Computer-Assisted Characterization of Prostate Cancer on Magnetic Resonance Imaging

Derek J. Soetemans
The University of Western Ontario

Supervisor
Dr. Aaron Ward
The University of Western Ontario

Graduate Program in Biomedical Engineering
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Engineering Science
© Derek J. Soetemans 2017

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Bioimaging and Biomedical Optics Commons

Recommended Citation
https://ir.lib.uwo.ca/etd/4504

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.
Abstract

Prostate cancer (PCa) is one of the most prevalent cancers among men. Early diagnosis can improve survival and reduce treatment costs. Current inter-radiologist variability for detection of PCa is high. The use of multi-parametric magnetic resonance imaging (mpMRI) with machine learning algorithms has been investigated both for improving PCa detection and for PCa diagnosis. Widespread clinical implementation of computer-assisted PCa lesion characterization remains elusive; critically needed is a model that is validated against a histologic reference standard that is densely sampled in an unbiased fashion. We address this using our technique for highly accurate fusion of mpMRI with whole-mount digitized histology of the surgical specimen. In this thesis, we present models for classification of malignant, benign and confounding tissue and aggressiveness of PCa. Further validation on a larger dataset could enable improved classification performance, improving survival rates and enabling a more personalized treatment plan.

Keywords

Magnetic resonance imaging, prostate cancer, machine learning, computer-assisted, Gleason grade, focal therapy, radiology, radiation oncology, pathology, histology
Co-Authorship Statement

This thesis is presented in an integrated article format, the second chapter of which is based on the following publication that is in preparation for submission:


My contributions to this work included defining the research questions, writing the software, formulating the experimental design, implementing and analyzing machine learning algorithms, and drafting the manuscript. A. D. Ward motivated the initial research direction and assisted with defining the research questions, formulating the case study analysis, interpreting the results and drafting the manuscript. All authors helped in reviewing and editing the manuscript. The work was performed under the supervision of A. D. Ward.
| Acknowledgments |

This thesis was completed thanks to the continued support of many people over the past few years. First I would like to thank my supervisor, Dr. Aaron Ward. Aaron’s passion, work ethic, and guidance are incredibly inspirational. These qualities have been invaluable in allowing me to achieve my full potential, both professionally and personally. Aaron’s emphasis on effort, communication and collaboration, and a proper work-life balance are reflected among my fellow lab members resulting in an amazing lab environment.

I would like to thank my collaborators and lab members I have had the pleasure to work with over the course of my graduate studies. Being part of a lab that fostered open communication and collaboration was essential to ensure the highest quality research could be generated. Special thanks to my fellow lab members who always made time to help out when needed.

Finally, I would like to thank my family for supporting over the years. Their continued support has been essential to the completion of this thesis.

Many sources of funding have supporting this work, including the Canadian Institutes of Health Research, Cancer Care Ontario, the Ontario Institute for Cancer Research, the National Sciences and Engineering Research Council of Canada Computer Assisted Medical Interventions CREATE Award, and the Western Graduate Research Scholarship.
Table of Contents

Abstract .................................................................................................................................................. i
Co-Authorship Statement ...................................................................................................................... ii
Acknowledgments ............................................................................................................................... iii
List of Tables ......................................................................................................................................... vi
List of Figures ....................................................................................................................................... vii
List of Abbreviations ........................................................................................................................... viii
1 Chapter 1 ........................................................................................................................................... 1
  1.1 Introduction ............................................................................................................................... 1
  1.2 Background ................................................................................................................................ 3
    1.2.1 Prostate anatomy .................................................................................................................. 3
    1.2.2 Prostate cancer epidemiology ............................................................................................. 4
    1.2.3 Prostate cancer diagnosis .................................................................................................... 5
    1.2.4 Prostate cancer treatment .................................................................................................... 7
    1.2.5 Prostate cancer imaging ....................................................................................................... 10
    1.2.6 Computer-aided diagnosis systems for cancer ................................................................. 14
    1.2.7 Prostate CAD systems ........................................................................................................ 14
    1.2.8 Reference standards ............................................................................................................ 17
    1.2.9 Image preprocessing ............................................................................................................ 18
  1.3 Feature extraction ...................................................................................................................... 19
    1.3.2 Intensity-based features ...................................................................................................... 21
    1.3.3 Edge-based features ............................................................................................................ 21
    1.3.4 Texture-based features ........................................................................................................ 21
    1.3.5 Haralick features ................................................................................................................. 22
List of Tables

Table 1.1: Summary of CAD systems with some of their defining characteristics. 15

Table 2.2: Radiologist true positives and false positives at different PIRADS thresholds. R1-R4 indicates the four radiologists. 53

Table 2.3: First and second order texture features used in this study. 55

Table 2.4: Classification of malignant and benign tissue. 59

Table 2.5: Classification of malignant and confounding tissue in the form of radiologist false positives. 60

