent of this study was not to provide a new T2 estimator. Nevertheless, some validation was performed to research the validity of the T2 estimates. Firstly, visual inspection was performed. Secondly, the median and variation for T2 relaxation time of the normal peripheral zone were computed. The results were compared with those found in literature. Thirdly, the method was compared with a multi-echo spin echo sequence, where the T2 relaxation curve was automatically fitted by a Siemens mono-exponential decay fitting algorithm. The sequence settings were 7 echo times (15.6–109.2msec), a spatial resolution of 1.2x1.2x3.0mm, matrix of 192x96, field of view of 230x115, TR of 2080, flipangle of 180 and 16 slices. Levator ani muscle was used as reference and the mean relaxation time was calculated for Muscle, Benign
PZ, Fat, Normal PZ, Normal transition zone (TZ) and Lesion PZ in a random patient case. Linear regression and Pearson correlation statistics were performed.

The effect of registration on the diagnostic performance of the CADx system was studied. In this experiment, T2-w values, T2 estimates without registration, T2 estimates after affine registration and T2 estimates after the implemented non-rigid registration, were extracted from the annotated regions and tested for their discriminative value. In the following experiment, the effect of parameter settings on the diagnostic performance was valuated. The method was tested with different constraint settings ($B$) and various Jacobian weighting factors ($\alpha$), to search for the optimal discriminative performance. In the final experiment the additional discriminative value of T2 estimates in a multimodal setup was tested and compared with the performance of the previous developed CADx system.

3.3 Results

In total, 34 consecutive patients with histologically proven adenocarcinoma of the prostate were recruited. Four patient studies were excluded because of bad dynamic data caused by large patient movement, coil artifacts and one due to large noise values. In total 39 malignant regions were annotated in the peripheral zone. The number of benign regions annotated in the peripheral zone was 19. The number of annotated regions in the normal peripheral zone was 29.

Fig. 3.1 shows an example case, where the coil profile is prominently visible in the T2-w image, but is removed by the method in the quantitative T2 image. In Fig. 3.2 the value of non-rigid registration is demonstrated. Patient movement during the examination as well as internal (bowel) movement were corrected by the method.

The median calculated T2 relaxation time was 82±7.5 msec for normal PZ, 68±15.4 msec for benign regions and 60±6.9 msec for malignant regions, at the best possible setting of the registration method and using levator ani muscle as reference tissue. In Fig. 3.3, T2 values of several tissues of one patient case are shown and compared to a Siemens T2 sequence acquired in that same study. A significant linear correlation can be observed.

Fig. 3.4 shows the histograms of the distributions of the T2-w values, T2 estimates without registration, T2 estimates after affine registration and T2 estimates after non-rigid registration, that were extracted from all normal, benign and malignant regions. It can be observed that the amount of overlap between the distributions (especially M and N) is lowest after the non-rigid registration.

Fig. 3.6 shows the effect of varying the parameters $\alpha$ and $B$ on the registration method. At low Jacobian weighting factors ($\alpha < 1.5$), the algorithm tends to be sensitive to the settings of $B$. Best settings were
Figure 3.2: Sample images illustrating the necessity for nonrigid registration. The acquired T2-w image (background image) and the PK parameter $K^{trans}$ (colour-coded transparent overlay) are misaligned in the first image. Typical enhancement patterns are observed peri urethral (arrow) and at the neurovascular bundles, but due to patient movement during the examination, the enhancement patterns do not match the underlying T2-w image (left). This can also be observed in the figure in the middle which shows the result of an affine registration, where after a global correction the neurovascular bundles match the enhancement patterns, but a peri urethral mismatch is still present. The right image shows the result of the non-rigid registration method. Note the correct location of enhancement at the peri urethral region.

Figure 3.3: Plot showing the mean T2 relaxation times for several tissues in a patient case, with T2 values generated by a traditional multi-echo sequence of the MR system on the x-axis and on the y-axis, the T2 values generated by the proposed method. The linear regression model and the Pearson correlation statistics are shown. Found at $B = \pm 25mm$ and $\alpha = 1.5$. With these parameter settings the diagnostic value of the T2 estimates was quantified by the discriminating performance of the CADx system. In a tumor localization setup, where normal PZ regions were discriminated from benign and malignant regions, a diagnostic performance of $0.97 (0.94-1.00)$ was achieved using non-rigid registration, as shown in Fig. 3.5. Using the non-rigid registration method, a significant improvement was shown compared to the affine registration method, as well as to not using registration. The T2-w values were of no diagnostic value to the CADx system. In a differentiation setup, where normal and benign PZ regions were discriminated from malignant PZ regions, the CADx system achieved a diagnostic performance of $0.85 (0.77-0.92)$ after non-rigid registration (Fig. 3.5). This accuracy was significantly better than not using registration ($p=0.03$), but did not show a significant improvement to an affine registration.

The accuracy of the CADx system for discriminating normal PZ and benign regions from malignant regions was $0.84 (0.76-0.92)$ using the PK parameters alone. Adding the estimated T2 relaxation times to the PK feature set resulted in an accuracy of $0.89 (0.81-0.95)$, which was a significant improvement.
3.4 Discussion

This study showed the feasibility of including T2-w MR in a multi-modal CADx system for prostate MR. T2 estimates were shown to be of a significant additional diagnostic value for the in-house developed CADx system [96]. Moreover, coil profile sensitivity, that was present in the T2-w images, was noticeably diminished, making these images potentially more beneficial for clinical interpretation during diagnostic viewing.

Registration had a strong influence on the diagnostic performance of the T2 estimates in discriminating normal and benign from malignant regions. This study is unique in comparing registration methods on the actual effect upon the diagnostic performance. While most registration literature is limited to visual observation or simulated deformations, the present study quantified the effect upon the diagnostic performance. The effect is demonstrated in Fig. 3.5. The non-rigid registration method showed the highest diagnostic ac-
Figure 3.5: ROC curves showing the discriminating performance of the CADx system using T2 estimates and T2-w values. The T2 estimates were first extracted when no registration was applied, second after affine registration and third after non-rigid registration. The left graph shows the result at a differentiation setup, where PZ and benign PZ regions are discriminated from malignant PZ regions. It can be noticed that normal PZ and benign regions are well differentiable from malignant regions. The right graph demonstrates the discriminating performance in a localization setup, where normal PZ is differentiated from both benign regions and malignant regions.

Figure 3.6: Parameters grid search for the B-Spline registration, with bound (B) the optimizer constraint on the displacement of the control points, $\alpha$ the Jacobian factor and $Az$ the diagnostic performance of the CADx system.

As an example, Fig. 3.2 illustrates organ movement that could not be compensated for using affine registration, but was resolved by free form deformation using cubic B-Splines [99]. A disadvantage of this technique is, that it can lead to loss of topology. Therefore, the Jacobian determinant as volume preserving constraint on the transformation was included. Fig. 3.6 demonstrates that including such a constraint, the diagnostic performance of the CADx system improves and leads to a more robust registration, as it is less
Computer assisted analysis of peripheral zone lesions using T2-w and DCE-MRI

Figure 3.7: ROC curves showing the diagnostic performance of the CADx system using the T2 estimates after non-rigid registration, versus only using PK parameters and the additional value of T2 estimates when they are combined with the PK parameters.

The T2 estimator in this study is a novel, simple method in a CADx context. Although a full validation is outside the scope of this study, some aspects indicate validity. Firstly, visual inspection showed that the coil profile was removed. Secondly, the inter-patient variability of T2 estimates in the normal PZ was small (±7.5 msec). Thirdly, a significant linear correlation was shown between the T2 estimates of our method and the multi-echo spin echo sequence (r=0.97, p=0.001). Finally, adding the T2 values as feature to the CADx system resulted in a significantly improved discriminating performance of 0.89 (0.81-0.95), compared to only using the PK features (0.84 (0.76-0.92)). The performance increase agrees well with previous literature on combining T2 estimates and PK parameters [57, 11, 50].

The calculated T2 relaxation times for the prostate (82±7.5 msec) were compared with those found in literature. In the small study of de Bazelaire et al. [112] (based on data from three healthy men), prostate and vertebra relaxation rates of 88±0 msec were calculated. Chan et al. [50] found relaxation rates of 128.3±42.9 for non-brachytherapy patients and 88±21 msec for post-brachytherapy patients in the normal PZ. More recently [113], using a different T2 map imaging sequence, they found, in 18 men with biopsy-proven PCa, 193±49msec in healthy tissue and 100±26msec in suspected cancer. Gibbs and Liney et al. [70] found in an early study relaxation rates of 96.2±15.2 msec in patients but in a more recent study 135.5±40.0 msec in the normal PZ of patients [114]. Two observations can be made from this literature survey. Firstly, there is no consensus on the absolute T2 value of normal PZ. Consequently, any T2 discrimination protocol or CADx system has to account for the T2 estimation method that was used. Secondly, a large inter-patient variability. The presented method has a much lower variability, which partly explains the good discriminative properties of the estimated T2 relaxation times. The reduced variability may be explained by the use of reference tissue (section 3.2.5). In doing so, the method calibrates on a per-patient basis which may be better in capturing machine dependencies. The effect of using calibration on the diagnostic performance was also demonstrated in a previous study [92], but then in a pharmacokinetic context.

The T2 estimator has a number of potential shortcomings. Firstly, muscle T2 variability can result in different T2 estimates, as it is used as reference tissue. The variability is an effect of physical activity and muscle training. In this study, we used levator ani muscle as reference tissue. In contrast to most skeletal muscles, levator ani muscles are non-voluntary, therefore intra-patient T2 variation is small. Secondly, the method is sensitive to B1 fluctuations. The volume of interest (prostate region), however, is in the middle of the image where B1 fluctuations are less present. Thirdly, the T2-w tse sequence has contributed data...
from different echo times (depending on the echo train length and k-space sampling) and therefore the estimated T2 values might not be in exact concordance with a multi-echo spin echo derived estimation. The slight inaccuracy, however, will be evident in all patients and all tissues evaluated. The inaccuracy will be compensated for by using reference tissue calibration (e.g., muscle). Fourthly, the T2-w and PD-w coil setup should be identical.

The current study has a number of limitations. One limitation is the amount of time needed for the non-rigid registration, which now ranges from 5 min to 15 min. This, however, will not conflict the intended application, since the radiologist does not evaluate the images directly after the acquisition and the registration can thus be performed offline. A second limitation is the manual annotation of levator ani muscle, which is needed for the reference tissue method. One solution is to have more information on actual gain coefficients and sequence models used by the MR system. Yet the reference tissue method may have a positive effect on the discriminating performance as stated above. Another potential method would be an automated segmentation of the levator ani muscle and is part of further research. Nevertheless, manual segmentation of the muscle and computing the T2 estimates is performed in seconds. Third limitation is that the registration method parameter Jacobian constraint weight ($\alpha$) requires tuning (Fig. 3.6). Different MR hardware (e.g., at 3 Tesla) and sequence settings may require different parameter settings.

In conclusion, the study demonstrated a simple T2 estimation method that has a diagnostic performance such that it complements a DCE T1-w based CADx system in discriminating malignant lesions from normal and benign regions with a significant improved accuracy of 0.89 (0.81-0.95) compared to only using DCE derived features.
Chapter 4

Automated calibration for computerized analysis of prostate lesions using pharmacokinetic magnetic resonance images

This chapter is based on the manuscript “Automated calibration for computerized analysis of prostate lesions using pharmacokinetic magnetic resonance images.” by Pieter C. Vos, Thomas Hambrock (MD), Jelle Barentsz, Henkjan Huisman. MICCAI, September 20-24, 2009, Proceedings, Part II, Volume 5761/2009, Pages 836-843.
The feasibility of an automated calibration method for estimating the arterial input function when calculating pharmacokinetic parameters from Dynamic Contrast Enhanced MRI is shown. In a previous study [96], it was demonstrated that the computer aided diagnoses (CADx) system performs optimal when per patient calibration was used, but required manual annotation of reference tissue. In this study we propose a fully automated segmentation method that tackles this limitation and tested the method with our CADx system when discriminating prostate cancer from benign areas in the peripheral zone.

A method was developed to automatically segment normal peripheral zone (PZ) tissue. Context based segmentation using the Otsu histogram based threshold selection method and by Hessian based blob detection, was developed to automatically select normal PZ as reference tissue for the per patient calibration. In 38 consecutive patients carcinoma, benign and normal tissue were annotated on MR images by a radiologist and a researcher using whole mount step-section histopathology as standard of reference. A feature set comprising pharmacokinetic parameters was computed for each ROI and used to train a support vector machine as classifier.

In total 42 malignant, 29 benign and 37 normal regions were annotated. The diagnostic accuracy obtained for differentiating malignant from benign lesions using a conventional general patient plasma profile showed an accuracy of 0.65 (0.54-0.76). Using the automated segmentation per patient calibration method the diagnostic value improved to 0.80 (0.71-0.88), whereas the manual segmentation per patient calibration showed a diagnostic performance of 0.80 (0.70-0.90). These results show that an automated per-patient calibration is feasible, a significant better discriminating performance compared to the conventional fixed calibration was obtained and the diagnostic accuracy is similar to using manual per-patient calibration.
4.1 Introduction

Several studies have indicated that multi-modal MRI increases the prostate cancer (PCa) localization accuracy of the radiologist. The accuracy is, however, dependent on the experience of the radiologist [57, 11, 50]. To help improve the diagnostic accuracy of the (unexperienced) radiologist, we are investigating the possible additional value of CADx. Previously [96], the feasibility was demonstrated of an in-house developed CADx system that calculates the malignancy likelihood of a given suspicious area in the peripheral zone of the prostate using T1-w DCE-MRI at 1.5T. Discrimination of malignant and benign regions was performed using a support vector machine (SVM) as classifier that was trained with features extracted from quantitative pharmacokinetic (PK) maps as well as T1 estimates. The study showed that a diagnostic accuracy of 0.83 (0.75-0.92) was obtained by a stand-alone CADx, which is comparable to an expert radiologist performance.

Pharmacokinetic (PK) DCE-MRI could further improve PCa differentiation by reducing inter patient and inter MR scanner fluctuations compared to conventional DCE-MRI. PK tissue parameters are estimated by fitting a tracer physiologic compartment model to the observed DCE-MRI data that is driven by a plasma profile. Various techniques for estimating plasma profiles exist. Quite some PK estimators do not include per patient calibration, but use a general patient plasma profile (fixed calibration) [115, 116]. Huisman et al. [117] demonstrated that the plasma profile varies per patient and thus, fixed calibration can cause fluctuation among patient when estimating the PK parameters. In [92], it was shown that the CADx system performs significantly better using per patient calibration instead of fixed calibration. The presented method was, however, dependent on manual annotation of healthy tissue before a malignancy likelihood could be calculated. This study addresses that limitation by presenting a more objective and automated calibration method and investigates its effect on the diagnostic accuracy of the CADx system.

The purpose of this study was to investigate the feasibility of a CADx system capable of objectively discriminating PCa from non-malignant disorders located in the peripheral zone of the prostate using an automated per patient calibration method.

4.2 Method

4.2.1 Pharmacokinetic modeling

Analysis of DCE-MRI data requires knowledge of the concentration of the contrast agent in the blood plasma. Without calibration (or fixed calibration), inter-patient plasma profile variability causes fluctuations in PK estimates, which are not related to the tissue condition. When using a power injector the most likely cause of differences in plasma curves are differences in body weight (total distributional volume), heart rate, vascular condition. Removing the plasma shape can be regarded as a form of patient calibration whereas fixed calibration uses a fixed plasma function over all patients.

The parametric model for analyzing contrast agent concentration time curves in DCE-MRI is the two compartment model of Tofts et al. [95]. The observed concentration-time curve can be expressed as:

\[ C_v(t) = h(t; t_0, V_e, K^{trans}, Washout) \otimes C_p(t), \]

where \( C_v(.) \) denotes the observed tracer concentration, \( h(.) \) the tissue impulse response, \( C_p(t) \) the plasma input function and \( t_0, V_e, K^{trans}, Washout \) are parameters from the model. The reference tissue method estimates the plasma input function by:

\[ C_p^*(t) = C_{ref,v}(t)/h_{ref,v}(t), \]

where \( C_{ref,v}(.) \) represents the observed plasma profile for tissue \( v \) and \( h_{ref,v}(.) \) a reference plasma profile for tissue \( v \) based on literature. The reference tissue method is considered to be a robust technique [93].

4.2.2 Automated per patient calibration

In a previous study [92], it was demonstrated that using PZ as reference tissue gave good results for estimating PK parameters. In this study a method was developed to auto segment PZ. The method is
divided into two stages. First, the location of the prostate is detected using a blob detection method. In the second stage, this location is further refined to segment a PZ region.