Table 2.6: Classification of high and low Gleason grade. 60
List of Figures

Figure 1-1: Prostate zones within the prostate in the sagittal plane. PZ = peripheral zone, CZ = central zone, TZ = transition zone, US = urethral sphincter, AS = anterior fibromuscular stroma. (Adapted from the American college of radiology web site.) ........ 3

Figure 1.2: (a) T2W MRI of a prostate, showing a cancerous lesion as a hypointensity (arrow) in the PZ. (b) ADC map of the same slice, showing the lesion as hypointensity (arrow)........................................................................................................................................................................ 11

Figure 1.3: (a) Whole-mount digitized histology, (b) Pathologist contours, (c) masks derived from whole-mount digitized histology......................................................................................................................... 14

Figure 1-4: Histogram showing some common first order texture features. ............... 20

Figure 1-5: Two images with the same mean intensity of 0.5 but second order energy of 0.309 and 0.5 for a) and b) respectively................................................................................................................................. 20

Figure 1-6: SVM (left), and KNN (right) decision boundaries. ................................. 28

Figure 2.5: Patient eligibility (left), and characteristics (right). .... Error! Bookmark not defined.

Figure 2.6: Illustration of extracted ADC texture features showing: a) energy, b) entropy, correlation and contrast calculated using a square window with a length and width of 2mm centered on each voxel........................................................................................................................................... 54
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D</td>
<td>Three-dimensional</td>
</tr>
<tr>
<td>ADC</td>
<td>Apparent diffusion coefficient</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial neural network</td>
</tr>
<tr>
<td>AUC</td>
<td>Area under the receiver operating characteristic curve</td>
</tr>
<tr>
<td>BPH</td>
<td>Benign prostatic hyperplasia</td>
</tr>
<tr>
<td>BT</td>
<td>Brachytherapy</td>
</tr>
<tr>
<td>CADe</td>
<td>Computer-aided diagnosis</td>
</tr>
<tr>
<td>CADx</td>
<td>Computer-aided detection</td>
</tr>
<tr>
<td>CG</td>
<td>Central gland</td>
</tr>
<tr>
<td>DCE</td>
<td>Dynamic contrast enhanced</td>
</tr>
<tr>
<td>DRE</td>
<td>Digital rectal examination</td>
</tr>
<tr>
<td>DS</td>
<td>Decision stump</td>
</tr>
<tr>
<td>DW</td>
<td>Diffusion-weighted</td>
</tr>
<tr>
<td>EBRT</td>
<td>External beam radiation therapy</td>
</tr>
<tr>
<td>FOV</td>
<td>Field of view</td>
</tr>
<tr>
<td>FPR</td>
<td>False positive rate</td>
</tr>
<tr>
<td>FNR</td>
<td>False negative rate</td>
</tr>
<tr>
<td>GLCM</td>
<td>Gray level co-occurrence matrix</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>HOG</td>
<td>Histogram of oriented gradient</td>
</tr>
<tr>
<td>KNN</td>
<td>K-nearest neighbours</td>
</tr>
<tr>
<td>LBP</td>
<td>Local binary pattern</td>
</tr>
<tr>
<td>LDA</td>
<td>Linear discriminant analysis</td>
</tr>
<tr>
<td>LLE</td>
<td>Local linear embedding</td>
</tr>
<tr>
<td>LR</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>LUTS</td>
<td>Lower urinary tract symptoms</td>
</tr>
<tr>
<td>mpMRI</td>
<td>Multi-parametric magnetic resonance imaging</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>MRF</td>
<td>Markov random field</td>
</tr>
<tr>
<td>MRS</td>
<td>Magnetic resonance spectroscopy</td>
</tr>
<tr>
<td>NB</td>
<td>Naïve Bayes</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PCa</td>
<td>Prostate cancer</td>
</tr>
<tr>
<td>PIN</td>
<td>Prostatic inter-epithelial neoplasia</td>
</tr>
<tr>
<td>PIRADS</td>
<td>Prostate imaging-reporting and data system</td>
</tr>
<tr>
<td>PNN</td>
<td>Probabilistic neural network</td>
</tr>
<tr>
<td>PSA</td>
<td>Prostate specific antigen</td>
</tr>
<tr>
<td>PZ</td>
<td>Peripheral zone</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>QDA</td>
<td>Quadratic discriminant analysis</td>
</tr>
<tr>
<td>RF</td>
<td>Random forest</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of interest</td>
</tr>
<tr>
<td>RP</td>
<td>Radical prostatectomy</td>
</tr>
<tr>
<td>SVM</td>
<td>Support vector machine</td>
</tr>
<tr>
<td>TFD</td>
<td>Texture fractal dimension</td>
</tr>
<tr>
<td>TPM</td>
<td>Trans-perineal mapping</td>
</tr>
<tr>
<td>TRUS</td>
<td>Transrectal ultrasound</td>
</tr>
<tr>
<td>T2W</td>
<td>T2-weighted</td>
</tr>
<tr>
<td>TE</td>
<td>Echo time</td>
</tr>
<tr>
<td>TR</td>
<td>Repetition time</td>
</tr>
<tr>
<td>WRS</td>
<td>Wilson rank-sum</td>
</tr>
</tbody>
</table>
1 Chapter 1.