Automated localization of the prostate

The prostate can be modelled as a large enhancing area (or blob) in the pelvis. Figure 4.1(a) demonstrates this model where the prostate can easily be detected by a human observer. Large and strong enhancements can be observed in the transition zone of the prostate making it suitable for detection. First experimental results showed however, that this assumption is not only true for the prostate. Because the acquisition time for the DCE-MRI can be rather long (3 min), contrast agent also arrives in the bladder, resulting in a comparably large enhancing blob, as demonstrated in figure 4.1(b). The prostate model is therefore extended by including the arrival time of the contrast agent ($t_0$ of $C_v(t)$). Otsu’s automatic threshold selection method from gray-level histograms ([118]) is used to segment early enhancing structures in the relative enhancement image $V(x)$:

$$V_O(x) = \begin{cases} V(x), & t_0(x) < th_{Otsu} \\ 0, & \text{otherwise} \end{cases}$$ \hspace{1cm} (4.3)

A common approach to detect blobs is to consider the Taylor expansion of $V_O$ at multiscale for a given neighborhood of pixel $x$ [119],

$$V_O(x + \delta x, \sigma) \approx V_O(x, \sigma) + \delta x^T \nabla_{\sigma} + \delta x^T H_{\sigma} \delta x,$$ \hspace{1cm} (4.4)

where $\nabla_{\sigma}$ and $H_{\sigma}$ are the gradient vector and Hessian vector of an image at scale $\sigma$. Here, $V_O$ is convolved using derivatives of Gaussians:

$$\frac{\delta}{\delta x} V_O(x, \sigma) = \sigma V_O(x) \frac{\delta}{\delta x} G(x, \sigma).$$ \hspace{1cm} (4.5)

Next, from $H_{\sigma}$ eigenvalues $\lambda_{\sigma,k}$ are computed, corresponding the the k-th normalized vector $\hat{u}_{\sigma,k}$ and analyzed to determine the likelihood of a pixel $x$ belonging to a blob. This analysis is based on the following likelihood function (for bright blob, dark background):

$$P(x, \sigma) = |\lambda_1(x)||\lambda_2(x)||\lambda_3(x)|,$$ \hspace{1cm} (4.6)
that is, all three eigen values should be large to represent a blob. A multiscale approach is adopted after which the maximum response is selected:

$$P(x) = \max_{\sigma_{\min} \leq \sigma \leq \sigma_{\max}} P(x, \sigma).$$ \hspace{1cm} (4.7)

The center location of the prostate $x_{pc}$ containing the highest probability is then selected by $x_{pc} = \arg \max_x P(x)$. In figure 4.2(a) a probability map is shown that is used for the prostate detection.

**Automated segmentation of normal peripheral zone tissue**

In the second stage of the method, a context based segmentation is performed to extract normal peripheral zone tissue. The method is based on the model that the $PZ$ is mainly dorsal located of $x_{pc}$. Thus, we define a box-mask below $x_{pc}$ with height, width and depth set to $\sigma$ to mask $V_G$. Here, $\sigma$ corresponds with the size of the prostate and is the scale at which $P(x_{pc})$ was found by the blob detector:

$$S(x_{pc}) = \arg \max_{\sigma} P(x_{pc}, \sigma)$$ \hspace{1cm} (4.8)

Figure 4.2(b) demonstrates this model.

Simple thresholding of extrema and removal of sharp edges using a gradient magnitude filter can now be applied to the box-mask which results in the segmentation of normal peripheral zone tissue, as demonstrated in figures 4.2(c) and 4.2(d).
4.2.3 CADx performance evaluation

The features $K^{\text{trans}}$, $V_e$ and $\text{Washout}$ were computed and used to train a SVM as classifier [95]. The features were combined into a single malignancy likelihood estimate using the SVM. The output of the classifier was used as a measure of likelihood of malignancy. The discriminating performance of the CADx system was estimated by means of the area under the receiver operator characteristics (ROC) curve (AUC). The prospective performance of the lesion analysis with per patient and fixed calibration were estimated by means of leave-one-patient-out (LOPO) cross validation. LOPO avoids training and testing on the same data and to emphasize the prospective value, one whole patient case was drawn from the set. The LOPO involves training on all but one case, estimating the likelihood of that left-out case, and repeating the procedure until each case has been tested individually. The bootstrap technique was used to compute 95% confidence intervals for the AUC and significance level for the paired difference [86].

4.2.4 Experiment

The study set consisted of 38 consecutive patients that were selected between January 2007 and October 2008. These patients had biopsy-proven PCa and underwent dynamic contrast-enhanced MR imaging at 3.0T, complementary to the routine staging MR imaging examination of the prostate. Patients were included in the study only if they were candidates for radical retropubic prostatectomy within 6 weeks after MR imaging. The study was approved by the institutional review board, and informed consent was obtained from all patients prior to MR imaging. Exclusion criteria were: previous hormonal therapy, lymph nodes positive for metastases at frozen section analysis, contraindications to MR imaging (e.g., cardiac pacemakers, intracranial clips), contraindications to endorectal coil insertion (e.g., anorectal surgery, inflammatory bowel disease).

Images were acquired with a 3.0T whole body MR scanner (TrioTim, Siemens Medical Solutions, Erlangen, Germany). A pelvic phased-array as well as a balloon-mounted disposable endorectal surface coil (MedRad®, Pittsburgh, PA, USA) inserted and inflated with approximately 80 cm$^3$ of Perfluorocarbon (FOMBLIN LC08), were used for receiving. The machine body coil was used for RF transmitting. 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.

High-spatial-resolution T2-weighted fast spin-echo imaging in the axial, sagittal and coronal planes, covering the prostate and seminal vesicles, was performed. 3D T1-weighted spoiled gradient echo images were acquired before and during an intravenous bolus injection of paramagnetic gadolinium chelate (0.1 mmol/kg, gadopentetate, Magnevist®; Schering, Berlin, Germany) using a power injector (Spectris, Medrad®, Pittsburgh, PA, US) with an injection rate of 2.5 ml/second followed by a 15 ml saline flush for 300 sec every 3 seconds. Fitting the DCE-MRI is described elsewhere [45].

Whole-mount step-section histology tumor maps were used as ground truth for annotating PCa (with a relevant diameter of at least 5mm), non-malignant suspicious enhancing (NS) and normal (N) regions on T2-w images for all patients in consensus by two readers.

4.3 Results

One patient case was excluded because the DCE examination had failed. In total 42 malignant regions were annotated in the peripheral zone. The number of NS regions annotated in the peripheral zone was 29. The number of normal peripheral zone regions was 38.

The effect of the per patient calibration on the diagnostic performance was first evaluated by pairwise scatterplots of PK parameters of the lesions. It is noticeable in figure 4.3(a) that without calibration the clusters overlap more than the clusters in figure 4.3(c) where manual calibration is included. Figure 4.3(b) shows similar results for automatic patient calibration. Furthermore, the $N$ and $NS$ clusters have a smaller covariance when patient calibration is used. An effect on the diagnostic performance can therefore be expected. The different distributions demonstrate the strong effect of the chosen calibration method.

The performance of discriminating malignant lesions from NS areas with fixed, manual and automatic calibration is demonstrated in the ROC curves shown in figure 4.4. Here, the focus is on the characterization of NS and malignant regions, because it is more challenging and clinically relevant. The diagnostic
Automated calibration for computerized analysis of prostate lesions using PK-MRI

Figure 4.3: Pairwise scatterplots of 2 kinetic parameters, $W_{ashout}$ versus $K_{trans}$, for the whole database with squares representing NS regions, spheres as malignant regions and triangles as N regions for the different calibration methods used. The ellipsoids summarize the three clusters by fitting a bivariate normal distribution and displaying the outline at 2 times standard deviation radius. It is noticeable that the clusters overlap one another when fixed calibration is used, whereas manual and automated per patient calibration demonstrate a noticeable clustering of features.

accuracy was 0.65 (95% confidence intervals = 0.54-0.76) when fixed calibration was used. The diagnostic accuracy improved significantly for both manual per patient calibration, $Az=0.80$ (0.70-0.90), as for automated per patient calibration, $Az=0.80$ (0.71-0.88). The marginal difference between the automated and manual calibration means that they perform similar, which was the intended goal.

Figure 4.4: ROC curves showing the discriminating performance of the CADx system using the different calibration methods fixed, automated per patient and manual per patient calibration.

4.4 Conclusion

In this study, we have demonstrated the feasibility of an automated calibration method for estimating the arterial input function when calculating pharmacokinetic parameters from DCE-MRI. The results show a significant better discriminating performance ($Az=0.80$ (0.71-0.88)) compared to the conventional fixed calibration. The performance is similar to using the manual per patient calibration.
Chapter 5

Automatic Parametric Multivariate Multi-object Segmentation of the Prostate in MR images Using Prior Knowledge

This chapter is based on the manuscript “Automatic Parametric Multivariate Multi-object Segmentation of the Prostate in MR images Using Prior Knowledge.” by Pieter C. Vos (MSc), Jelle Barentsz (PhD, MD), Nico Karssemeijer, Henkjan Huisman (MS, PhD) Submitted to IEEE Transactions on Medical Imaging, 2011.
Abstract

Automatic prostate segmentation is a crucial step for Computer-Aided Detection (CAD) systems to reduce the number of false positive cancer candidates. Fully automatic prostate segmentation in MR images is challenging, mainly due to the lack of well-defined edges, similar intensity profiles with surrounding organs and its relative small size in the male pelvis. In this paper, a fully automatic parametric multivariate multi-object segmentation method is presented. A parametric model describing the prostate and surrounding anatomical structures is first fitted to multiple MR images simultaneously. The obtained anatomical information about structure shape, position and appearance is then used as prior knowledge in a Bayesian framework to classify the voxels of the multivariate MR images and to segment the prostate. The segmentation performance is compared to the results obtained when (1) information from a single MR image is used and (2) pelvic model fitting is not performed as a first step. The results show that a multivariate segmentation including fitting significantly outperforms the aforementioned approaches with a dice coefficient index of 0.77±0.05. To the best of our knowledge, this is the first time that a method which fits a parametric multi-object model to multivariate MR images is presented for prostate segmentation.
5.1 Introduction

Prostatic cancer (PCa) is the most commonly diagnosed cancer among American men and remains the second leading cause of cancer related deaths among males [120]. Prostate Specific Antigen (PSA)-based screening programs may reduce mortality from prostate cancer with 20% to 30% but they are associated with overdiagnosis and overtreatment of indolent disease [8, 9]. Prostate Magnetic Resonance Imaging (MRI) can be used in PSA based screening scenarios in order to reduce the rate of overdiagnosis and overtreatment. Several studies showed that combining anatomical, functional and metabolic MRI information leads to a sufficient high PCa localizing and detection accuracy of up to 92% [11, 12, 13, 14, 15]. The reported accuracy of prostate cancer detection on MR images varies widely, and controversy persists regarding recommendations for the use of MR imaging in the initial diagnosis of high-risk patients or in patients with previous negative biopsy findings but persistently high PSA level (62). Despite the promising results, it has been shown that the accuracy is highly dependent on the experience of the radiologist and the interpretations suffer from observer variability [16].

Computer aided detection (CAD) systems can help to improve the diagnostic accuracy of the radiologist. CAD aims to automatically highlight cancer suspicious regions, leading to a reduction of search and interpretation errors, as well as the variation between and within observers [49]. However, for CAD to be useful in a prostate screening environment, the following requirements are essential as well as challenging: it should be fully automated; it needs to be robust to the large population variation; and it should be fast enough in a high volume screening setting.

Automatic prostate segmentation is a crucial step within the workflow of a CAD system. That is, it reduces a substantial number of false positive cancer candidates that are detected in the initial step. Automatic prostate segmentation in MR images has been a subject of considerable research as it is a difficult task to accomplish. Challenges include the presence of imaging artifacts due to air in the rectum and inhomogeneities of the magnetic field, the large anatomical variability between subjects, the differences in rectum and bladder filling, a lack of well-defined edges in the apex and base, similar intensity profiles with surrounding organs and its relative small size in the male pelvis. Most prostate segmentation methods in literature are single-object methods of which the majority only operate on univariate images, e.g., on a single MR image. In a single-object prostate segmentation, only properties of the prostate are modeled and do not include information from surrounding anatomy. As a result these methods are typically semi-automatic and are inaccurate at base and apex part of the prostate [121, 122, 123, 124].

Recently, multi-object segmentation has been adopted by several authors [125, 126, 127, 128]. The key advantage of this multi-object approach is that organs with low-contrast can be accurately segmented when nearby well-defined organs are taken into account. In [127], an atlas-based method was proposed for prostate segmentation in univariate (T2-weighted) images. Although good results were obtained, the method strongly depends on the training set used and can fail on unforeseen cases. Costa et al. [126] showed that statistical shape and appearance models (ASM) are a fast and robust approach. In their method, a coupled bladder and prostate segmentation in CT images is used to achieve an automatic prostate segmentation. The disadvantage of those models is that they strongly depend on correct landmark correspondence by manual segmentation in the training phase as well as the quality of the initialization [129, 128].

Both approaches require large training efforts and training databases in order to capture the high anatomical variability of the pelvis. An alternative approach is to model shape and appearance using parametric shape models. This technique does not require the existence of anatomical landmarks and requires fewer training samples. Although complex shapes can be modeled with many parameters controlling local deformations, it is desirable to reduce the number of parameters and the range of each parameter used in a model to obtain a compact representation [130]. In [131], the authors demonstrated that with a relatively simple shape model good prostate segmentation results were obtained. Prostate segmentation was achieved by first fitting a deformable superellipsoid model to the prostate in ultrasound images. Subsequently, the obtained shape information was integrated into a Bayesian prostate segmentation algorithm. However, initialization of the model was manual, the method was 2D, did not include information of surrounding structures and cannot handle multiple MR images simultaneously.

In this paper we present a fully automatic novel parametric multivariate multi-object method to segment the prostate. The first aim of the paper is to demonstrate how to construct a multi-object model with few, sufficiently independent parameters describing the absolute and relative shape as well as the appearance
Chapter 5

of pelvic structures. The second aim is to investigate if the supervised method can generalize on small training datasets. The performance of the complete method is analyzed by comparing to expert manual segmentations. To the best of our knowledge, this is the first time that a method which fits a multi-object parametric model to multivariate MR images is presented for prostate segmentation.

5.2 Method

The proposed method consists of two steps: (1) model building and fitting; (2) object segmentation with prior model constraints. In the first stage, a model is constructed with multiple parametric objects that define the shape and multivariate appearance of each anatomical structure. Next, the model is fitted to multivariate images simultaneously. Furthermore, a realistic anatomical structure representation is obtained by constraining the parameters within a population model. Relationships between the defined objects (e.g. position relative to other object) is captured beforehand by a transformation of the model parameters. This prevents the need of a large training database to capture correlation between parameters such they can be assumed independent, reduces the complexity of the model and maintains the intuitive parameter setup of the model. In the second stage, a Bayesian framework is used to obtain the final segmentation. Here, the fitted model is used as a prior to the Bayes classifier such that prior information about spatial and multivariate appearance relations between anatomical structures is efficiently taken into account.

5.2.1 Model Fitting

In this section, a method is described to fit the model $\Phi(\zeta)$ with parameters $\zeta$ to $M$ images $\{I_1, \ldots, I_j, \ldots, I_M\}$ from a patient. The model is fitted by finding the parameters that optimize the correspondence between model and images:

$$\zeta = \arg \max \zeta \{ F(I_\Phi(x; \zeta), I_j(x), \ldots, I_M(x)) \},$$

(5.1)

where $F(.)$ represents the objective function to be optimized, $I_\Phi(x; \zeta)$ a 3D synthetic image representation of $\Phi(\zeta)$ with same dimensions as $I_j(x)$ and $x = (x, y, z)$ denotes a voxel location in cartesian space.

Let the model $\Phi(\zeta)$ be composed of $N$ objects. Examples of objects are pelvic structures such as bladder, prostate, rectum or pelvic bone. The appearance of object $i$ in $I_\Phi$ is modeled with a vector $v_i$ of mean grey values $\{v_{i1}, \ldots, v_{ij}, \ldots, v_{iM}\}$, where $i = 1, \ldots, N$. The grey value $v_{ij}$ depends on the image $I_j$ as well as the object it represents. Additionally, the spatial distribution of an object $i$ is considered the same in the different images by means of image registration. A mutual information registration strategy was applied to correct for patient movement [132]. Assuming that any object $i$ in the model conforms to a multivariate Gaussian shape in $I_\Phi$, then the spatial distribution $\phi_i$ of object $i$ can be defined as a multivariate normal distribution:

$$\phi_i(x; c_i, \Sigma_i) = N(x; c_i, \Sigma_i)$$

$$= (2\pi)^{-\frac{k}{2}} |\Sigma_i|^{-\frac{1}{2}} \exp \left[ -\frac{(x - c_i)^T \Sigma_i^{-1}(x - c_i)}{2} \right],$$

(5.2)

where $c_i$ is the center of the object and $\Sigma_i$ defines its shape. For this study, it is assumed that objects cannot be rotated in space. The covariance matrix $\Sigma_i$ then becomes:

$$\Sigma_i = \text{diag}(\sigma_{i1}^2, \sigma_{i2}^2, \sigma_{i3}^2).$$

(5.3)

A full covariance matrix could be used in order to generalize the model for object rotation but at the expense of increasing the model complexity. An additional background model is defined with uniform spatial distribution $\phi_0$ defined as follows:

$$\phi_0 = \frac{1}{\sum_1^m 1},$$

(5.4)

where $m$ is the total amount of voxels in $I_j$.

The synthetic representation $I_\Phi(x; \zeta)$ is then defined as:

$$I_\Phi(x; \zeta) = \{ v_i; \text{if} L(x; \zeta) = i \},$$

(5.5)
where $L(x; \zeta)$ is a label image defined as:

$$L(x; \zeta) = \arg \max_{\mathbf{a}} \phi_{\mathbf{a}}(x; \mu, \Sigma).$$

(5.6)

In the label image, the voxels take discrete values $i$, each corresponding to an object in $\Phi(\zeta)$. Although individual objects are defined as an ellipsoid in $I_\Phi(x; \zeta)$, they become more realistic shaped in $L(x; \zeta)$ at the boundaries where the objects overlap each other. An example of $I_\Phi(x; \zeta)$ is shown Fig. 5.1.