1.1 Introduction

Prostate cancer (PCa) is one of the most prevalent cancers among men [1]. PCa makes up one in five new cancer diagnoses and is the second leading cause of cancer related death among men [1]. Initial PCa diagnosis is important to create an optimal treatment plan for each patient. Initial diagnosis consists of a digital rectal examination, prostate specific antigen (PSA) test, and an analysis of primary and secondary Gleason grade, if available. Patients with a Gleason score $\leq 6$, PSA $< 10$ and stage T1c to T2a [2] are usually placed on an active surveillance disease management strategy, where patients are monitored to see if the cancer becomes aggressive [2]. The T stage measures the location and extent of PCa. T stages $\leq 2$ reflect PCa that is confined within the organ. Living with untreated cancer presents a psychological burden for some patients, resulting in many patients choosing to undergo a procedure that may not have been needed. Prostatectomy is a procedure consisting of the removal of the entire prostate through surgery. Overtreatment with radical prostatectomy (RP) is estimated to occur in 20 to 60% of cases [3], resulting in significant deterioration of urinary and sexual function [3].

For a subset of patients who have localized PCa, therapies targeting PCa while sparing healthy tissue may be an intermediate option with fewer side effects, suitable for many patients. These therapies rely on pre-treatment target volume contours containing aggressive PCa. Many approaches have been investigated to create these pre-treatment target volume contours. Radiologist-identified regions from multi-parametric magnetic resonance imaging (mpMRI) have shown promise for detection of PCa [4], but substantial inter-observer variability in lesion contouring has been observed [5]. Many
computer-assisted diagnosis (CAD) models have been developed to improve lesion classification [6] [7] [4] [8] [9]. These systems can be used as second observers to improve physician performance, especially among novice radiologists [10]. Models that classify healthy tissue vs. malignant tissue have reached an area under the receiver operating characteristic curve (AUC) of 0.97 in the peripheral zone (PZ) [7] [11]. However, models that classify cancer from confounding regions have had lower accuracies, with Litjens et al. [12] achieving AUCs of 0.73, 0.75, 0.63, and 0.69 for classification of tumours vs. prostatic intraepithelial neoplasia (PIN), atrophy, inflammation and benign prostatic hyperplasia (BPH), respectively. These studies have limitations which have made widespread clinical implementation of computer-assisted PCa lesion characterization elusive, including small sample sizes and registration errors in reference standards due to spatial biases. Spatial biases may be the result of false-positive low-intensity benign regions, as low-intensity regions are indicative of prostate cancer on T2-weighted (T2W) MRI and apparent diffusion coefficient (ADC) maps derived from diffusion-weighted MRI, by the PIRADS guidelines [13].

Registration for CAD models is often done by radiological consensus, where radiologists confer with each other and/or a pathologist to determine tumour location and boundaries. This consensus mapping approach is most straightforward when identifying true-positive mpMRI lesions corresponding to histologic foci, but identifying mpMRI false-positive and false-negative lesions is more challenging and the attendant uncertainty in the error in spatial mapping remains.
Figure 1-1: Prostate zones within the prostate in the sagittal plane. PZ = peripheral zone, CZ = central zone, TZ = transition zone, US = urethral sphincter, AS = anterior fibromuscular stroma. (Adapted from the American college of radiology web site.)

Widespread clinical implementation of computer-assisted PCa lesion characterization remains elusive. Critically needed is a model that is validated against a histologic reference standard that is densely sampled in an unbiased fashion. Using our technique for highly accurate fusion of mpMRI with whole-mount digitized histology of the surgical specimen, we can address these biases [5].

1.2 Background

1.2.1 Prostate anatomy

The prostate is divided into 4 anatomical zones. The PZ covers 70% of the gland and extends from the base to the apex of the prostate. The central 25% of the gland that wraps around the urethra is the central zone (CZ). The transition zone (TZ) is made up of
the remaining 5% of the gland. The final zone is the non-glandular anterior fibromuscular stroma.

1.2.2 Prostate cancer epidemiology

Prostate cancer is the most common cancer identified in North American men [1]. In 2016, prostate cancer (PCa) is estimated to account for 21% of new cases of cancer and is the second leading cause of death of cancer in men in the United States, with one in eight men being diagnosed with PCa during his lifetime [1]. PCa progression is highly variable [14]. The variable nature of disease progression limits treatment options to either a radical therapy or a watch and wait strategy. 70% of cancer appears in the PZ, 25% in the TZ and 5% in the CZ [15]. Up to 90% of patients with recently diagnosed PCa are treated even when the disease is unlikely to have caused any major negative effects on their quality of life [16]. Additionally, many patients who elect for a watch and wait strategy have to endure the psychological burden of living with an untreated cancer.

Thus, a key challenge in PCa treatment selection is predicting which patients do and do not need treatment. Computer models may have a role in addressing this challenge as models have been shown to improve the diagnostic accuracy of PIRADS scoring when combining the system score with a radiologist score [10], as well as classifying Gleason grade 6 vs. Gleason grade $\geq 7$ cancer [9, 10, 17]. Models may also be able to address secondary objectives including accurate stratification of patients through optimized targeted biopsy procedures, monitoring low-risk patients through longitudinal measurements of tumour burden, and targeting curative-intent therapies for both aggressive and less aggressive treatments.
1.2.3 Prostate cancer diagnosis

Diagnosis of PCa follows various pathways depending on the patient’s history and symptoms.

1.2.3.1 Lower urinary tract symptoms

If a patient presents lower urinary tract symptoms (LUTS) a primary care provider will assess and treat the patient. PCa treatment may exacerbate LUTS, and therefore the severity and impact on quality of life should be carefully considered before diagnosis or treatment [18].

1.2.3.2 Unexplained symptoms

If a patient is 40 years old or older and has unexplained lower back pain with reproducible percussion tenderness, severe bone pain, or weight loss, especially in the elderly, they will receive a digital rectal exam (DRE) and prostate specific antigen (PSA) test [19].

1.2.3.3 Digital rectal exam

In the context of PCa screening, sensitivity and specificity are used to measure the fractions of positives and negatives that are correctly identified. A DRE is a procedure in which a physician manually palpates the patient’s prostate with a gloved finger in order to identify abnormally dense or asymmetric regions [20]. DRE has a sensitivity of 52.3% and specificity of 83.6% [21] [22].

1.2.3.4 PSA testing

PSA is a protein primarily produced by prostate cells. PSA testing measures the level of PSA in a blood sample. A PSA level above 10 ng/mL is considered suspicious
for cancer. However, other conditions such as prostatitis and benign prostatic hyperplasia (BPH) can also increase PSA levels [23]. Prostatitis comes in the form of an infection or inflammation. Prostatitis is classified into four categories: acute bacterial, chronic bacterial, chronic nonbacterial and chronic pelvic pain syndrome [24]. BPH is a condition that enlarges the prostate. The enlarged prostate can squeeze the urethra, causing urinary incontinence [25]. BPH occurs in almost all men as they age and is not considered to be a major health concern. PSA testing has a sensitivity of 72.1% and specificity of 93.2% [22].

1.2.3.5 Biopsy
If a patient presents with symptoms of intermediate or aggressive PCa, they will receive a prostate biopsy. These symptoms include a PSA level > 10 and/or suspicious nodes found during the DRE. Gleason scores are used to determine the aggressiveness of the disease. A Gleason score is determined by inspection of architectural patterns, size and spacing of the tumour glands [26]. During a biopsy procedure, tissue samples are taken from the prostate through the rectal wall or less commonly through the perineum using a biopsy needle. Biopsy can be performed using transrectal ultrasound and/or MRI for guidance and targeting. 6–12 biopsy samples are obtained primarily in the PZ, as that is where most cancers are found [27]. Tissue obtained from the biopsy is then examined to determine the number of cores containing cancer and the extent of cancer in each core. A Gleason grading system is used to identify the aggressiveness of the cancer obtained from the biopsy [26]. The Gleason grading system is used to assign a Gleason grade to tissue using the microscopic architecture and differentiation of prostate glands on a five point scale [28, 29]. Grading biopsy tissue is a challenge. There is a 40% rate of discrepancy
between biopsy and prostatectomy Gleason grade [30, 31]. By adding the primary and secondary Gleason grades of tissue, a Gleason score can be obtained. Many studies have used the Gleason score for classification of tissue.

1.2.4 Prostate cancer treatment

Once a patient has been diagnosed with prostate cancer, an appropriate treatment plan will be selected with assistance from their physician. Treatment options will vary depending on the risk level associated with a particular patient.

1.2.4.1 Watchful waiting and active surveillance

Watchful waiting (also referred to as conservative management, expectant management and deferred treatment) is a palliative disease management strategy for monitoring patients while not treating them [2]. Active surveillance is intended as a curative disease management strategy that initiates curative treatment if and when necessary [32]. In active surveillance, patients are closely monitored for signs of progression that include frequent biopsies and PSA tests [33]. If progression is suspected, these patients can proceed to radical treatments [33]. Active surveillance is recommended for patients who have low-risk cancer. Low risk cancers have a Gleason score \( \leq 6 \), PSA\(<10\text{ng/ml} \) and clinical stage T1c–T2a [33].