![Figure 5.1: Cross-sectional sagittal views of an example synthetic image $I_\Phi(x; \zeta)$ simulating ADC and T2 maps. The scalar values are population averages of several anatomical structures visible in the ADC and T2 maps below. Window and level settings were set equally. Considering the synthetic T2 map, bladder, prostate and the rectum are displayed from left to right (Anterior to Posterior) with the rectum as the most dark anatomical structure. In the ADC map and the synthetic representation, the bladder and urine stand out as a bright anatomical structure.](image)

Fitting the model can now be considered an optimization problem, where the goal is to search for the values of the model parameters that best fit the multivariate images. Thus, the optimization problem defined in Eq. 5.1 needs to be solved using the parameter vector:

$$\zeta = \{c_1 \ldots c_n, v_1 \ldots v_n, \Sigma_1 \ldots \Sigma_n \}.$$ (5.7)

Let $F(.)$ be defined as a function consisting of a similarity term and a penalty term:

$$F(I_\Phi(x; \zeta), I_1(x); \ldots, I_M(x)) =$$

$$= \sum_{j=1}^{M} \omega_j S(I_\Phi(x; \zeta), I_j(x; \zeta)) - (1 - \sum_{j=1}^{M} \omega_j) C(\zeta),$$

(5.8)

where $S(\ldots)$ measures the similarity between $I_\Phi$ and $I_j$; $C(\zeta)$ is a penalty term to constrain $\zeta$ within the population distribution; $\omega$ is a weight to control the balance between the two components and $M$ a median filter with kernel $\kappa$ in order to reduce noise and preserve edges.

$C(\zeta)$ is a penalty term to constrain the parameter vector $\zeta$ to be within a population distribution. We define the penalty term as a multivariate Gaussian with mean and covariance estimated from a training population (explicit prior knowledge). In that way, a realistic fit of the model is obtained as spurious parameter values outside the population range are penalized more heavily.

---

**Figure 5.1:** Cross-sectional sagittal views of an example synthetic image $I_\Phi(x; \zeta)$ simulating ADC and T2 maps. The scalar values are population averages of several anatomical structures visible in the ADC and T2 maps below. Window and level settings were set equally. Considering the synthetic T2 map, bladder, prostate and the rectum are displayed from left to right (Anterior to Posterior) with the rectum as the most dark anatomical structure. In the ADC map and the synthetic representation, the bladder and urine stand out as a bright anatomical structure.
5.2.2 Transformation of the Parameter Vector

Spatial and appearance relations between the objects in the model are captured beforehand by redefining the parameter vector \( \zeta \) to \( \zeta' \). The transformation should reduce the complexity of the model and aims to reduce dependencies between the parameters. The transformation is firstly based on knowledge about topology, i.e. spatial relation between objects. Secondly, it is based on knowledge on the appearance relation between objects. Applied to the pelvis, the coordinate system of the model is placed in the center of the prostate.

Assuming that objects can only move in the sagittal and coronal plane with respect to the prostate, two parameters need to be calculated for each object: the distance \( d_i \) from the object to the prostate center and the angle \( \alpha_i \) in the sagittal plane. Next, the object intensity vector \( v'_i \) is calculated with respect to the intensity vector of the prostate as follows: \( v'_i = v_{pr} - d v_i \), where \( v_{pr} \) is the intensity vector of the prostate and \( d v_i \) is a vector of shift values for the object \( i \).

5.2.3 Anatomical Structure Segmentation

This section will describe the Bayesian framework used to obtain the final anatomical structure segmentation. The fitted model is incorporated into the segmentation framework as prior knowledge about shape and appearance of the objects in the model. Under the maximum a posteriori (MAP) criterion, the optimal segmentation field is the result which maximizes the posterior probability:

\[
\hat{\ell}_x = \arg \max_i \{ p(\ell_x = i | I_1(x), \ldots, I_M(x)) \} ,
\]

(5.9)

where \( \hat{\ell}_x \) is the optimal segmentation result under the MAP criterion, \( \ell_x \) is the label of voxel \( x \) and \( p(\ell_x = i | I_1(x), \ldots, I_M(x)) \) is the probability of \( \ell_x \) to belong to class \( i \) given the intensity values \( I_1(x), \ldots, I_M(x) \) of the voxel \( x \) in the images. Using the Bayes rule, \( p(\ell_x = i | I_1(x), \ldots, I_M(x)) \) can be rewritten as:

\[
p(\ell_x = i | I_1(x), \ldots, I_M(x)) \propto p(I_1(x), \ldots, I_M(x) | \ell_x = i)p(\ell_x = i).
\]

(5.10)

In the proposed method, the anatomical structure segmentation is achieved by incorporating the estimated model \( \Phi(\zeta) \) as a priori knowledge. Assuming that the grey values of the images within the objects are independent, \( p(I_1(x), \ldots, I_M(x) | \ell_x = i) \) can be formulated as:

\[
p(I_1(x), \ldots, I_M(x) | \ell_x = i) = \prod_{j=1}^M p(I_j(x) | \ell_x = i).
\]

(5.11)

Assuming further that the intensity distribution of each object can be modeled as a Gaussian distribution, then the likelihood function \( p(I_j(x) | \ell_x = i) \) is defined as:

\[
p(I_j(x) | \ell_x = i) = \mathcal{N}(I_j(x); \hat{v}_j^i, \sigma_j^2),
\]

(5.12)

where \( \hat{v}_j^i \) is the estimated intensity value of object \( i \) in \( I_j \) obtained after fitting the model and \( \sigma_j^2 \) is the corresponding standard deviation obtained from training.

The prior probability \( p(\ell_x = i) \) in Eq. 5.12 is defined as the estimated spatial distribution of the object \( \phi_i \) obtained after fitting:

\[
p(\ell_x = i) = \phi_i(x; \hat{e}_i, \hat{\Sigma}_i),
\]

(5.13)

The final segmentation \( B(x) \) is then achieved as follows:

\[
B(x) = \begin{cases} 
1, & \text{if } \hat{\ell}_x = i_{as} \\
0, & \text{otherwise}
\end{cases}
\]

(5.14)

where \( i_{as} \) is the index corresponding to the anatomical structure that is to be segmented.
5.3 Data

The evaluation of the proposed method was performed on a prostate multivariate MR dataset acquired from a cohort of 55 clinical patients. These patients had elevated PSA levels and one negative biopsy; and were scheduled for a prostate cancer MR detection between January and April 2009 at the RUNMC radiology department. Images were acquired with a 3.0T whole body MR scanner (TrioTim, Siemens Medical Solutions, Erlangen, Germany) using a pelvic phased-array surface coil. For our experiments, MR data from each patient comprised a T2-weighted (T2-w) axial volume (with dimensions 256x256x15 and voxel size 0.75mmx0.75mmx4mm), a proton density-weighted (PD-w) volume (with dimensions 128x128x12 and voxel size 1.8mmx1.8mmx4mm) and an ADC map (with dimensions 136x160x10 and voxel size 1.625mmx1.625mmx3.6mm) that was generated by the MR scanner. The T2-w and PD-w volumes were used to generate a T2 map, as explained in [132].

The collected patient data was randomly divided into two groups. The first group consisted of prostate MR data from 40 patients and was used as a training set for the proposed method. The remaining 15 patients were used to evaluate the prospective segmentation performance.

For evaluation, each prostate was manually annotated in the T2-w images by a radiologist and researcher in consensus. The T2-w images were used because they usually have high spatial resolution and provide the most accurate prostate contours. The annotation was performed using an in-house developed software in which the user employed a brush tool to select prostate pixels in each MR slice. The annotation started at the MR slice where the transition of the external sphincter to the prostate apex was visible and ended at the upper part of the transition zone. Additionally, the model was manually fitted to the T2-w images to obtain prior knowledge of the optimal model parameter setting. The optimal parameter settings were used to determine the population mean and covariance.

5.4 Settings and Experiments

5.4.1 Settings

ADC maps and generated T2 maps were used as input for the segmentation method \((M = 2)\). The T2 maps were registered and resampled to the image space of the corresponding ADC maps [133]. To make the objects in the images homogeneous and suppress noise, the kernel size of \(M\) was empirically set to 5x5x3 voxels. The mean of squared differences was used as similarity measure \(S(\ldots)\). The weights \(\omega_j\) in the objective function \(F(.)\) (see Eq. 5.8), were empirically set to \(\omega_1 = \omega_2 = 0.1\).

The L-BFGS-B [110], a limited-memory, quasi-Newton minimization package was used as the optimizer in Eq. 5.1 until termination criteria were satisfied. The limited-memory method is well suited for optimizing objective functions with a large number of parameters and provides an additional advantage in that it allows bound constraints on each parameter. Minimum and maximum values of the model parameters can be controlled individually. For all experiments, each parameter was hard constrained to be within mean \(\pm 2\sigma\) of the population distribution. Methods that are Newton based require a gradient estimate during the optimization. In this study the common finite-difference approximation approach is used.

5.4.2 Evaluation Measures

The segmentation results were evaluated by voxel-based comparing the automatically generated prostate segmentations with the manual annotated ground truths. As performance measure, the Dice similarity coefficient (DSC) was used [134]:

\[
DSC = \frac{2TP}{TP + FP + TP + FN},
\]

where \(TP\), \(FP\) and \(FN\) are, respectively, the number of true positive voxels, false positive voxels and false negative voxels. A DSC value of 1 indicates a perfect overlap, whereas a value of 0 means no overlap at all.
Table 5.1: *A-PRIORI* knowledge of the model parameters that was obtained from literature about pelvic anatomy. The relative parameter values are calculated with respect to the corresponding prostate value.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Parameter Description</th>
<th>Value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{\text{v}11}$</td>
<td>Bladder ADC relative intensity value (mm²/sec)</td>
<td>$0.7 \pm 0.24 \times 10^{-3}$</td>
<td>Reference [136]</td>
</tr>
<tr>
<td>$v_{\text{v}11}$</td>
<td>Prostate ADC intensity value (mm²/sec)</td>
<td>$1.57 \pm 0.09 \times 10^{-3}$</td>
<td>Reference [136]</td>
</tr>
<tr>
<td>$d_{\text{v}31}$</td>
<td>Rectum ADC relative intensity value (mm²/sec)</td>
<td>$-0.77 \pm 0.09 \times 10^{-3}$</td>
<td>Reference not found. The value was chosen based on the radiologist’s experience.</td>
</tr>
<tr>
<td>$d_{\text{v}12}$</td>
<td>Bladder T2 relative intensity value (ms)</td>
<td>$20.0 \pm 10$</td>
<td>Reference not found. The value was chosen based on the radiologist’s experience.</td>
</tr>
<tr>
<td>$v_{\text{v}22}$</td>
<td>Prostate T2 intensity value (ms)</td>
<td>$91.0 \pm 13.3$</td>
<td>Reference [114]</td>
</tr>
<tr>
<td>$d_{\text{v}32}$</td>
<td>Rectum T2 relative intensity value (ms)</td>
<td>$-40.0 \pm 10$</td>
<td>Reference not found. The value was chosen based on the radiologist’s experience.</td>
</tr>
<tr>
<td>$a_{\text{i}}$</td>
<td>Bladder angle (Degrees)</td>
<td>$114 \pm 9.5$</td>
<td>Reference not found. The value was chosen based on the radiologist’s experience.</td>
</tr>
<tr>
<td>$c_{\text{v}1}$</td>
<td>Bladder Center (X,Y,Z) (mm)</td>
<td>$(0 \pm 10.0, 0 \pm 7.5, 0 \pm 4.5)$</td>
<td>Origin. The standard deviation was based on the radiologist’s experience.</td>
</tr>
<tr>
<td>$d_{\text{v}3}$</td>
<td>Rectum relative distance (mm)</td>
<td>$24 \pm 4.5$</td>
<td>Reference not found. The value was chosen based on the radiologist’s experience.</td>
</tr>
<tr>
<td>$o_{\text{a}}$</td>
<td>Rectum angle (Degrees)</td>
<td>$-18 \pm 4.5$</td>
<td>Reference not found. The value was chosen based on the radiologist’s experience.</td>
</tr>
<tr>
<td>$S_{\text{i}1}$</td>
<td>Bladder volume (cc)</td>
<td>$179 (42–582)$</td>
<td>Reference [135]. The bladder radius is computed from the ellipsoid volume formula: Volume = height x width x length x 0.52.</td>
</tr>
<tr>
<td>$S_{\text{v}2}$</td>
<td>Prostate volume (cc)</td>
<td>$55 (21–154)$</td>
<td>Reference [137]. The prostate radius is computed similar to the bladder radius.</td>
</tr>
<tr>
<td>$S_{\text{v}3}$</td>
<td>Rectum volume (cc)</td>
<td>$108 (28–223)$</td>
<td>Reference [137]. The rectum radius is computed similar to the bladder radius.</td>
</tr>
</tbody>
</table>

5.4.3 Experiments

In this paper, three experiments were carried out to evaluate the proposed method. In the first experiment, the effect of fitting the model to MR images and training size were evaluated. The segmentation performance was measured twice: firstly after setting the model parameters equal to a population mean that was obtained from training; secondly after fitting of the model was performed and the estimated parameters were set to the model. For each scheme (with and without fitting), a group of $T$ patients from the pool of 40 training cases were randomly selected to train the proposed method, where $T \in \{2, 5, 10, 15, 20, 25, 30, 35\}$. To reduce the influence of the random selection, the experiment was repeated 10 independent times and the results were averaged at each training size $T$. The influence of training size, training selection and fitting on the segmentation performance was analyzed by means of linear modeling. Segmentation performances were compared using ANOVA [135] with the Statistical Package for the Social Sciences, Version 15 (SPSS, Inc; Chicago, Illinois).

In the second experiment, the additional value of the multivariate approach ($M > 1$) was evaluated. The prostate segmentation performance was compared to the results achieved using a univariate approach ($M = 1$), using only the ADC map. Similar to the previous experiment, for both schemes (multivariate and univariate), a group of $T$ patients from the pool of 40 training cases were randomly selected to train the method. This experiment was also repeated 10 independent times. The segmentation results were averaged at each $T$ and compared using ANOVA.

In the last experiment, the effect of the selection of prior knowledge on the segmentation performance was investigated. For this purpose, we included prior knowledge that ranged from general to more specific. The performance was first evaluated when the proposed method was initialized using *a-priori* knowledge that was not obtained from a training population. This *a-priori* knowledge was acquired from literature, as it is summarized in Table 5.1, providing a general information of the population distribution. Parameters of the model can be set intuitively. Therefore, parameters that could not be obtained from literature were set by a radiologist based on experience. The result was then compared to the performance obtained in the first experiment, in which the method was initialized using *a-priori* knowledge obtained from a training set. In this case, the selection of *a-priori* knowledge is more specific to our clinical environment. The results were compared using ANOVA and a Bonferroni correction was applied.
5.5 Results

The results of the first experiment are shown in Fig. 5.2(a). The prostate segmentation performance using fitting outperforms significantly (p<0.001) the results obtained when fitting was not performed. Linear modeling showed that the segmentation performance was not influenced by training selection. A trend is visible in both approaches for which using fitting requires less training (no significant improvement at training size $T = 10$) compared to without fitting (performance is not significantly different at training size $T = 20$).

Fig. 5.2(b) shows the results of the second experiment. In this case, the segmentation performance was significantly improved when a multivariate approach was adopted.

Table 5.2 summarizes the DSC and the corresponding 95% confidence intervals of the third experiment. Fig. 5.3 and Fig. 5.4 show examples of the final prostate segmentation. The example of Fig. 5.4 is a difficult case because of tumor in the ventral part of the prostate, which changed the intensity value in both T2 and ADC maps. The univariate method clearly fails as it depends too much on the single variate grey value, whereas the multivariate approach succeeds by combining the two appearances.

Manually fitting the model to the patient data provides insights on the most accurate segmentation achievable which resulted in an average DSC of 0.84.

The average computation time for fitting the model to the T2 and ADC maps and segmenting the prostate was 3.2 minutes (range: 2 to 4 minutes) on a Intel Core Duo CPU 2.33GHz.
Table 5.2: Dice similarity coefficients and the corresponding 95% confidence intervals for the proposed segmentation method with and without fitting using prior knowledge from different sources, which ranged from general to more specific. * indicates the performance is significantly different compared to the result obtained using fitting and training size \( T = 20 \). † indicates the performance using literature as prior information is significantly different compared to the result obtained using a training size \( T = 2 \). Bonferroni correction was applied.

<table>
<thead>
<tr>
<th>Prior knowledge source</th>
<th>Literature ( (T = 2) )</th>
<th>Training ( (T = 20) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>without fitting</td>
<td>0.612±0.062*†</td>
<td>0.731±0.033*</td>
</tr>
<tr>
<td>with fitting</td>
<td>0.734±0.045*</td>
<td>0.734±0.033*</td>
</tr>
</tbody>
</table>

Figure 5.3: Example of pelvic segmentation. All views have a color coded overlay with red representing bladder, blue represents prostate and green the rectum. (a) ADC map of the pelvic in a transversal view. The prostate is successfully segmented by not including loose connective and adipose tissue containing the periprostatic venous plexus and neurovascular components that separates muscles from the prostatic capsule. (b) Corresponding T2-w image of the pelvic in a transversal view. (c) Corresponding T2-w image of the pelvic in a sagittal view. The hole in the bladder segmentation is due to very high ADC values of the urine at that location.