1.2.4.2 Prostate-focused therapies

Radical prostatectomy and radical radiotherapy are the main treatment choices for localized prostate cancer. These treatment methods treat the entire prostate to the same degree. Radical prostatectomy is a procedure where the entire prostate is removed through surgery. Radical prostatectomy is used to treat patients with cancer that does not extend beyond the prostate capsule. The primary side effects of a radical prostatectomy
include high rates of erectile dysfunction (34–90% [34, 35]), urinary incontinence (8–13% [34, 36]) and other complications (10–30% perioperatively and 4–10% within a year [34, 37]). Radiation therapy can refer to multiple procedures that irradiate the prostate in order to damage DNA. The first type of radiation is external beam radiation therapy (EBRT). This is a procedure in which high energy x-rays generated outside the body are focused on the prostate from multiple angles. A second type of radiation therapy is low-dose-rate brachytherapy (BT). This is a procedure where radioactive seeds are permanently inserted into the prostate. High-dose-rate BT is another type of radiation therapy where a highly radioactive emitter is moved within the prostate during the procedure and then removed. Radiation therapy patients have high rates of erectile dysfunction. A meta-analysis of patients with localized PCa who underwent radiotherapy found the probability of maintaining erectile dysfunction after treatment was 80%, 69%, 68%, 22%, 16% and 13% for BT alone, BT + EBRT, EBRT alone, nerve sparing RP, standard RP, and cryotherapy, respectively [35]. Brachytherapy resulted in urinary incontinence in 6.6% of men after the procedure. For external beam radiation, minor to moderate urinary symptoms were present in 30% of men.

1.2.4.3 Lesion-focused therapies

In order to reduce the side effects of radical therapy, emerging treatments that preferentially target the prostate cancer lesions are being investigated. These treatments may allow for strategically escalating treatment to tumours in patients to achieve better outcomes.
1.2.4.3.1 Whole-gland radiation with focal boost

A local recurrence may occur in the prostate after a radical radiotherapy, typically at the site of the dominant lesion [38]. Technological advances for planning and delivering radiation enable radiation oncologists to deliver different doses to specific regions within the prostate [39, 40].

1.2.4.3.2 Prostate-sparing and lesion-only therapies

Although prostate cancer is multifocal in 67–82% of patients [41-43], there is a subset of patients who have a dominant lesion surrounded by non-cancerous tissue. There is also evidence that the size of the largest cancer focus is the primary predictor of risk of recurrence and biochemical failure, even in multifocal prostate cancers [43]. These findings have led to the development of treatment methods that target the cancer while sparing outside healthy tissue. Some of these treatment modalities include high-intensity focused ultrasound [44], cryotherapy [45], photodynamic therapy [46], laser ablation [47] and radiation therapy [48]. Early evidence from studies of several focal therapies suggests that they successfully reduce the side effects prevalent in radical therapies [49-51]. Cryotherapy studies have reported incontinence rates lower than 5% and erectile dysfunction rates of 10–35% [45]. High-intensity focused ultrasound studies have reported incontinence rates less than 10% and erectile dysfunction rates of 5–11% [52]. A recent consensus panel suggested that delineation of cancerous regions should be performed using mpMRI [53]. These suggestions imply prostate cancer imaging and lesion delineation could play an important role in focal therapy planning.
1.2.5  Prostate cancer imaging

There are many methods of imaging used to aid in PCa diagnosis and treatment planning. The focus of this thesis is on mpMRI for lesion classification.

1.2.5.1  Magnetic resonance imaging

MRI is used to provide rich anatomical and functional information. It allows for control of contrast using different acquisition sequences. MRI is being used in many areas of PCa diagnosis. Detection [13], staging [54, 55], localization [56-58], therapy planning [40, 53, 59], and therapy monitoring [60] are all areas where localized prostate cancer treatment and planning have been improved.

Inter-observer variability in localization of prostate cancer on MRI is very large [53]. The following sections describe sequences and imaging protocols that have been investigated for imaging localized prostate cancer.

1.2.5.2  T2W MRI

T2-weighted MRI is used to visualize the structural components of the prostate. T2-weighted MRI measures the transverse relaxation due to the dephasing of atomic nuclei. In the PZ, tumours appear hypo-intense relative to benign peripheral zone tissue [61]. Tumours in the central gland are generally homogenous with low-intensity, ill-defined irregular boundaries, and a lenticular shape [62]. Figure 1.2(a) shows a T2W MRI of a prostate, with a hypo-intense region in the PZ (indicated by an arrow) showing a cancerous lesion.
Figure 1.2: (a) T2W MRI of a prostate, showing a cancerous lesion as a hypointensity (arrow) in the PZ. (b) ADC map of the same slice, showing the lesion as hypointensity (arrow).

1.2.5.3 Diffusion-weighted MRI

Diffusion-weighted (DW) MRI is used to measure the rate of diffusivity of water molecules in tissue. In tumours, diffusivity drops due to the increased cellular density associated with cancer [63]. DW MRI post-processing is used to create apparent diffusion coefficient (ADC) maps computed from multiple DW MR images taken with different b-values [13]. A b-value measures the strength and timing of the gradients used to generate the images. Higher b-values represent a stronger diffusion effect. $S_0$ is the MR signal at baseline, $D$ is the diffusion coefficient and the signal, $(S)$ after diffusion gradients have been applied is given by:

$$S = S_0 e^{-bD} \quad (1)$$

Prostate cancer appears hypo-intense on ADC maps. DW MRI has been shown to have higher sensitivity for central gland tumours compared to other MRI sequences [63, 64]. ADC values have also been shown to be correlated with Gleason score [65-67]. The main
limitation of DW MRI is low resolution, [13, 63] but ADC also suffers from distortion artifacts [68, 69]. Figure 1.2(b) shows an ADC map of the same slice of the prostate as shown in Figure 1.2(a), with a hypo-intense region in the PZ (indicated by an arrow) showing a cancerous lesion.

1.2.5.4 Dynamic contrast-enhanced (DCE) MRI

DCE MRI is used to detect changes in vascular characteristics associated with angiogenesis due to cancer [70, 71]. A DCE MRI pulse sequence rapidly images the patient using a sequence of T1-weighted MR images prior to and for 5-10 minutes after introducing a gadolinium-chelate contrast agent. DCE MRI creates contrast based on how quickly the contrast agent washes in and out of tissue. In cancerous tissue, the contrast agent washes in and out faster than in non-cancerous tissue. Pharmacokinetic parameters can be modelled by combining faster sequences with estimates of the arterial input function [70, 71]. Pharmacokinetic parameters such as the contrast transfer coefficient $k_{\text{trans}}$ and the rate constant $k_{\text{ep}}$ can be used to detect the presence of cancer [72]. Patient motion is the main challenge for DCE MRI. Secondary challenges include the presence of confounding non-cancerous abnormalities that mimic cancer, such as prostatitis in the peripheral zone, and highly vascularized BPH in the central gland [71].

1.2.5.5 MR spectroscopy

MR spectroscopy (MRS) is an imaging method that measures concentrations of metabolites that can be distinguished by the resonance frequencies of their protons. These protons include choline, creatine, and citrate [73]. Elevated choline and decreased citrate are typically found in cancerous tissue. Choline and creatine are difficult to separate from each other [70], so a (choline+creatine)/citrate ratio is used for diagnosis. This ratio has
been shown to be correlated with Gleason score [74]. MRS interpretation is challenging because of contamination of the signal from nearby lipids, high sensitivity to shimming, operator variability, and common abnormalities that mimic cancer such as inflammation and BPH [70, 72, 73, 75].

1.2.5.6 Transrectal ultrasound

Transrectal ultrasound (TRUS) imaging is used mainly to estimate prostate volume and for image guidance during biopsy and needle-based therapy procedures. TRUS imaging allows for rapid visualization of the prostate gland and the boundary between prostate zones. However, the sensitivity for detecting prostate cancer on TRUS is low at 31% [76-79].

1.2.5.7 Histology in prostate cancer

A histological examination is used at several points in the clinical workflow. Histological examination of biopsy tissue is used to follow up on initial detection in screening populations, for ongoing monitoring in active surveillance and to confirm recurrence in post-treatment populations. Histological examination is also used to assess prognosis for patients post-prostatectomy [80], and to inform decisions about adjuvant therapy [81]. The Gleason grading system categorizes tissue based on its microscopic morphological appearance including glandular differentiation and growth patterns within the stroma. The system can be used to measure the aggressiveness of a particular tissue which has been shown to correlate with patterns on mpMRI.
1.2.6 Computer-aided diagnosis systems for cancer

Computer-aided diagnosis (CAD) techniques have been investigated for many types of cancer diagnosis. These types include breast [82, 83], lung [84-86], and colorectal [87]. CAD techniques for PCa have not been investigated until recently [88].