5.6 Discussion

In this paper, we have presented a novel fully automatic parametric multivariate multi-object segmentation method and applied it to a cohort of prostate MRI data acquired during prostate cancer screening. The method can be generalized to any modality and application, e.g. combining ultrasound and MR images. Additionally, the amount of objects, their shapes and appearances can be easily modified in the parametric model. Moreover, the proposed model allows a general and intuitive setup, with clear object definitions, which makes it easy to tune and update.

The proposed method achieved good segmentation performances on small training datasets. Fig. 5.2(a) demonstrates that an optimal segmentation performance was obtained using 10 training cases only. The good results were achieved by fitting the model to the patient under study, which provides a better spatial distribution of the pelvic structures for the Bayesian framework. Table 5.2 supports the idea that the nature
of this \textit{a-priori} information has an important effect on the final results. Using only information obtained from a different population distribution, such as the one extracted from literature, deteriorated the results when the method was applied to the provided test cases. The segmentation performance increased when more specific information was included using training samples from the same clinical environment. By fitting the model and thus making the \textit{a-priori} knowledge more patient specific, the segmentation performance significantly improved. Table 5.2 also shows that including fitting reduces the dependency on a large training size, because a maximum segmentation performance was obtained with less training samples. This fact shows that fitting is of paramount importance for the final results. The use of this patient-specific information can overcome the limitation encountered for other authors when the training set was not representative enough to deal with the large variability of prostate appearance and shape [127, 131, 128].

Information obtained from multivariate image helps to increase the segmentation performance compared to a univariate approach. Similar to the workflow of a radiologist, we hypothesized that using the appearance of objects in multivariate MR images will improve the segmentation performance. The results support this assumption as demonstrated in Fig. 5.2(b): using a T2 and ADC map in a multivariate approach significantly outperforms the approach where only the ADC map was used. When a univariate approach is used, anatomical structures that are adjacent and with similar intensity may be segmented as one object. Adding information from another image, where the intensity ranges for those anatomical structures differ, makes the segmentation accuracy improve. Therefore, combining information from multivariate MRI helps to overcome the drawbacks that each image presents separately. Fig. 5.4 illustrates the idea that using information from multivariate MRI will improve the segmentation performance.

The obtained results for prostate segmentation were slightly lower than those reported previously [138, 127, 123]. However, a direct comparison is not possible due to the lack of publicly available multivariate MR data. The reported studies have a strong focus on radiotherapy, where bladder and rectum are considered as the organs-at-risk that should be protected against high dosage of radiation during treatment of prostate cancer. Therefore, their aim is to have an accurate segmentation at prostate boundaries at the cost of this \textit{a-priori} information.
of computation time. The focus of this study, however, was on the development of a segmentation method that can be used in a prostate screening environment. In such an environment, prostate size, shape and appearance are highly variable as it consists of healthy, benign and tumor patients. To capture this with atlas-based and ASM methods would require large training efforts. It was demonstrated that the proposed method obtained already good segmentation performances with small training efforts. Additionally, the model generalized with good segmentation results without being influenced by the chosen training set.

In [128], it is demonstrated that the use of a spatial constraint increases the robustness of the deformable model comparatively to a deformable surface that is only driven by an image appearance model. The spatial constraint was obtained by registering a pelvic atlas to the target image. The authors discussed that most serious errors occur in prostates which are dissimilar to the atlas. They hypothesized that using a multi-atlas approach can improve the segmentation performance. This would come, however, at the costs of large training efforts. We believe that our proposed method can be used in a similar fashion, as the method can be used instead of an atlas based method for generating the spatial constraint. This can potentially reduce the training effort needed to capture the wide range of shape sizes.

Despite the encouraging results achieved by the proposed method, we have identified a number of enhancements. Firstly, computation time can be reduced in the fitting stage. The optimization method requires derivatives for each parameter which were calculated using the common finite-difference approximation. The common finite-difference approximation is computational expensive and sensitive to noise. Using an analytic gradient would make the algorithm potentially faster and more robust. Secondly, more robustness can be obtained by including more structures in the model and by including more images. For example, adding the muscles that surround the prostate or including a T1 map could provide an improved segmentation accuracy and robustness. However, extending the model or adding more multivariate images is at the expense of more computational effort. Thirdly, anatomical structure segmentation can be performed more accurately by extending their object definition. For example, the prostate is modeled as a homogeneous region which in fact consists of several zones. Extending the prostate model in different zones with their own shape and appearance properties would make the segmentation method potentially more accurate. However, this can result in more parameters of the model which leads to more computational effort. Fourthly, similarity between the synthetic image representation of the model and the multivariate images is calculated using the mean of squared differences. The appearance of the anatomical structures that are modeled in the synthetic image representation are assumed to be homogenous. Typically, MR images that are T1-weighted or T2-weighted show a coil profile such that this assumption does not hold. The method that was presented in [133] was used to remove the coil profile. However, choosing a different similarity measure could provide similar results such that the correction is not needed.

In conclusion, a fully automatic novel automatic parametric multivariate multi-object segmentation method was presented and applied to the task of segmenting the prostate. The proposed method generalized good on small training datasets and was not influenced by training selection.
Chapter 6

Automatic Computer Aided Detection of Prostate Cancer Based on Multiparametric Magnetic Resonance Image Analysis

This chapter is based on the manuscript “Automatic Computer Aided Detection of Prostate Cancer Based on Multiparametric Magnetic Resonance Image Analysis” by Pieter C. Vos, Jelle O. Barentsz, Nico Karssemeijer and Henkjan J. Huisman Submitted to Physics in Medicine and Biology, 2011.
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 Coefficients 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 free response operating characteristic (FROC) methodology. The results show that the CAD method obtained sensitivities of 0.41, 0.65 and 0.74 at false positive levels of 1, 3 and 5 per patient, respectively. In conclusion, this study showed that it is feasible to automatically detect prostate cancer at an acceptable false positive rate. The CAD method can assist the radiologist to detect all prostate cancer locations and could potentially guide biopsy towards the most aggressive part of the tumour.
6.1 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 [1]. Early detection of PCa can be life saving [139]. 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 [8, 9].

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 levels 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 techniques involves systematic sampling of multiple areas in the prostate during TRUS-guided biopsy regardless of the presence of hypoechoic lesions. Over 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. 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 [140, 141]. To prevent patient anxiety, more accurate methods need to be found to detect or rule out significant disease [142].

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% [11, 12, 13, 14, 15]. Furthermore, multiparametric MRI can target biopsies towards regions determined to be suspicious of cancer [143]. Unfortunately, prostate MRI analysis requires a high level of expertise and suffers from observer variability [16]. 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 [17].

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 [49]. CAD research has been successfully pursued in other diagnostic areas such as mammography [18, 19], CT chest [144, 21, 22], CT colonography [23, 24] as well as retinal imaging [25]. CAD systems generally consist of multiple sequential stages. In the initial stage, lesion candidates are selected within a likelihood map that was generated by 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 false positives that were localized in the initial stage.

Most prostate CAD researchers have focused on the initial voxel classification stage [50, 51, 52, 53, 54, 55]. They obtained likelihood maps by combining information from multiparametric MR images using mathematical descriptors. These studies showed on a voxel basis that the discrimination between benign and malignant tissue is feasible with good performances. Recently, we presented work that focused on the regional classification stage [132]. 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 system 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 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 screening population that included patients with elevated PSA levels. The ultimate goal was to detect all tumors regions at an acceptable false positive rate to potentially guide biopsy towards the most aggressive part of the tumor.
Figure 6.1: Dataflow diagram of the implemented CAD system. A Hessian based filter initially detects possible lesion candidates at multiple scales. 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.

6.2 Method

6.2.1 Overview

The proposed CAD method is schematically outlined in figure 6.1. 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 Coefficients (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 section 6.2.

6.2.2 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 less diffuse property of PCa in a 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 $k = 1, 2, 3$ [119]. The likelihood of a voxel at scale $\sigma$ to belong to a blob is defined by [145]:

$$P(x, \sigma) = \begin{cases} \frac{\lambda_{\sigma,1}(x) \lambda_{\sigma,1}(x)}{|\lambda_{\sigma,3}(x)|} & \lambda_{\sigma,k}(x) < 0 \quad k \in 1, 2, 3 \\ 0 & \text{otherwise} \end{cases}$$  \hspace{1cm} (6.1)

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 ranging from 8 mm to 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:

$$L(x) = \max_{\sigma_{\min} \leq \sigma \leq \sigma_{\max}} L(x, \sigma).$$  \hspace{1cm} (6.2)
6.2.3 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 [146]. 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 multiple MR images, e.g., ADC and T2 maps. 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. Here, the fitted model is used as a prior to the 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 labeled with the corresponding organ. An example of the segmentation method is shown in figure 6.2.

![ADC map and T2 map](image)

**Figure 6.2:** Cross-sectional transversal views of multiple MR images that were used to segment the bladder (red), prostate (blue) and rectum (green) simultaneously.

6.2.4 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 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 neighborhood $\Omega$ with diameter $d$; and the difference between $L(x_i)$ and the mean neighborhood value should be more than the squelch threshold $\epsilon$. Let $\phi$ be the group of final selected candidates, then $x_i \in \phi$ when:

$$B(x_i) = 1 \wedge L(x_i) > \tau \wedge (L(x_i) - mean(\Omega(x_i), r)) > \epsilon.$$  \hspace{1cm} (6.3)

For this paper, the diameter $d$ was set to 5mm which represents the minimal size of a significant prostate tumour, $\epsilon$ and $\tau$ were empirically set to 0.1 and 1, respectively. An example of the candidate detection procedure is shown in Figure 6.3.

Candidates were removed when the corresponding MR values were outside biological meaningful thresholds. The following thresholds were obtained from literature: the ADC value was below 200 or above normal prostate intensity value of 1600 mm/s² [136]; the T2 intensity value was above normal prostate 100 msec or below muscle 36 msec [114]; and the interstitial volume $V_e$ was below normal prostate of 20 mmHg [81].
Chapter 6

Figure 6.3: Cross-sectional transversal views of an example ADC map of a patient with a tumour in the transition zone. Multiple dark blob like regions are visible in and outside the prostate. Lesion candidates are detected within the prostate segmentation. The red arrow indicates the location of a tumour with Gleason grade 3+4 which was detected by the peak detector with the highest likelihood of all detected blobs.

6.2.5 Local feature analysis

Prostate cancers can be discriminated from benign abnormalities by their heterogenous nature. That is, hotspots are visible in multiparametric MR images at different locations within the extent of the tumour. Because lesion segmentation methods typically work on a monoparametric MR image 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 6.4 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 lesion candidate $i$ at location $x$, such that analysis can be performed on multiple MR images simultaneously.

Figure 6.4: 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.
Analysis was performed within region $R$ by combining histograms of intensity values from T2, pharmacokinetic, T1 and ADC maps with texture based features. Percentiles 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 practice at our hospital, where prostate cancer diagnosis is performed on a daily basis [11]. Based on the preferences in our clinic, we expect the following intensity features as most descriptive:

$f_1 \text{ 25 \% percentile T2}:$

The 25 \% percentile was extracted from the T2 map as it has been established that prostate cancer typically demonstrates lower signal intensity than normal prostate tissue on T2-weighted MR images [147].

$f_2 \text{ 25 \% percentile ADC}:$

The 25 \% percentile was extracted from 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 [148, 149, 47].

$f_3 \text{ 75 \% percentile } K_{\text{trans}}:$

The pharmacokinetic parameter $K_{\text{trans}}$ or transfer constant (1/min) relates to permeability surface area. The permeability (or leakiness) surface area 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 tumours. An increased capillary permeability is observed in prostate cancer [73]. The upper quartile captures the presence of hotspots.

$f_4 \text{ 75 \% percentile } V_e:$

In the extravascular, extracellular space (EES) of normal tissue, pressure is near atmospheric ($25 mmHg$) values, whereas in tumours 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 tumour itself. This results in an increase of the EES. The EES, defined as percentage per unit volume of tissue, is therefore considered a descriptive parameter. An increased interstitial leakage space is observed in tumour, hence the upper quartile is used to capture these hotspots [81].

$f_5 \text{ 25 \% percentile } Wash - out:$

The kinetic parameter $Wash - out$ quantifies the slope of the curve after the first wash-in phase. Although it does not directly correlate to physiological parameters, like pharmacokinetic parameters, the presence of washout 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 tumor [73].

$f_6 \text{ 50 \% percentile T1 map}$

Post-biopsy hemorrhage mimics high tumour vascularity. Fortunately, hemorrhage is clearly visible as a high-intensity area on a T1-w image. As biopsy hemorrhage is often visible as a large homogeneous area, the 50 \% percentile is extracted from the T1 map.

The following texture features were extracted:

$f_7 \text{ Peak value}:$

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

$f_8 \text{ Mean neighborhood value}:$

The mean neighborhood value of $\Omega(.)$ at location $x_i$.

$f_9 \text{ Squelch value}:$

The squelch value is the difference between the peak value $f_7$ and the mean neighborhood 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 [132].

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

### 6.2.6 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 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 [135] are independently selected and a Linear Discriminant Analysis (LDA) classifier [150] is trained to remove spurious candidates from the data. Note that the FD ratio analyzes 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. The threshold is set such that no true positives are lost in the first phase for the training set. Before performing the feature selection, the feature values are transformed using the Box-Cox transformation [151] in order to approximate the feature distribution to a normal distribution.

In the second stage, a classifier is trained using a selection of features from the complete feature set. After pilot experiments with several classifiers, an LDA classifier was chosen in favour of k Nearest Neighbour [152] classifier, Quadratic Discriminant Analysis (QDA) classifier [150] and Support Vector Machines [80]. The selection of features is carried out by Sequential Forward Floating Selection (SFFS) [153] 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 6.2 summarizes the selected features in order of selection.

### 6.3 Data and Experiments

#### 6.3.1 Data

Imaging data was 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). A pelvic phased-array surface coil was used for receiving. The machine body coil was used for RF transmitting. 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, MR data from each patient comprised a T2-weighted (T2-w) axial volume (with dimensions 256x256x15 and voxel size 0.75mmx0.75mmx4mm), a proton density-weighted (PD-w) volume (with dimensions 128x128x12 and voxel size 1.8mmx1.8mmx4mm), contrast enhanced 3D T1-weighted (T1-w) spoiled gradient echo images and an ADC map (with dimensions 136x160x10 and voxel size 1.625mm x1.625mm x3.6mm) that was generated by the MR scanner. Additionally, gadolinium chelate concentration curves were calculated and dynamic contrast enhanced derived parameter maps ($K_{\text{trans}}$, $V_z$, $\text{Wash-out}$) were generated at a dedicated workstation [96]. All MR data was automatically normalized such that quantitative assessment was possible for the pharmacokinetic, T1 and T2 map using a previously presented method [133, 132].

The reference standard for lesion localization and pathology was established by combining the findings of one experienced 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, that 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 labeled as low grade when histopathology confirmed a tumour of Gleason grade less than six. Regions were labeled as high grade tumours when histopathology confirmed a tumour of Gleason grade seven or more. When the Gleason score was six or when no histopathology was available and the radiologist indicated presence of tumour, the region was labeled as intermediate.

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

6.3.2 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 three-fold cross validation. Cross validation folds were obtained by drawing whole patient cases.

A tumour was considered as detected when the initial 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 false positives. The specificity was computed using those patients where no presence of PCa was found based on radiology reports, follow-up reports and MR biopsy outcome. In this way a false positive 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 false positive 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 analyze the improvement obtained by adding additional information from multiple MR images. The difference in performance between the initial voxel classification stage and the local feature analysis was evaluated by jackknife FROC analysis (JAFROC) [154]. In the third experiment, the detection performance for prostate cancer was analyzed taking into account the malignancy grade. Therefore, the detection performance was evaluated for all tumours, high grade tumours, intermediate tumours and low grade tumours. The results were compared to the performance of a method that detect malignant regions based on complete random guess (random detection based on average prostate size and tumor size). All detection performances were evaluated at false positive levels of 1, 2 and 5 per healthy patient.

6.4 Results

6.4.1 PSA Levels and Histopathological findings

The study set consisted of 200 consecutive patients (mean age, 60 years; range, 50–69 years). The mean prostate specific antigen level was 13.6 ng/mL (range, 1–58 ng/mL), mean Gleason score was 7.3 (range, 5–9). MRI was performed on average 3 weeks after 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 can-
cers 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).

6.4.2 CAD performance

The candidate detection step generated 6227 candidates of which 44 were true positives (TPs), resulting in a sensitivity of 92%. These candidates were used for training.

The results of the first experiment are shown in table 6.1. 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.

<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>

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 state two-stage classification approach in order of selection are summarized in table 6.2.

<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% $K_{trans}$</td>
</tr>
<tr>
<td>$f_1$</td>
<td>75% T2</td>
</tr>
</tbody>
</table>

Figure 6.5 shows the detection results for the second experiment. The results demonstrate that using information from multiple MR images significantly improves the detection performance ($P<0.05$).

The detection performances for the different tumour grades were measured at a false positive level of 1, 3 and 5 per patient. At these levels, the CAD method obtained a sensitivity of 0.48, 0.73, 0.88 respectively for the detection of high grade tumours. The sensitivity for detecting all malignant regions was 0.41, 0.65 and 0.74 respectively. Detecting intermediate grade tumour resulted in sensitivities of 0.41, 0.65, 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 6.6 demonstrates the detection results of the different datasets that were obtained by FROC analysis.

6.5 Discussion

In this paper, we have presented a novel fully automatic CAD method and applied it to a cohort of prostate MRI data acquired during PCA screening. This study showed that it is feasible to automatically detect PCa using information from multiple MR images simultaneously at an acceptable false positive rate.