1.2.7 Prostate CAD systems

The first model was proposed by Chan et al. in 2003 [89]. Since then, there have been over 50 studies proposed for various tasks related to PCa diagnosis. Table 1.1 shows the main studies and their principal characteristics. The first row of the table describes the work presented in this thesis. Note that it is distinguished from all of the other studies in the table because it classifies malignant, benign, and confounding tissue and high and low Gleason grade on T2W and ADC MRI which has not been done before.
Table 1.1: Summary of CAD systems with some of their defining characteristics: probabilistic neural network (PNN), linear discriminant analysis (LDA), support vector machine (SVM), Markov random field (MRF), random forest (RF), naive Bayes (NB), decision stump (DS), logistic regression (LR), artificial neural network (ANN), quadratic discriminant analysis (QDA), cross validation (X-val), leave-one-out (LOO), Wilson rank-sum (WRS), MRI modalities (Mods) regions of interest (ROIs) malignant (M), benign (B), confounding tissue (C), radiologist false positives (RFP), biopsy (B), whole-mount digitized histology (WM), whole-mount digitized histology registration with quantified registration error (WMQ) and guided biopsy (GB). The first row of the table describes the work presented in this thesis.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Classifiers</th>
<th>X-val</th>
<th>Accuracy</th>
<th>Mods</th>
<th>n</th>
<th>Field</th>
<th>Zone</th>
<th>ROIs</th>
<th>Ref. Std.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soetemans</td>
<td>2016</td>
<td>SVM, KNN</td>
<td>k-fold</td>
<td>AUC 0.96</td>
<td>T2W</td>
<td>25</td>
<td>ADC</td>
<td>PZ,</td>
<td>M, B,</td>
<td>WMQ,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ADC</td>
<td></td>
<td></td>
<td>CG,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ampeliotis</td>
<td>2008</td>
<td>PNN</td>
<td>LOO</td>
<td>AUC: 0.9</td>
<td>T2W</td>
<td>25</td>
<td>DCE</td>
<td>1.5</td>
<td>PZ,</td>
<td>M, B, B</td>
</tr>
<tr>
<td></td>
<td>[8]</td>
<td></td>
<td></td>
<td></td>
<td>DCE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ampeliotis</td>
<td>2007</td>
<td>PNN</td>
<td>LOO</td>
<td>AUC: 0.87, 0.84</td>
<td>T2W</td>
<td>25</td>
<td>DCE</td>
<td>1.5</td>
<td>PZ,</td>
<td>M, B, B</td>
</tr>
<tr>
<td></td>
<td>[90]</td>
<td></td>
<td></td>
<td></td>
<td>DCE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antic [91]</td>
<td>2013</td>
<td>LDA</td>
<td>LOO</td>
<td>AUC: 0.94</td>
<td>T2W</td>
<td>53</td>
<td>ADC</td>
<td>1.5</td>
<td>PZ,</td>
<td>M, B, WM</td>
</tr>
<tr>
<td>Artan [92]</td>
<td>2010</td>
<td>SVM, MRF</td>
<td>LOO</td>
<td>Sen: 0.66, Spe: 0.74</td>
<td>T2W</td>
<td>21</td>
<td>DCE</td>
<td>1.5</td>
<td>PZ,</td>
<td>M, B, WM</td>
</tr>
<tr>
<td>Artan [93]</td>
<td>2009</td>
<td>SVM, MRF</td>
<td>LOO</td>
<td>Sen: 0.74, Spe: 0.82</td>
<td>T2W</td>
<td>10</td>
<td>DCE</td>
<td>1.5</td>
<td>PZ,</td>
<td>M, B, B</td>
</tr>
<tr>
<td>Chan [89]</td>
<td>2003</td>
<td>LDA, SVM</td>
<td>LOO</td>
<td>AUC: 0.84</td>
<td>T2W</td>
<td>15</td>
<td>ADC</td>
<td>1.5</td>
<td>PZ,</td>
<td>M, B, B</td>
</tr>
<tr>
<td>Giannini [94]</td>
<td>2015</td>
<td>SVM</td>
<td>LOO</td>
<td>Sen: 0.79, Spe: 0.84</td>
<td>T2W</td>
<td>56</td>
<td>ADC</td>
<td>1.5</td>
<td>PZ,</td>
<td>M, B, WM</td>
</tr>
<tr>
<td>Giannini [95]</td>
<td>2013</td>
<td>NB</td>
<td>LOO</td>
<td>AUC: 0.87</td>
<td>T2W</td>
<td>10</td>
<td>DCE</td>
<td>1.5</td>
<td>PZ,</td>
<td>M, B, WM</td>
</tr>
<tr>
<td>Kelm [96]</td>
<td>2007</td>
<td>LDA, RF, Kernel, SVM</td>
<td>LOO</td>
<td>N/A</td>
<td>MRSI</td>
<td>24</td>
<td></td>
<td>PZ,</td>
<td>M, B,</td>
<td>B</td>
</tr>
<tr>
<td>Kwak [97]</td>
<td>2014</td>
<td>SVM</td>
<td>WRS</td>
<td>AUC: 0.89</td>
<td>T2W</td>
<td>244</td>
<td>ADC</td>
<td>WP</td>
<td>M, B, C B</td>
<td></td>
</tr>
<tr>
<td>Langer [98]</td>
<td>2009</td>
<td>LR</td>
<td>WRS</td>
<td>AUC: 0.71</td>
<td>T2W</td>
<td>25</td>
<td>ADC</td>
<td>1.5</td>
<td>PZ,</td>
<td>M, B, WM</td>
</tr>
<tr>
<td>Litjens [4]</td>
<td>2014</td>
<td>LDA, RF</td>
<td>LOO</td>
<td>AUC: 0.81</td>
<td>T2W</td>
<td>347</td>
<td>ADC</td>
<td>3</td>
<td>WP,</td>
<td>M, B, GB</td>
</tr>
<tr>
<td>Litjens [99]</td>
<td>2012</td>
<td>KNN, SVM</td>
<td>LOO</td>
<td>Sen: 0.75, Spe: 0.83</td>
<td>T2W</td>
<td>288</td>
<td>ADC</td>
<td>3</td>
<td>PZ,</td>
<td>M, B, GB</td>
</tr>
</tbody>
</table>