The proposed method achieved a good performance for the detection of prostate cancer in a challenging cohort of patient MR data. The MR data was acquired during screening for prostate cancer for patients with elevated PSA levels. As a result, the patient database consisted of both healthy and prostate cancer patients.
Figure 6.5: FROC curve showing the detection performances of the CAD method for different selected features. The horizontal axis shows the number of false positive detections per healthy patient and the vertical axis shows the sensitivity that is achieved at this specificity level. The red dotted FROC curve corresponds to the detection performance when only the blob likelihood (f7) was used as feature. The blue solid FROC curve corresponds to the detection performance using the two-stage classification approach. 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. There is a significant improvement of the detection performance when using a the two-stage approach (P<0.05).

In figure 6.6 it can be observed that high grade tumours are detected with a sensitivity of 0.734 at only a false positive rate of 2.3 per patient. When detecting tumour of all grades, the sensitivity was 0.60. Although intermediate grades 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 high and intermediate tumours. The very difficult to detect low grades tumours were detected with a sensitivity of 40 with 2.3 FPs, which is still considerably better than randomly detecting PCa.

In figure 6.5, 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 then only using feature $f_7$ for classification, i.e., the peak value obtained after the initial voxel classification stage. An 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. They are, however, questionable in their relevance. 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 examination. 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 analyzed the detection performance using FROC methodology. Most studies
Figure 6.6: FROC curves showing the detection performances of the CAD method for different tumour grades. The horizontal axis shows the number of false positive 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.

Presented in literature evaluated a discriminating performance of malignant and benign voxels using ROC analysis [50, 51, 52, 53, 54, 55]. ROC analysis, however, misses information about the number of false positive candidates. Therefore, the method may have a high discriminative performance though presents many false positive candidates. This can have a negative influence on the detection performance of the radiologist. Furthermore, the studies presented in 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 reflects more a screening population as both benign and malignant patients were present in the database.

Noguchi et al. [32] demonstrated that 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 to potential to guide prostate biopsy towards the most aggressive and representative part of the tumour [143]. 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.

Limitation of this study was that the reference standard for some patients could not be accurately established, as follow-up studies were not yet performed and no biopsy was performed. The reference standard was for those patients was graded as intermediate which in fact could be different. Also, the patient data used to represent a screening population is 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 prostate cancer at an acceptable false positive rate. The CAD method can assist the radiologist to detect all prostate cancer locations and could potentially guide biopsy towards the most aggressive part of the tumour.
Chapter 7

Computer-aided diagnosis of prostate cancer using multiparametric 3T MR imaging: Effect on Observer Performance

This chapter is based on the manuscript “Computer-aided diagnosis of prostate cancer using multiparametric 3T MR imaging: Effect on Observer Performance” by Thomas Hambrock, Pieter Vos, Christina Hulsbergen-van de Kaa, Jelle Barentsz, Henkjan Huisman Submitted to Radiology, 2011.

Abstract

The purpose of the study was to determine the effect of computer-aided diagnosis (CAD) on observer performance in differentiating benign and malignant prostate lesions on multiparametric 3T MRI. 34 Consecutive patients with biopsy proven prostate cancer who received a multiparametric 3T MRI incl. T2-weighted, diffusion weighted (DWI) and dynamic contrast enhanced (DCE) imaging prior to radical prostatectomy, were retrospectively included. Six less-experienced and four experienced prostate radiologists were asked to characterize different cancer suspicious regions on multiparametric MRI, first without and subsequently with the use of in-house developed CAD software. The effect of CAD was statistically analyzed using a multiple reader, multi-case, receiver operating characteristic analysis as well as linear mixed-model analysis. In 34 patients, a total of 206 pre-annotated regions which included 67 malignant and 64 benign regions in the peripheral zone (PZ) and 19 malignant and 56 benign regions in the transition zone (TZ) were evaluated. Stand alone CAD had an overall area-under the receiver operating characteristic curve (AUC) of 0.90, in discriminating PZ lesions an AUC=0.92 and for the TZ an AUC=0.87. Overall, less-experienced observers (LEO) had a pre-CAD AUC of 0.81 which significantly increased to 0.91 (p=0.001) after CAD while for experienced observers (EO) the pre-CAD AUC was 0.88 which increased to 0.91 (p=0.17) after CAD. In discriminating PZ lesions, LEOs increased their AUC from 0.86 to 0.95 (p<0.001) after CAD while EOs showed an increase from 0.91 to 0.93 (p=0.13). Similar, for the TZ lesions, LEOs significantly increased their performances from 0.72 to 0.79 (p=0.01) after CAD and EOs from 0.81 to 0.82 (p=0.42). In conclusion, the addition of CAD significantly improved the lesion discriminating performance for LEOs who reached similar performances as EOs. The stand alone performance of CAD is similar to experienced observers.
Chapter 7

7.1 Introduction

Multiparametric MR imaging both at 1.5T and 3T, has proven its value in detection, localization and characterization of prostate cancer [12, 15, 13, 14]. Despite high-resolution imaging (e.g., when using an endorectal coil for MR imaging at 3 T), the signal intensities on anatomic T2-weighted MR images show overlap among prostate cancer, post-biopsy hematoma, benign prostatic hypertrophy, fibrosis and prostatitis. Therefore, functional imaging modalities like dynamic contrast enhanced (DCE), diffusion weighted imaging (DWI) and spectroscopic imaging have been added. The multiparametric approach has been proven to be the most useful for evaluation.

DCE MRI which utilizes high-temporal resolution scanning, allows both qualitative and quantitative estimations of the perfusion, capillary surface space, extravascular extracellular space, features which all change under the influence of angiogenic processes akin to neoplastic disease. Different techniques for compartment modeling, arterial input determinations and gadolinium concentration estimation have been studied to improve the objectivity and reproducibility of quantitative pharmacokinetic parameters [45, 155]. Reference tissue techniques have shown promising results in providing a more robust and accurate estimations [92, 93]. DWI which utilizes different gradient strengths to measure the restriction capabilities of free proton movement has increasingly been used in prostatic imaging not only for detection and localization of tumor but also for assessment of cancer aggressiveness [156, 157]. The apparent diffusion coefficient values calculated from the DWI allow a more objective quantitative assessment of the tissue micro-environment. However, despite the quantitative nature of pharmacokinetic DCE and ADC, prostate cancer analysis on multi-parametric MRI (MP-MRI) is still challenging, requires a high level of expertise and suffers from observer variability [158]. A need therefore exist to aid radiologists in improving their assessment of MP-MRI and to reduce inter-observer variability.

Computer assisted diagnosis (CAD) for radiological assessment of various malignancies including breast cancer[159], lung cancer [144, 21] and colorectal cancer [23] have been developed. Especially their value in supporting less-experienced radiologists in improving overall tumor detection capabilities have been emphasized [160, 161]. For prostate cancer, studies that incorporate various features on MRI have shown the feasibility of CAD to discriminate benign and malignant lesions [50, 52]. However, no CAD system has been evaluated yet in a reader study with radiologists to determine the effect on improving observer’s performances for characterization of prostate lesions on MRI.

The CAD performances reported previously are sufficiently high, that a positive effect on reader performance can be expected. The purpose of this study therefore was to clinically evaluate the effect of CAD on the ability to characterize prostate lesions for both less-experienced and experienced observers using information obtained from quantitative pharmacokinetic DCE parameters and ADC values on 3T MR imaging.

7.2 Materials and Methods

7.2.1 Study population

Between January 2008 and January 2009, 50 consecutive patients with biopsy proven prostate cancer scheduled for radical prostatectomy, were referred from the department of urology at the Radboud University Nijmegen Medical Centre (RUNMC). All patients received a clinically routine MRI for tumor localization and staging prior to radical prostatectomy. The ethical committee waived the need for informed consent. Inclusion criteria were: a) MR imaging performed at 3 Tesla using an endorectal coil combined with pelvic phased array coils; b) MR imaging included three-directional T2-weighted (T2-w) imaging, diffusion weighted imaging (DWI) and dynamic contrast enhanced (DCE) MR imaging; c) prostatectomy performed in our institution and analyzed by one expert prostate pathologist. Exclusion criteria were: previous hormonal or radiotherapy or substantial imaging artifacts related to patient movement.

7.2.2 MR Imaging

MR imaging was performed using a 3T MR scanner (Siemens Trio Tim, Erlangen, Germany) with the use of an endorectal coil (Medrad, Pittsburgh) and pelvic phased array coil. The endorectal coil was filled
with a 40-mL perfluorocarbon preparation (Fomblin; Solvay-Solexis, Milan, Italy). Peristalsis was suppressed with an intramuscular administration of 20-mg Butylscopolaminebromide (Buscopan; Boehringer-Ingelheim, Ingelheim, Germany) and 1 mg of glucagon (Glucagen; Nordisk, Gentofte, Denmark). The imaging sequence parameters are shown in Table 7.1. Gadopentetate dimeglumine (Dotarem; Guerbet, Paris, France), of which 15 ml was administered with a power injector (Spectris; Medrad) at 2.5 mL/s and followed by a 30-mL saline flush, was used as contrast agent.

Table 7.1: MR imaging sequence parameters. T2-w: T2-weighted; T1-w: T1-weighted; DWI: Diffusion weighted imaging; DCE: Dynamic contrast enhanced imaging; TSE: Turbo Spin Echo; SE-EPI: Spin Echo - Echo Planar Imaging; TR: Repetition Time; TE: Echo Time; GRAPPA: Parallel imaging factor; GRE: Gradient echo imaging; FLASH: Fast low angle shot imaging.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence type</th>
<th>Slice thickness</th>
<th># Slices</th>
<th>In-plane resolution</th>
<th>TR</th>
<th>TE</th>
<th>Averages</th>
<th>GRAPPA</th>
<th>b-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2-w</td>
<td>TSE</td>
<td>3 mm</td>
<td>15-19</td>
<td>0.4 x 0.4 mm</td>
<td>4260 ms</td>
<td>99 ms</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Axial</td>
<td>TSE</td>
<td>3 mm</td>
<td>15-19</td>
<td>0.5 x 0.5 mm</td>
<td>3590 ms</td>
<td>98 ms</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2-w</td>
<td>TSE</td>
<td>3 mm</td>
<td>15-19</td>
<td>0.5 x 0.5 mm</td>
<td>4290 ms</td>
<td>105 ms</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coronal</td>
<td>TSE</td>
<td>3 mm</td>
<td>15-19</td>
<td>1.0 x 1.0 mm</td>
<td>2600 ms</td>
<td>81 ms</td>
<td>10</td>
<td>3</td>
<td>0, 50, 500, 800 mm²/s</td>
</tr>
<tr>
<td>DWI</td>
<td>SE-EPI</td>
<td>3 mm</td>
<td>15-19</td>
<td>1.0 x 1.0 mm</td>
<td>34 ms</td>
<td>1.4 ms</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>T1-w</td>
<td>GRE (FLASH 3D)</td>
<td>5 mm</td>
<td>16</td>
<td>1.0 x 1.0 mm</td>
<td>34 ms</td>
<td>1.4 ms</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

7.2.3 Pharmacokinetic MRI

Pharmacokinetic maps were generated off-line by an in-house developed system in three steps. Firstly, kinetic DCE imaging parameters were estimated by fitting each MR signal enhancement-time curve to a general exponential signal enhancement model as described previously [37]. Secondly, the signal enhancement-time curves were converted to tracer concentration [mmol/ml] time curves by applying the approach of Hittmair et al. [45]. Thirdly, inter-patient plasma profile variability was removed using the reference tissue method presented by Kovar et al. [93]. The reference tissue method assumes that a tissue area within the patient is available with a known tissue model based on literature values. For this experiment, the reference tissue was manually determined in consensus by two radiologists, by placing a region of interest (ROI) of 5x5 mm in the normal appearing peripheral zone which was visually characterized by a high-signal intensity on T2-w and ADC as well as homogeneous enhancement after contrast. Estimation of pharmacokinetic parameters was thereafter performed, conform to the theoretical derivations [95]. An extensive description of the method can be found in Vos et al. [96].

7.2.4 Histopathological evaluation of prostatectomies

Following radical prostatectomy, prostate specimens were uniformly processed and entirely submitted for histological investigation. Immediately after surgical resection, specimens were fixed in 10% neutral-buffered formalin, using fine needle formalin injections and fixation overnight. Subsequently, the entire surface was marked with ink using three different colours, after which the entire prostate specimen was cut into serial transverse 4 mm thick slices, perpendicular to the dorsal-rectal surface and all slices were macroscopically photographed with a CCD-camera. After histological staining all specimens were evaluated by one expert urological pathologist (C.H, 17 years experience). Tumors were outlined on the microscopic slides and subsequently mapped on the macroscopic photographs to allow reconstruction of tumor extent and multi-focality.
7.2.5 Standard of Reference

Histology tumor maps were used as ground truth. Annotations of MR images were performed in consensus by two radiologists. The morphology of the central gland, peripheral zone, cysts, calcifications, and urethra were used as landmarks to find the corresponding MRI slices. Translation techniques as described previously were furthermore used [47]. The anatomy of the prostate is best imaged on T2-w images which were therefore used for correlation to histopathological maps. First, based on histopathology, all tumor region were identified and a region of interest (ROI) placed in the peripheral and transition zone corresponding to tumor on histopathology. Only tumors > 0.2 cc in volume were annotated. Additional benign, but tumor suspicious regions, were annotated when: a) focal low-signal intensity on T2-w images and/or b) focal restriction on ADC maps and/or c) suspicious irregular enhancing areas were evident on images and the underlying histopathological analysis showed no evidence for tumor. Therefore all areas vaguely to strongly suspicious of malignancy based on current known features on multiparametric imaging were annotated. To allow exact spatial matching of the different imaging sets, a manual registration was applied for all patients, to correct for patient related movement.

7.2.6 CAD system

An in-house developed CAD system was used to assist the radiologist in the diagnosis of prostate lesions. An extensive description of the system can be found in previous publications [96, 132]. Briefly, the CAD system can visualize multi-parametric MRI and derived maps simultaneously in multiple linked views either as background or as transparent color coded overlays. Figure 7.1 demonstrates the CAD system with a dedicated prostate hanging protocol as it was used in the experiment. The CAD system characterizes a region of interest by extracting a relevant feature set from the available quantitative derived DCE and ADC maps. The extracted set of features is presented to a trained classifier which calculates a malignancy likelihood. Hereafter, the calculated likelihood is presented to the radiologist to assist in their diagnosis, as shown in Figure 7.2. For this observer study, two linear discriminant analysis classifiers were trained separately for the PZ and TZ. For the two classifiers, selection of features was carried out by Sequential Forward Floating Selection (SFFS) [153] to establish the most discriminant features. The SFFS procedure uses leave-one-patient-out training and testing with the area under the Receiver Operating Characteristic (ROC) curve as the criterion to be optimized. For the PZ, the 25th percentile of ADC values, 75th percentiles of $K_{trans}$ and Ve and 25th percentile of WashOut were selected. For the TZ, the 25th percentile of ADC values and WashOut were selected. The bootstrap resampling approach with 1000 iterations was used for estimating the bootstrap mean area-under-the (AUC) receiver operator characteristic curve as well as the 95% confidence intervals [86].

7.2.7 Observer Study

All anonimized patients and ROIs were shown in identical order to 9 observers. The observers varied in their level of experience: 6 less experience observers (LEO) in MR imaging of the prostate (< 50 prostate MRIs evaluated) and 4 prostate experienced observers (EO) (> 100 prostate MRIs evaluated). Observers were informed that all patients had biopsy proven prostate cancer followed by prostatectomy. The CAD system was designed with an experimental environment where predefined ROIs were automatically displayed for characterization by the observers. For each ROI to be evaluated, the axial, coronal and sagittal T2-w images, the ADC map, the pre-contrast T1-w images and as color coded transparent overlays, the DCE parameters: $K_{trans}$, WashOut and Ve were shown. An automated hanging protocol ensured that all images and maps of each patient were displayed at identical positions. A lookup table was displayed at each viewpoint to represent the window and level settings which were automatically set and fixed to a predefined intensity range, corresponding to absolute ADC values (0.5 - 1.5x10^-3 mm/s^2) and DCE enhancement values: $K_{trans}$ (1-3/s); Washout (-1 - -10/s) and Ve (20 - 70%). See Figure 7.1 for an example. For every ROI shown, the observer was expected to provide a malignancy likelihood on a scale of 0-100%. An interactive tool was displayed on top of the CAD system to guide the observer through the successive ROIs of each patient. The tool recorded a (pre-CAD) malignancy likelihood entered by the observer for a given ROI. Hereafter, features were extracted from multiple parametric images simultaneously within the ROI. A supervised classifier summarized the features into a CAD determined malignancy likelihood. The CAD
Figure 7.1: Example patient case of the observer study. The MRCAD observer hanging protocol shows on the top row from left to right a) T2-w axial b) T2-w coronal c) T2-w sagittal and d) ADC map. On the bottom row T1-w transversal images are displayed with the pharmacokinetic maps as transparent color overlays representing e) $K_{trans}$, f) $V_e$, g) WashOut and h) native T1-w image prior to contrast. The separate window shows the scoring widget were the observer is asked to enter a malignancy likelihood for the provided region of interest.

likelihood was displayed to the observer in combination with a distribution of the predicted likelihoods that was obtained during training of the classifier in relation to their reference standard, as demonstrated in Figure 7.2. Subsequently, the observer entered an additional (post-CAD) malignancy likelihood before the next ROI was shown. Observers were provided with the discriminating performances (as AUC) of the CAD system for the PZ and TZ respectively. Prior to the study, all observers were trained and familiarized with the CAD system, evaluating 4 cancer patients with a total of 25 different ROIs in the PZ and TZ with and without CAD.

7.2.8 Statistical Analysis

A multiple-reader multiple case (MRMC) ROC analysis (DBM MRMC 2.2, Kurt Rossmann Laboratories, Chicago, U.S.A) was performed. The average AUC for LEOs and EOs for the whole prostate as well as the PZ and TZ separately, were established before and after CAD. As MRMC analysis cannot provide statistical significance in repeated observational studies, additional linear mixed model analysis (using SPSS version 17) was performed to determine the significance. P values less than .05 were considered to indicate a significant difference.

7.3 Results

7.3.1 Sample Characteristics

A total of 34 patients were included with a mean age of 64 years (range 53-74); mean PSA of 7.5 (range 3.4 - 21.8) and median Gleason Score of 7 (range 5-9). In these patients, 120 benign and 86 malignant lesions
Figure 7.2: The scoring widget used for the observer study. The observer is asked to provide a malignancy likelihood for the provided region after which the observer presses the Score button. After the observer provides the pre-CAD malignancy likelihood, the CAD system calculates a malignancy score which is presented in a density plot. The green area in the density plot summarizes the smoothed distribution of all the calculated likelihoods for all benign regions from the database used for training the classifier. Likewise, the blue and red area corresponds with the normal and malignant regions, respectively. Hereafter, the observer can enter a post-CAD malignancy likelihood while taking the CAD prediction into account.

were annotated and evaluated. Of the 120 benign lesions, 64 were in the PZ and 56 in the TZ. Of the 86 malignant lesions, 67 were in the PZ and 19 in the TZ.

7.3.2 CAD Stand-Alone Performance
The overall CAD stand-alone AUC was 0.90 (CI 0.83-0.96) while for the PZ this was 0.92 (CI 0.88-0.96) and for the TZ, 0.87 (CI 0.78-0.96).

7.3.3 Observer Performance without CAD
LEOs had an overall pre-CAD performance of AUC=0.81 (CI 0.76-0.85); for the PZ this was AUC=0.86 (CI 0.83-0.88) and for the TZ, AUC=0.72 (CI 0.66-0.77). Experienced observers had an overall performance of AUC=0.88 (CI 0.85-0.93). A pre-CAD AUC=0.91 (CI 0.89-0.93) and AUC=0.81 (CI 0.69-0.94) was determined for the PZ and TZ respectively.

7.3.4 Observer Performance with CAD
When the observers were allowed to change their ratings depending on CAD predictions, the overall average AUC for less-experienced observers improved significantly to 0.91 (CI 0.90-0.93; p=0.001) and for experienced observers to 0.91 (CI 0.86-0.97; p=0.17). For less-experienced observers this was more evident for the PZ where the average AUC increased to 0.95 (CI 0.94-0.95; p<0.001) compared to the TZ, where the average AUC improved to 0.79 (CI 0.76-0.83; p=0.01). Experienced observers revealed no significant improvement in overall PZ (post-CAD AUC=0.93 [0.90-0.97]; p=0.13) or TZ lesion characterization (post-CAD AUC=0.82 [0.68-1.00]; p=0.42). A summary of the pre- and post-CAD performances are shown in Table 7.2 and Table 7.3 as well as Figure 7.3.
Table 7.2: Summary of mean pre- and post-CAD performances for readers grouped into less-experienced and experienced readers (* denotes a statistical significance).

<table>
<thead>
<tr>
<th></th>
<th>Pre-CAD performance (AUC)</th>
<th>Post-CAD performance (AUC)</th>
<th>Significance (p-values)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Less-experienced observers</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>0.81 (0.76-0.85)</td>
<td>0.91 (0.90-0.93)</td>
<td>0.001 *</td>
</tr>
<tr>
<td>PZ</td>
<td>0.86 (0.83-0.88)</td>
<td>0.95 (0.94-0.95)</td>
<td>&lt; 0.001 *</td>
</tr>
<tr>
<td>TZ</td>
<td>0.72 (0.66-0.77)</td>
<td>0.79 (0.76-0.83)</td>
<td>0.01 *</td>
</tr>
<tr>
<td><strong>Experienced observers</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>0.88 (0.85-0.93)</td>
<td>0.91 (0.86-0.97)</td>
<td>0.17</td>
</tr>
<tr>
<td>PZ</td>
<td>0.91 (0.89-0.93)</td>
<td>0.93 (0.90-0.97)</td>
<td>0.13</td>
</tr>
<tr>
<td>TZ</td>
<td>0.81 (0.69-0.94)</td>
<td>0.82 (0.68-1.00)</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Table 7.3: Summary of the individual performances pre- and post-CAD for all readers.

<table>
<thead>
<tr>
<th></th>
<th>Pre-CAD performance (AUC)</th>
<th>Post-CAD performance (AUC)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Less-experienced observers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PZ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reader 1</td>
<td>0.83</td>
<td>0.95</td>
</tr>
<tr>
<td>Reader 2</td>
<td>0.84</td>
<td>0.95</td>
</tr>
<tr>
<td>Reader 3</td>
<td>0.84</td>
<td>0.93</td>
</tr>
<tr>
<td>Reader 4</td>
<td>0.87</td>
<td>0.95</td>
</tr>
<tr>
<td>Reader 5</td>
<td>0.88</td>
<td>0.95</td>
</tr>
<tr>
<td>Reader 6</td>
<td>0.87</td>
<td>0.94</td>
</tr>
<tr>
<td>TZ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reader 1</td>
<td>0.73</td>
<td>0.81</td>
</tr>
<tr>
<td>Reader 2</td>
<td>0.72</td>
<td>0.77</td>
</tr>
<tr>
<td>Reader 3</td>
<td>0.65</td>
<td>0.80</td>
</tr>
<tr>
<td>Reader 4</td>
<td>0.65</td>
<td>0.77</td>
</tr>
<tr>
<td>Reader 5</td>
<td>0.80</td>
<td>0.85</td>
</tr>
<tr>
<td>Reader 6</td>
<td>0.74</td>
<td>0.76</td>
</tr>
<tr>
<td><strong>Experienced observers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PZ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reader 7</td>
<td>0.92</td>
<td>0.95</td>
</tr>
<tr>
<td>Reader 8</td>
<td>0.89</td>
<td>0.94</td>
</tr>
<tr>
<td>Reader 9</td>
<td>0.91</td>
<td>0.96</td>
</tr>
<tr>
<td>Reader 10</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>TZ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reader 7</td>
<td>0.89</td>
<td>0.92</td>
</tr>
<tr>
<td>Reader 8</td>
<td>0.87</td>
<td>0.88</td>
</tr>
<tr>
<td>Reader 9</td>
<td>0.80</td>
<td>0.88</td>
</tr>
<tr>
<td>Reader 10</td>
<td>0.71</td>
<td>0.69</td>
</tr>
</tbody>
</table>
7.4 Discussion and Conclusion

In this study we have demonstrated the clinically value of CAD to characterize prostate lesions by radiologists using information obtained from quantitative pharmacokinetic DCE parameters and ADC values on 3T MR imaging. The discriminating performance significantly improved for LEOs when they were assisted by CAD. For both the PZ and the TZ, this improvement was noted. Furthermore, when CAD was used, LEOs improved their discriminating ability to such an extent that similar performances were achieved as those of EOs. CAD however, did not significantly alter the performance of EOs. To our knowledge this study is the first to evaluate the effect of a CAD system for MR prostate evaluation by radiologists.

Multiparametric MR imaging (MP-MRI) is still the most accurate imaging technique available to detect, localize and stage prostate cancer. The additional value of information obtained from DWI is also gaining importance as a tumor aggressiveness biomarker, especially at 3T. T2-weighted imaging can still be regarded as the cornerstone of prostate evaluation as anatomy is exquisitely well depicted. Prostate cancer usually reveals a low-signal intensity region but this lacks specificity as numerous benign conditions, can have similar appearances. In addition, low-grade tumors might not be depicted and therefore detected as easily as high-grade tumors [147]. The overall accuracy of T2-weighted imaging has therefore been rather low with AUC ranging between 0.68 and 0.81 [162, 163]. Unfortunately, quantitative T2-w imaging is not easily performed and requires excessively long imaging times [52]. DCE-MR imaging suffers from a similar lack of specificity as prostatitis, high-grade PIN and normal BPH can also show increased vascularization and perfusion. For DWI, BPH and fibrosis also reveal increased proton movement restriction. Therefore, a multi-parametric approach combining all three imaging modalities has been shown to be most optimal [14, 164]. These authors showed that information obtained from the combination of the different imaging techniques provides a better discriminating performance then each technique individually. Yet, MP-MRI evaluation remains challenging, is largely dependent on experience and substantial inter-observer variability in interpretation exists.

As most MR imaging modalities lack specificity, the goal of our CAD approach was to improve the characterization accuracy of tumor suspicious lesions incorporating information from multiple features on DCE and DWI. This was done using linear discriminant analysis to determine the likelihood that a region represent malignancy or not. The current standard paradigm for the use of CAD systems is to use CAD as a second reader. After the radiologist has evaluated the multiple imaging sets, CAD offers the tumor likelihood for a given suspicious region, therefore aiding in differentiation. Our developed CAD system...
has a discriminating performance (AUC TZ: 0.87, PZ: 0.92) similar to that of an experienced radiologist (AUC TZ: 0.81, PZ: 0.91). In clinical practice, radiologists tend to evaluate MR images without quantitative analysis. For example, on DCE-MRI, enhancement patterns that indicate the presence of tumor are compared to the relative enhancement of the normal surround prostate tissue. In addition, ADC maps are only visually evaluated by looking at a focal area of relative restriction which may be indicative of tumor. Because of this relative assessment, no clear cut-off values for the presence of tumor are routinely used. As a result, the evaluation of multiparametric MRI requires a high level of experience and induces observer variability [16]. This was confirmed by our study which demonstrated that LEOs have substantially lower performances compared to EOs and that the inter-observer variability was large (CI 0.76-0.85). CAD not only improved the overall performance to the level of experienced observers, but also decreased the inter-observer variability (CI 0.90-0.93).

Transition zone tumors are known to have different genetic mutations, biological behavior and overall prognosis compared to PZ tumors [165, 166]. TZ tumors are furthermore often larger in volume and associated with higher PSA values. On the contrary they often are of lower grade and more likely to be confined to the prostate. Histologically the PZ consists of more glandular components compared to the TZ, which due to common BPH formation, has a larger stromal component including compact muscle fibers. On MR imaging this results in a lower T2-w signal and ADC values compared to the normal PZ where higher values are seen. In addition, the higher vascular features associated with BPH result in enhancement patterns on DCE similar to that of tumor [97]. The differences in MR appearances of TZ tumors compared to PZ tumors have been reported before [167, 168]. For this reason we have developed a CAD system which consists of two classifiers for characterization of the PZ and TZ lesions separately. The results of this study confirm that the TZ is indeed more challenging to evaluate than the PZ, as a lower stand-alone CAD performance (AUC 0.87 vs. 0.91 in the PZ) as well as lower overall observer performances (AUC 0.72-0.81) were achieved compared to the PZ (AUC 0.86-0.91).

Our study has a number of limitations. Firstly, we have used multiple observation per patient which may hamper a straightforward interpretation of the results. A linear mixed model analysis which incorporates findings from multiple observation in the same patient was performed to determine significance. Secondly, the number of TZ tumors was fairly low compared to PZ tumors. This is in accordance with the known overall lower prevalence (30%) of these tumors [169] identified on a normal basis. Therefore the overall performance of both the CAD system as well as the readers, may rather constitute a PZ dominated effect. Thirdly, as an integral part of DCE MRI quantification, the reference tissue calibration method requires an annotation of normal PZ. For this study this was done manually prior to the experiment. Ideally such annotation should be performed automatically without requiring any user interaction. Previous studies have shown that despite the fact that automatic calibration performs better compared to a general estimate, manual calibration is still superior at this stage [92]. Finally, all regions scored by the observer were annotated and predefined beforehand. This was done to prevent observer variability caused by annotating. In a clinically routine setup, either CAD should localize regions suspicious for cancer beforehand or provide a tumor likelihood based on radiologists interactive assessment, indicating regions suspicious of malignancy to be evaluated by CAD.

In conclusion, we have shown that the addition of CAD on multi-parametric 3T MP-MRI, significantly improves the discriminating performance for LEOs for both the PZ and TZ. Furthermore LEOs assisted by CAD, reached similar performance compared to EOs. Therefore CAD appears to be a promising method for implementation into routine clinical environment for MR imaging assessment of prostate cancer.
Chapter 8

Summary and General Discussion
Introducing Computer Aided Diagnosis (CAD) in the clinical workflow of a radiology department to assist in prostate cancer diagnosis is an enormous challenge. The lack of standardized MR sequences, consensus in interpretation of the acquired data and available objective quantitative features to diagnose prostate cancer obstructs the implementation of prostate CAD in medical centers. Furthermore, to become successful in a clinical environment, the CAD method should be fully automated, robust to the large population variation and fast enough for a typical screening production of say 30 to 40 cases a day. In this thesis, several methods are described that allow diagnosis of prostate cancer using objective quantitative features. Furthermore, a complete CAD scheme has been developed that assist radiologists in the diagnosis of prostate cancer and its additional value to the human observer was evaluated. In this final chapter, the results are summarized and the current status and future directions for prostate CAD methods in MRI are discussed.

8.1 Summary

Chapter 2 This chapter introduces a novel automated computerized scheme that was developed to determine a likelihood measure of malignancy for cancer suspicious regions in the prostate based on dynamic contrast-enhanced MRI (DCE-MRI) images. To evaluate the method, 34 consecutive patients with histologically proven adenocarcinoma in the peripheral zone of the prostate were collected. Both carcinoma and non-malignant tissue were annotated in consensus on MR images by a radiologist and a researcher using whole mount step-section histopathology as standard of reference. The annotations were used as regions of interest (ROI). A feature set comprising pharmacokinetic parameters and a T1 estimate was extracted from the ROIs to train a support vector machine as classifier. The output of the classifier was used as a measure of likelihood of malignancy. Diagnostic performance of the scheme was evaluated using the area under the ROC curve. The diagnostic accuracy obtained for differentiating prostate cancer from non-malignant disorders in the peripheral zone was 0.83 (0.75-0.92). This suggests that it is feasible to develop a CAD system capable of characterizing prostate cancer in the peripheral zone based on DCE-MRI.

Chapter 3 In this chapter, the CAD scheme is extended by extracting additional features from T2-weighted images in an attempt to improve the discriminating performance of the CAD method. Two issues arise when incorporating T2-w images in a CAD system: T2-w values are position as well as sequence dependent and images can be misaligned due to patient movement during the acquisition. A method was developed that computes T2 estimates from a T2-w and proton density value and a known sequence model. A mutual information registration strategy was implemented to correct for patient movement. Global motion is modeled by an affine transformation, while local motion is described by a volume preserving non-rigid deformation based on B-Splines. The additional value to the discriminating performance of a DCE T1-w based CAD system was evaluated using Bootstrapped ROC analysis. T2 values were successfully computed in 29 patients. T2 values were extracted and added to the CAD system from 39 malignant, 19 benign and 29 normal annotated regions. T2 values alone achieved a diagnostic accuracy of 0.85 (0.77-0.92) and showed a significantly improved discriminating performance of 0.89 (0.81-0.95), when combined with DCE T1-w features. In conclusion, the study demonstrated a simple T2 estimation method that has a diagnostic performance such that it complements a DCE T1-w based CAD system in discriminating malignant lesions from normal and benign regions. Additionally, the T2 estimate is beneficial to visual inspection due to the removed coil profile and fixed window and level settings.

Chapter 4 This chapter continues with the research on pharmacokinetic parameters derived from DCE-MRI. The feasibility of an automated calibration method for estimating the arterial input function when calculating pharmacokinetic parameters from DCE-MRI is shown. In chapter 2, it was demonstrated that the CAD system performs optimal when per patient calibration was used, but required manual annotation of reference tissue. In this study we propose a fully automated segmentation method that tackles this limitation and tested the method with our CAD system when discriminating prostate cancer from benign areas in the peripheral zone. A method was developed to automatically segment normal peripheral zone (PZ) tissue. Context based segmentation using the Otsu histogram
Summary and General Discussion

A feature set comprising pharmacokinetic parameters was computed for each ROI and used to train a support vector machine as classifier. In total 42 malignant, 29 benign and 37 normal regions were annotated. The diagnostic accuracy obtained for differentiating malignant from benign lesions using a conventional general patient plasma profile showed an accuracy of 0.65 (0.54-0.76). Using the automated segmentation per patient calibration method the diagnostic value improved to 0.80 (0.71-0.88), whereas the manual segmentation per patient calibration showed a diagnostic performance of 0.80 (0.70-0.90). These results show that an automated per-patient calibration is feasible, a significant better discriminating performance compared to the conventional fixed calibration was obtained and the diagnostic accuracy is similar to using manual per-patient calibration.

Chapter 5

Automatic prostate segmentation is a crucial step for CAD systems to reduce the number of false positive cancer candidates. Robust and fully automatic prostate segmentation in MR images is challenging, mainly due to the lack of well-defined edges, similar intensity profiles with surrounding organs or its relative small size in the male pelvis. In this chapter, a fully automatic parametric multivariate multi-object (PAMMO) segmentation method is presented that is generalizable to any modality and application including multiple MR modalities of any kind, i.e. combining ultrasound and MR images, as well as the freedom to define the amount of objects and their shapes and appearances. Segmentation is performed in two stages. The first stage is a novel approach where a PAMMO model is fitted to multiple MR images. Information from the multivariate MR images is simultaneously used during fitting of the model while its parameters are constrained within a population model leading to realistic solutions. In the second stage, the obtained knowledge about organ shape, position and appearance is used as prior knowledge in a Bayesian framework to classify the voxels of the multivariate MR images and to segment the prostate. The performance of the method is compared to approaches that (1) use information from a single MR image and (2) do not incorporate pelvic model fitting in their segmentation process. The results show that a multivariate segmentation including fitting significantly outperforms the other approaches with a dice coefficient index of 0.77±0.05.

Chapter 6

This chapter presents a fully automatic computer aided detection (CADe) method for the detection of prostate cancer. The CADe 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 Coefficients 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 is evaluated using methodology. The results show that the CADe method obtained sensitivities of 0.41, 0.65 and 0.74 at false positive levels of 1, 3 and 5 per patient, respectively. The study showed that it is feasible to automatically detect prostate cancer using information from multiple MR images simultaneously.

Chapter 7

In this chapter, the effect of computer-aided diagnosis (CAD) was determined on observer performance in differentiating benign and malignant prostate lesions on multiparametric 3T MRI. 34 Consecutive patients with biopsy proven prostate cancer who received a multiparametric 3T MRI incl. T2-weighted, diffusion weighted (DWI) and dynamic contrast enhanced (DCE) imaging prior to radical prostatectomy, were retrospectively included. Six less-experienced and four experienced prostate radiologists were asked to characterize different cancer suspicious regions on multiparametric MRI, first without and subsequently with the use of in-house developed CAD software. The effect of CAD was statistically analyzed using a multiple reader, multi-case, receiver operating characteristic analysis as well as linear mixed-model analysis. In 34 patients, a total of 206 pre-annotated regions which included 67 malignant and 64 benign regions in the peripheral zone (PZ) and 19 malignant and 56 benign regions in the transition zone (TZ) were evaluated. Stand alone CAD had an overall area-under the receiver operating characteristic curve (AUC) of 0.90, in discriminating PZ...
lesions an AUC=0.92 and for the TZ an AUC=0.87. Overall, less-experienced observers (LEO) had a pre-CAD AUC of 0.81 which significantly increased to 0.91 (p=0.001) after CAD while for experienced observers (EO) the pre-CAD AUC was 0.88 which increased to 0.91 (p=0.17) after CAD. In discriminating PZ lesions, LEOs increased their AUC from 0.86 to 0.95 (p<0.001) after CAD while EOs showed an increase from 0.91 to 0.93 (p=0.13). Similar, for the TZ lesions, LEOs significantly increased their performances from 0.72 to 0.79 (p=0.01) after CAD and EOs from 0.81 to 0.82 (p=0.42). The study concluded that the addition of CAD significantly improved the lesion discriminating performance for LEOs who reached similar performances as EOs. The stand alone performance of CAD is similar to experienced observers.

8.2 General discussion

The challenge to introduce CAD in the clinical workflow with a prostate cancer screening environment has been and still is enormous. The lack of standardized sequences, interpretation variability and objective quantitative features of PCa has been an important obstruction to develop adequate CAD methods for prostate MR diagnosis. This thesis presented a number of methods that allow diagnosis of prostate cancer using objective quantitative features obtained from multiparametric MRI. Furthermore, a complete prostate CAD system was developed and showed to be beneficial to the radiologist when diagnosing prostate cancer. In the next coming years, research towards automatic detection of prostate cancer will proceed further. Currently, multi-center studies are evaluating standardized MR imaging protocols for prostate cancer diagnosis and international prostate cancer societies emphasize on the importance of a scoring mechanism similar to the BIRADS score. The European Society of Urogenital Radiology is currently working on new guidelines that will have the purpose to standardize the current imaging techniques and diagnosis for prostate cancer. This is a convenient trend for the implementation of prostate CAD methods in the clinic, as they typically rely on a contained environment. Furthermore, standardization of prostate MRI and a structured way of reporting will benefit prostate cancer research enormously as they will produce large and well described datasets. In that way, prostate CAD systems can be trained and tested on larger datasets which is, of course, beneficial for the robustness of the CAD system.

The following key aspects can be identified from this thesis for future work:

**Additional features:** In this thesis, pixel intensity values were most often used for feature analysis. Especially in a multiparametric MR setup, histogram analysis on the pixel intensity values showed to be of benefit to discriminate tumor from benign conditions. In chapter 3 it was demonstrated that combining information from different MR images improves the discriminating performance, compared to using information from each image type separately. This concept is further demonstrated in chapter 6, where feature analysis was performed on three different MR images, i.e., T2, DCE, and ADC. The discriminating performance could potentially increase even more when information obtained from spectroscopy is included. It stands to reason that texture features that are similar to the criteria used in the clinic, will further improve the discriminating performance. For example, in T2-weighted images the criterion of smoothness in combination with a low T2 signal is used. Another example is the DCE symmetry criteria in the transition zone. Asymmetrical enhancement in the transition zone is a marker for prostate cancer whereas symmetrical enhancement is typical for BPH [97]. Additionally, pattern recognition techniques can be applied for a final boost of the discriminating performance of the CAD method. Examples that have been published in literature are Markov random fields and fractals [50, 51, 52, 55]. Anatomical features can be added which are extracted from e.g. T2-weighted MR images. In clinical practice they still form the basis of prostate cancer diagnosis, where the radiologist searches for shape features such as bulging or an irregular shaped prostate capsule. Another way to improve the CAD system would be to employ contextual information, such as patient age, PSA level and history.

**Prostate cancer aggressiveness:** This thesis primarily focussed on the capability of detecting clinically significant prostate cancer and the ability to differentiate prostate cancer from benign abnormalities. In a clinical perspective however it is becoming more important to only detect the more aggressive tumors or the ability to differentiate high grade from low grade tumors. Therefore, several authors
have attempted to establish a correlation between the Gleason score and image modality [47, 170, 157, 171]. However, a substantial overlap between the different Gleason scores and the information obtained from the image modality was present. In chapter 6, primarily results were presented of the developed CAD system with the ability to detect tumor with different grades. As the results are promising, future work could focus on the evaluation of the correlation between Gleason score and information obtained from multiparametric MRI.

**MR CAD image guidance:** The precise localization and grading of prostate cancer is becoming a very important aspect when diagnosing prostate cancer. For example in case of focal therapy, the exact location and most aggressive part of the tumor is of paramount importance. Another example is biopsy, where it is essential to target the most aggressive and thus representative part of the tumor. Future research would be whether MR CAD image guidance can assist in localizing the dominant tumor and objectively guide, e.g. biopsy, towards the most aggressive part of the tumor. Chapter 6 presented a method that detects prostate cancer by localizing the most dark region of the tumor in the ADC map. This is similar performed at an MR guided biopsy, where the operator determines the most aggressive part of the tumor by visual inspections of the ADC map. The developed CAD system is able to provide exact coordinates of the most aggressive part of the tumor, which can be transferred towards the biopsy targeting system. Another application would be to incorporate the prostate CAD results into in TRUS guided biopsy system by means of e.g. image registration.
Bibliography


96

BIBLIOGRAPHY


100

BIBLIOGRAPHY


Appendix A

MRCAD: a In-house Developed Toolkit for Prostate MR Computer Aided Diagnosis

This chapter is based on the paper “MRCAD for daily clinical analysis of prostate MR” by Henkjan Huisman, Pieter C. Vos, Published at The Kitware Source, April 2010.
A.1 Introduction

The Diagnostic Image Analysis Group (DIAG) from the Department of Radiology at Radboud University Nijmegen Medical Centre in The Netherlands researches computer aided diagnosis of breast, lung and prostate cancer. The research in prostate cancer is a collaborative effort with the clinical prostate group in the department and is focused on high resolution anatomical, functional (diffusion and dynamic contrast enhanced) and metabolic (spectroscopy) MR. During various prostate and breast MR projects an image analysis workstation, MRCAD, has been gradually developed based on VTK, ITK and IncrTCL. This workstation provides dedicated image analysis tools and display facilities for various state-of-the-art MR sequences and hardware as well as a platform for testing new computerized image analysis techniques. Currently, MRCAD is in daily clinical use at our clinic and at a few other sites. MRCAD is being used to reporting prostate MR and in various research projects and publications. This article will address some of the design features and explain how VTK, ITK and IncrTCL have been used to achieve a robust, versatile, clinical research platform.

A.2 Screenshots

To get an impression of MRCAD, first two example patient cases. Fig A.1 shows an MR detection study of a patient with elevated PSA, where prior ultrasound biopsy could not confirm the presence of cancer. The detection protocol includes: t2-w, ADC, and DCE-MRI sequences. The radiologist marks a finding, which is suspicious for cancer, with a transparent red sphere (vtkUnstructuredGrid) and is presented with CADx output to assist in differentiating between benign and malignant. A malignancy score of .33 (green and red are scores of benign and malignant training findings) indicates that the region is unlikely to contain cancer. This patient is currently in an active surveillance program.

Fig. A.2 shows a 59-year-old patient with a PSA of 12 ng/ml and a confirmed biopsy Gleason 8 tumor in the right peripheral zone. The MRCAD hanging protocol shows T2-w transversal (right view) and in color overlays: Pharmacokinetic DCE-MRI; and ADC-map (left upper row); and choline metabolite concentration; coronal T2-w image (left bottom row); The separate window shows a time-concentration curve and an MR Spectroscopy spectrum at the cursor location. The reporting radiologist rated all 4 modalities 5 on a 5 point scale: where 1 is no tumor, and 5 is definitely tumor. These MR findings were confirmed retrospectively with a stage T3a determined after prostatectomy.
A.3 Overview of the software

The workstation is running an OpenSuse Linux OS. A DCMTK server daemon is receiving DICOM imaging that are automatically forwarded by the PACS system. A polling mechanism triggers automatic pre-processing of MR studies (e.g. fitting DCE-MRI curves). The studies are then ready for viewing and further analysis. MRCAD features are: compressed, multiframe DICOM Reading/Writing; multi-modal color overlay; DCEMR pharmacokinetic modeling and registration; manual and semi-automatic segmentation; region of interest statistics; CADx, CADe; MRS (using LCModel), and Structured Reporting.

The MRCAD software has a layered structure as shown in Fig. A.3. The main programming language is Tcl/Tk with the object oriented extension IncrTcl. Processing and display are performed in VTK pipelines that are setup using the Tcl wrapped interface. Local VTK classes provide additional high computational functionality at the C++ level or as means to include wrappable ITK functionality.

A.4 Annotation and Structured Reporting

Clinical findings and ground truth for CAD are annotated and stored in XML. Example tags are: tumor location, tumor grade, scorings, access time, etc. For structured reporting findings can be collected and
Chapter A

added to the XML database. A pdf file can be generated from the XML file (using LaTeX) and, after inspection, send to an Electronic Patient Database using SOAP requests. The XML data is additionally used for training CAD systems, and performing clinical observer studies. An XML example of an observer scoring is:

```xml
<markdatasetlists
PatientsName="Anonymous" PatientID="12345"
StudyDate="20100217" InstitutionName="UMC_St_Radboud"
StudyDescription="ABDOMEN-ONDERBUIK">
  <marklist username="pieter" mode="detect"
repeat="1" studysetname="Default">
    <mark id="mark1">
      <Time>2010.02.24 - 11:51:05</Time>
      <Position>-25.7904 9.30513 -53.7623</Position>
      <AlignedModalityBG>T2T1r1</AlignedModalityBG>
      <Type>Tumor</Type>
      <T2>5</T2><DCE>4</DCE><DIFF>4</DIFF
      <Link id="Link1">
        <StudyDate>20091220</StudyDate>
      </Link>
    </mark>
  </marklist>
</markdatasetlists/>
```

The node `markdatasetlists` is the root node for all annotations of the MR study. In this example (user 'pieter' in the `marklist` node) a tumor is detected and is rated 5-4-4 on 3 MR modalities. Furthermore, the finding is linked to another finding in a prior MR study (Link1). The patient diagnosis is in the `markpatient` node: a T2c tumor. After reporting the patient case, MRCAD can use the findings stored in the XML database to collect screenshots and format the information into a PDF file using dedicated stylesheets. This PDF structured report is then send to the Electronic Patient Database using SOAP: set token

```bash
[:http::geturl $posturl -type "text/xml; boundary='"" -headers [list Cookie $sessionid] -query [Soaprequest asXML]]
set body [http::data $token]
http::cleanup $token
```

A.5 Coordinate system

MRCAD uses three coordinate systems to handle the variety of image volume and voxel dimensions and orientations in an MR scan: ijk voxel location in an MR series, xyz patient position, uvw viewport coordinate. A vtkDCMTKImageReader directly reads DICOM images in single slice, multiframe, and/or compressed format and produces a vtkImageData and a vtkTransform (ijk2xyz). The latter transform maps voxel location ijk to an xyz patient position. The viewer uses an additional uvw2xyz transform to handle: zooming, panning and slice selection. This allows for accurate (sub-voxel) overlay of any image (e.g transforms 0.5x0.5x3mm) onto any other orientated image (oblique (15deg) sagittal (1x1x4mm)). Arbitrarily shaped 3D volume annotations are created and stored in xyz coordinates using the vtkUnstructuredGrid format. These annotations can be added and overlayed to any view and are also used to compute region of interest statistics in the various multi-model images.

A.6 Data and widget factory

The core of the MRCAD application is designed as a factory method pattern. The approach ensures fast initializing of the application, by constructing data objects on demand. As a result, VTK pipelines are only created when needed. An IncrTcl coded part of the factory:
class MRCADDATA FACTORY()

public {
    # static method that returns the singleton
    proc Instance {}{
    # check whether the data object 'mapname' is registered
    method IsAvailable {mapname}
    # Register the data object named 'mapname' to the factory
    method AddData {mapname data}
    # Returns the data object from the factory
    method GetData {mapname}
}

    # Implementation
    body MRCADDATAFACTORY::GetData {mapname} {
        set data $.rgContainer($mapname)
        set object [find object *$data]
        if { $object == ""} {
            # create the object only once
            set object [lindex $data 0]  
            set $.rgContainer($mapname) $object
        }
        return [namespace which $object]
    }

    The factory is implemented as a singleton and can be queried using the static member function Instance.
    The factory is filled with possible data during MRCAD initialization, for example a diffusion-w series might be present, then:

    # Register a DiskImageData object
    [MRCADDATAFACTORY::Instance] AddData "Diff1" [list DiskImageData $files]

    adds an imagedata object with mapname "Diff1", which is constructed by a specialized class DiskImage-Data (an IncrTcl wrapper of the vtkDCMTKImageReader).
    To view the diffusion-w image in the application:

    if {[MRCADDATAFACTORY::Instance] IsAvailable "Diff1"} {
        # Get the data object and
        # create the VTK pipeline necessary to produce the data
        set obj [MRCADDATAFACTORY::Instance] GetData "Diff1"
        # Display the image in the viewer
        v1SetImage BG [Obj GetVolume]
        v1SetTransform BG [Obj GetTransform]
    }

    Data objects can be accompanied by data widgets to allow user interaction with the data object. In the image shown below the spectroscopic metabolite image data object output is displayed as a transparent color-coded overlay on top of a T2-w image. Multiple metabolites map are available and the accompanying data widget (shown to the right) allows selection of the metabolite and provides other interactions with the data object. In Fig. A.4 is an example where the 3th available metabolite, choline (Cho), is selected:

A.7 MRCAD used for research

MRCAD is being used in a number of research projects. MRCAD was used to detect and annotate the Dominant Intraprostatic Lesion for radiotherapy IMRT planning [172]. A successful Computer Aided Diagnosis system has been researched and implemented to discriminate benign from malignant suspicious regions using region of interest statistical features derived from DICOM images directly or processed parametric maps [96]. Pattern recognition as well as ROC analysis was performed using the statistical package R.
Pharmacokinetic features computed from Dynamic Contrast Enhanced MRI provided the highest diagnostic accuracy. Subsequently in [132], more modalities/features were added to improve the discriminating performance. An ITK based registration method was extended with a local incompressibility term to add MR series features that were misaligned by possible patient movement during acquisition. Currently, an initial system for fully automatic detection of prostate cancer is being developed using a locally developed vtkClass based on the itkHessianRecursiveGaussianImageFilter to detect lesions [133, 173]. MRCAD is also used in several observer studies that are performed by radiologists to determine the diagnostic value of a certain MR modality or protocol. In [11] MRCAD was used for scoring DCE-MRI derived parameters and establish the value for localizing prostate cancer. Recent use includes detecting MR suspicious lesions for MR or ultrasound guided biopsy and detecting recurrence after radiotherapy [143, 174].
Publications

Papers in international journals


- Thiele Kobus, **Pieter C. Vos, Thomas Hambrock, Christina Hulsbergen-Van de Kaa, Jelle Barentsz, Arend Heerschap, Tom Scheenen** Combined use of MR Spectroscopy and Diffusion Weighed Imaging for in-vivo assessment of prostate cancer aggressiveness Submitted to Radiology.

- **Thomas Hambrock and Pieter C. Vos, Jelle O. Barenstz, Henkjan J. Huisman** Computer Aided Diagnosis in Localizing Prostate Cancer: an observer Study Submitted to Radiology.


- **Pieter C. Vos, Thomas Hambrock, Jelle O. Barenstz, Henkjan J. Huisman** Computer assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI Phys Med Biol. 2010 Mar 21;55(6):1719-34. Epub 2010 Mar 2. PMID: 20197602


Papers in conference proceedings

- Henkjan J. Huisman, Pieter C. Vos, Geert Litjens, Thomas Hambrock, Jelle Barentsz. *Computer aided detection of prostate cancer using T2w, DWI and DCE-MRI: methods and clinical applications.* MICCAI 2010 Prostate Cancer Imaging Workshop, Beijing


Abstracts in conference proceedings

- Robust pharmacokinetic prostate MRI using an automated per-patient reference tissue artery input function estimator. ECR March 2010 Viena

- Computer-aided Detection of Prostate Lesions at Diffusion-weighted MR Using a Dedicated Hessian Matrix-based Detection Scheme. RSNA, November 2009 Chicago

- Computer assisted analysis of peripheral zone prostate with T2-weighted and dynamic contrast enhanced T1-weighted MRI. NVPHBV November 2008

- Computer assisted analysis of peripheral zone prostate lesions with T2-weighted and dynamic contrast enhanced T1-weighted MRI. ISMRM Benelux, December 2008, Antwerpen
Appendix B

Samenvatting
Het introduceren van een computer werkstation dat ondersteuning geeft bij de diagnose van prostaatkanker in MR beeldvorming blijkt, in de klinische praktijk van een radiologieafdeling, een enorme uitdaging. Problemen als het ontbreken van gestandaardiseerde MR sequenties, geen consensus tussen radiologen bij de interpretatie van de verkregen beeld gegevens en het missen van objectieve kwantitatieve kenmerken voor de diagnose van prostaatkanker in de beelden belemmert de uitvoering van computer ondersteunende diagnose (CAD) van prostaatkanker in medische centra. Om een prostaat CAD workstation verder succesvol te laten opereren in een klinische omgeving, dient het volledig geautomatiseerd te zijn, robuust berekeningen te kunnen uitvoeren om de grote populatie variatie te kunnen ondervangen en snel genoeg te zijn om een grote productie van zeg 30 tot 40 prostaatkanker patiënten per dag te kunnen diagnosticeren.

Dit proefschrift beschrijft verschillende methoden die diagnose van prostaatkanker in MR beelden mogelijk maakt voor radiologen met behulp van objectieve kwantitatieve computer analyses. Een volledig prostaat CAD systeem is beschreven in deze proefschrift en de aanvullende waarde voor de radioloog is gevalueerd.

Dit laatste hoofdstuk presenteert een samenvatting van de resultaten.

B.1 Samenvatting

Hoofdstuk 2

Dit eerste hoofdstuk presenteert een nieuw computer systeem dat werd ontwikkeld om een kans op maligniteit automatisch vast te stellen voor regio’s in de prostaat die kanker verdacht zijn. Het systeem werkt op basis van automatische interpretatie van dynamisch contrast-versterkende MRI (DCE-MRI) beelden. Het evalueren van het systeem werd gedaan op een dataset bestaande uit een cohort van 34 patiënten met histologisch bewezen adenocarcinoom in de perifere zone van de prostaat. Zowel carcinoom als niet-kwaadaardig weefsel werd gemarkeerd op MR beelden door een radioloog en een onderzoeker in consensus met histologisch geheel ingebed prostaat weefsel als gouden standaard. De markeringen werden gebruikt ter referentie (ROI) voor verdere evaluatie. Een feature set, bestaande uit perfusie en diffusie parameters en een T1 schatting, werd geextract uit iedere ROI. De feature set werd gebruikt om een support vector machine te trainen als classifier. De output van de classifier werd gebruikt als een maat voor kans op maligniteit. De diagnostische waarde van het systeem werd gevalueerd met behulp van de opvlakte onder de statistische receiver operating characteristic (ROC) curve. De diagnostische nauwkeurigheid voor de differentiatie van prostaatkanker en niet-kwaadaardige aandoeningen in de perifere zone resulteerde in 0,83 (0,75 - 0,92). Hieruit kan worden geconcludeerd dat het CAD systeem in staat was prostaatkanker van niet-kwaadaardig weefsel in de perifere zone te onderscheiden gebruikt makend van DCE-MRI beelden.

Hoofdstuk 3

In dit hoofdstuk wordt uitgelegd hoe het CAD systeem werd uitgebreid met de analyse van T2-gewogen (T2-w) beelden in een poging om het discriminerend vermogen te verbeteren. De integratie van mbox T2-w beelden in het CAD systeem heeft twee kwesties die opgelost dienden te worden: 1) de T2-w waarden worden beïnvloed door de keuze van MR sequentie als mede het spoel profiel; 2) doordat beweging van de patiënt tijdens de MR beeldvorming lastig voorkomen kan worden, zijn er positionele verschillen aanwezig tussen de MR beelden. In het hoofdstuk wordt een methode beschreven dat kwantitatieve T2 berekend uit een mbox T2-w en protonendichtheid MR beeld tezamen met een sequentie model. Een Mutual Information registratie strategie was geïmplementeerd om bewegingen van de patiënt te corrigeren. In deze strategie werden globale bewegingen gemodelleerd door middel van een affine transformatie algoritme en lokale bewegingen werden gomodelleerd door van een niet-rigide transformatie algoritme met behoud van volume. De additionele waarde van de kwantitatieve T2 informatie op het CAD systeem werd gevalueerd met behulp van een Bootstrapped ROC analyse. T2 schattingen werden succesvol berekend voor alle 29 patiënten. 39 kwaadaardige, 19 goedkwaadaardige en 29 normale gebieden werden gomarkeerd door een radioloog waaruit T2 schattingen en DCE parameters werden gecomputeerd. De geëxtraheerde informatie werd gebruikt om het CAD systeem voor zijn discriminerende vermogen te evalueren. Uit de experimenten bleek dat met T2 schattingen afzonderlijk, het CAD systeem een diagnostische nauwkeurigheid behaald van 0,85 (0,77-0,92). Een sterke verbetering was zichtbaar wanneer DCE parameters werden toegevoegd: 0,89 (0,81-0,95). Het hoofdstuk concludeert dat de informatie van de T2 schattingen een aanvulling is voor het CAD systeem omdat het diagnostisch vermogen drastisch
Verbeterd. Daarnaast zijn de kwantitatieve T2 schattingen gunstig voor visuele inspectie aangezien het spoel profiel wordt verwijderd en een objectief vaste contrast instelling gebruikt kan worden.

**Hoofdstuk 4** In hoofdstuk 2, werd aangetoond dat het CAD systeem optimaal presteert wanneer een per-patiënt kalibratie werd gebruikt. Echter, de methode vereiste een handmatige markering van referentie weefsel. In dit hoofdstuk wordt de haalbaarheid van een automatische kalibratie methode onderzocht en getest met behulp van CAD systeem. De methode betreft een volledig geautomatiseerde segmentatie van gezonde perifere zone weefsel dat gebruikt wordt ter referentie. De methode omvat een context gebaseerde segmentatie bestaande uit een Otsu histogram drempel selectiemethode en een Hessian gebaseerde bol detectie. De database voor de evaluatie bestond uit een cohort van 38 patiënten met carcinoom. Een radioloog en onderzoeker markeerde in consensus kwaadaardig, goedaardig en normaal weefsel op MR beelden. Hierbij werd histologisch geheel ingebonden deze als gouden standaard. Voor iedere ROI werd een feature set berekend bestaande uit PK parameters. De feature set werd gebruikt om een support vector machine te trainen voor de classificatie van de verschillende gebieden. In totaal, 42 kwaadaardige, 29 goedaardige en 37 normale regionen werden gemarkeerd. De diagnostische nauwkeurigheid voor de differentiatie van kwaadaardige en goed aardige laesies werd bepaald voor drie verschillende kalibratie technieken: 1) een conventioneel algemene patiënt plasma profiel; 2) handmatige segmentatie per patiënt kalibratie; en 3) automatische segmentatie per patiënt kalibratie. De experimenten lieten zien dat het behoud van een conventioneel algemene patiënt plasma profiel een nauwkeurigheid van 0.65 (0.54-0.76) kan worden gehaald. De geautomatiseerde segmentatie per patiënt kalibratie methode liet een duidelijke verbetering zien van het discriminierend vermogen: 0.80 (0.71-0.88). De handmatige segmentatie per patiënt kalibratie methode toonde een gelijkvaardig diagnoseprestatie van 0.80 (0.70-0.90). Deze resultaten tonen aan dat een volledig geautomatiseerde methode voor een automatisch per patiënt kalibratie haalbaar is, de automatische methode een significant betere discriminerende prestatie laat zien ten opzichte van de conventionele algemene patiënt plasma profiel en vergelijkbaar presteert met een handmatig per patiënt kalibratie methode.

**Hoofdstuk 5** Automatische segmentatie van de prostaat is een cruciale stap voor het CAD systeem om de het aantal detecteerde false-positive kandidaten te beperken. Een volledig geautomatiseerde prostaat segmentatie methode in MR beelden is een enorme uitdaging, veroorzaakt door het gebrek aan goed gedefinieerde randen, overlappende intensiteit waarden met omliggende organen en de relatief kleine omvang in het mannelijke bekken. Het hoofdstuk beschrijft een volledig geautomatiseerde parametrische multivariate multi-object (PAMMO) segmentatie methode dat generaliseerbaar is naar elke modaliteit en toepassing, bijvoorbeeld het combineren van echografie en MR beelden. De methode is toepasbaar op een willekeurig aantal objecten en laat de definitie van hun vormen en opmaak volledig vrij. De segmentatie wordt in twee fasen uitgevoerd. De eerste fase betreft een nieuwe benadering, waarbij een PAMMO model wordt geïntegreerd op meerdere MR beelden tegelijk. Informatie van de MR beelden wordt door het PAMMO model gebruikt bij het genereren van vergelijkbare synthetische MR beelden. Alleen realistische oplossingen worden gegenereerd doordat de PAMMO parameters binnen de grenzen van een populatie model worden gehouden. In de tweede fase wordt de verkregen kennis over orgaan vorm, positie en opmaak gebruikt als voorkennis in een Bayesians strategie om de voxels van de MR beelden classificeren en de prostaat te segmenteren. In het hoofdstuk worden experimenten besproken waarbij de prestatie van de PAMMO methode wordt vergeleken met (1) de prestatie van een segmentatie methode dat alleen informatie uit een enkele MR beeld haalt en (2) een segmentatie methode dat het fitten achtwege laat. De behaalde resultaten toonen aan dat het PAMMO segmentatie methode inclusief fitting aanzienlijk beter presteert dan de andere benaderingen met een Dice coëfficiënt index van 0.77 ± 0.05.

**Hoofdstuk 6** Dit hoofdstuk presenteert een volautomatische computergestuurde detectie (CADE) algoritme voor de detectie van prostaatkanker. De procedure om locaties op te sporen die verdacht zijn voor prostaatkanker bestaat uit meerdere opeenvolgende fases. In de eerste fase wordt een voxel classificatie uitgevoerd met behulp van een Hessians gebaseerde bol detectie algoritme, welke op meerdere schalen wordt uitgevoerd op een Apparent Diffusion Coefficients map. Vervolgens wordt het PAMMO segmentatie (zie hoofdstuk ?? methode toegepast. De daaruit resulterende segmentatie
wordt gebruikt om de kandidaat-detectie tot de prostaat te beperken. De overige kandidaten worden geclasseerd door middel van histogram analyse op de multiparametrische MR beelden. De geëxtraheerde kenmerken worden vervolgens samengevat tot een kans op kwaadaardigheid gebruik makend van een supervised classifier in een two-stage classificatie methode. De prestaties voor het detecteren van prostaatkanker werden geëvalueerd met behulp van free response operating characteristic (FROC) methodologie. De resultaten tonen aan dat het CAD algoritme gevoeligheden van 0,41, 0,65 en 0,74 haalt bij vals positieve ratios van 1, 3 en 5 per patiëntrespectievelijk. De studie toonde aan dat het mogelijk is om volledig automatisch prostaatkanker te detecteren met behulp van informatie uit verschillende typen MR beelden.

Hoofdstuk 7 In dit hoofdstuk werd het effect van computer-aided diagnose (CAD) bepaald op radiologen bij het differentiëren van goedaardige en kwaadaardige prostaat afwijkingen op multiparametrische 3T MRI. Het hoofdstuk beschrijft een experiment met 34 opeenvolgende patiënten die door biopsie bewezen prostaatkanker hebben. Bij deze patiënt was voorafgaand aan radicale prostatectomie, een multiparametrische 3T MRI inclusief T2-gewogen, diffusie gewogen (DWI) en DCE beeldvorming uitgevoerd. Zes minder ervaren en vier ervaren prostaat radiologen werden gevraagd om met kanker verdachte regio’s te karakteriseren met behulp van de multiparametrische MRI, eerst zonder en vervolgens met het gebruik van het ontwikkelde CAD systeem. Het effect van CAD was statistisch geanalyseerd met behulp van multiple reader, multi-case ROC analyse en een lineaire mixed-model analyse. Bij de 34 patiënten werden in totaal 206 regio’s geraadpleegd bestaande uit 67 kwaadaardige en 64 goedaardige regio’s in de perifere zone (PZ) en 19 kwaadaardige en 56 goedaardige regio’s in de transitie zone (TZ). Het discriminerend vermogen in van het CAD systeem op deze dataset had een totale oppervlakte onder ROC curve (AUC) van 0,90, bij het onderscheiden van PZ laesies een AUC van 0,92 en voor de TZ een AUC van 0,87. Over het geheel genomen hadden minder ervaren radiologen (LEO) een pre-CAD AUC van 0,81 die aanzienlijk werd verbeterd tot 0,91 (p = 0,001) na het gebruik van CAD, terwijl voor ervaren radiologen (EO) de pre-CAD AUC 0,88 was, die verhoogd naar 0,91 (p = 0,17) na het gebruik van CAD. In de PZ, LEO’s stegen in hun AUC van 0,86 naar 0,95 (< 0,001) na het gebruik van CAD, terwijl EO een stijging lieten zien van 0,91 naar 0,99 (p = 0,13). In de TZ, LEO’s stegen aanzienlijk in hun prestaties van 0,72 naar 0,79 (p = 0,01) na CAD en EO een stijging lieten zien van 0,81 naar 0,82 (p = 0,42). Het hoofdstuk concludeert dat de toevoeging van CAD significant het discriminerend vermogen van minder ervaren radiologen verhoogd naar een vergelijkbare prestatie van ervaren radiologen. Bovendien presteert het CAD systeem zelfstandig op gelijke voet met ervaren radiologen.
Appendix C

Acknowledgements
It was a beautiful Tuesday in summer 2005 as I rushed towards the climbing hall on my bike. My mind was wondered off not realizing someone was about to bump into me. I was working at a software house where, after two years of employment, I realized that I missed a certain challenge I could not describe. Henkjan Huisman completely surprised me by stopping in front of my bike, not letting me pass further. This was the turning point of my life. I had never considered doing a PhD career but Henkjan offered me the possibility of doing so. I had no idea what so ever what a PhD career looked like, but the enthusiasm of Henkjan on the scientific challenge completely realized me what I was missing as a software engineer: the hunger for knowledge. From here on, I followed and trusted Henkjan fully into an academic career that I now so proudly continue.

Dr. ir. H.J. Huisman, dear Henkjan and good friend, I hope you allow me to consider you as the best supervisor I could ever wish for. Your patience, expert knowledge and perfectionism has helped me towards the academic level I now have. That, what seemed impossible to some, has been become successful because of your persistent trust in my capabilities.

Prof. dr. ir. N. Karssemeijer, dear Nico, I am honoured to be the first candidate that will obtain his PhD during your academic career as a Professor. I loved to listen to your intriguing anecdotes and stories of your vivid live.

Prof. dr. J.O. Barentsz, dear Jelle, I think one cannot learn more about prostate cancer diagnosis from such an experienced radiologist such as you. The software tool MRCAD has brought us to many discussion which lead to a state of the art development of the MRCAD software. Thank you for your clinical insights that lead to the many publications and this final thesis.

Dr. T. Hambrock, dear Thomas and good friend, thank you very being my partner in this thesis. I believe the strategy of a medical and computer scientist team has worked out quite well. The clinical expertise you obtained together with my computer skills has lead to a very successful and long list of publications. I hope we will continue to work together in the future.

Dear colleagues at the radiology department and not to forget the secretary, thanks for the wonderful time together as a colleague.

Drs. M. Samulski, dear Maurice and neighbour colleague. Most of my PhD time I spent next to you. I think we have survived the cold winters and hot summers quite well in the ‘barak’. Thanks for listening to the many compiling frustrations I had during development. And of course the nice time we spent at conferences and free time around it.

Dear colleagues at the DIAG group, as a group we stand strong in an academic world full of competition. Thank you all for the good help and positive input.

Dear friends with whom I climb every Tuesday and Thursday, Desiree, Leonard and Brigiet. Although you do not seem to understand why I continue to suffer as a researcher, I thank you a million for listening to all more issues and troubles during the finalization of my thesis. Thank you for the many escapes.

Dear friends with whom I enjoy live, thank you for all the support that I received.

Finally I would like to thank my family, for their everlasting trust, support and sweetness towards me. Thank you for making this happening.

But above all stands my sweet Clarisa, without you I would never have finished my thesis. No day passes without a thought on you. Thanks for your patience and believe such that I was able to go through one of my most difficult achievements so far: finishing this thesis.