The following classifiers are used:
- SVM: Support Vector Machine
- KNN: K-Nearest Neighbors
- LDA: Linear Discriminant Analysis
- RF: Random Forest
- NB: Naive Bayes
- DS: Decision Stump
- LR: Logistic Regression
- ANN: Artificial Neural Network
- QDA: Quadratic Discriminant Analysis
- X-val: Cross Validation
- LOO: Leave-One-Out
- WRS: Wilson Rank-Sum
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Model</th>
<th>Method</th>
<th>Parameters</th>
<th>T2W</th>
<th>DCE</th>
<th>ADC</th>
<th>Score</th>
<th>Class</th>
<th>WM</th>
<th>M</th>
<th>B</th>
<th>GB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Litjens [100]</td>
<td>2011</td>
<td>RBSVM</td>
<td>k-fold</td>
<td>Sen: 0.75, Spe: 0.83</td>
<td>T2W</td>
<td>DCE</td>
<td>ADC</td>
<td>188</td>
<td>3</td>
<td>WP</td>
<td>M, B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>Litjens [10]</td>
<td>2015</td>
<td>RF</td>
<td>AUC: 0.81</td>
<td>T2W</td>
<td>DCE</td>
<td>ADC</td>
<td>130</td>
<td>3</td>
<td>PZ, CG</td>
<td>M, B, GG</td>
<td>GB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liu [101]</td>
<td>2013</td>
<td>SVM</td>
<td>AUC: 0.83</td>
<td>T2W</td>
<td>DCE</td>
<td>ADC</td>
<td>54</td>
<td>3</td>
<td>PZ, CG</td>
<td>M, B</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liu [102]</td>
<td>2009</td>
<td>SVM, MRF</td>
<td>Sen: 0.9, Spe: 0.88</td>
<td>T2W</td>
<td>DCE</td>
<td>ADC</td>
<td>11</td>
<td>1.5</td>
<td>PZ</td>
<td>M, B</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lopes [6]</td>
<td>2011</td>
<td>AdaBoost</td>
<td>SVM</td>
<td>k-fold</td>
<td>Sen: 0.85, Spe: 0.93</td>
<td>T2W</td>
<td>27</td>
<td>1.5</td>
<td>PZ, CG</td>
<td>M, B</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lv [7]</td>
<td>2009</td>
<td>DS</td>
<td>AUC: 0.97</td>
<td>T2W</td>
<td>55</td>
<td>1.5</td>
<td>PZ</td>
<td>M, B</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matulewicz [103]</td>
<td>2012</td>
<td>ANN</td>
<td>k-fold</td>
<td>Sen: 0.6, Spe: 0.66</td>
<td>MRSI</td>
<td>18</td>
<td>3</td>
<td>PZ, CG</td>
<td>M, B</td>
<td>WM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mazzetti [104]</td>
<td>2011</td>
<td>NB, LOO</td>
<td>Sen: 0.82, Spe: 0.82</td>
<td>DCE</td>
<td>10</td>
<td>1.5</td>
<td>PZ</td>
<td>M, B</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moradi [105]</td>
<td>2012</td>
<td>SVM</td>
<td>LOO</td>
<td>AUC: 0.96</td>
<td>DCE</td>
<td>29</td>
<td>3</td>
<td>PZ</td>
<td>M, B, GG</td>
<td>WM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Niaf [106]</td>
<td>2014</td>
<td>SVM, LDA, KNN, NB</td>
<td>LOO</td>
<td>AUC: 0.89</td>
<td>T2W</td>
<td>DCE</td>
<td>ADC</td>
<td>49</td>
<td>1.5</td>
<td>PZ</td>
<td>M, B</td>
<td>WM</td>
<td></td>
</tr>
<tr>
<td>Niaf [107]</td>
<td>2012</td>
<td>SVM, LDA, KNN, NB</td>
<td>LOO</td>
<td>AUC: 0.86</td>
<td>T2W</td>
<td>DCE</td>
<td>ADC</td>
<td>30</td>
<td>1.5</td>
<td>PZ</td>
<td>M, B</td>
<td>WM</td>
<td></td>
</tr>
<tr>
<td>Niaf [108]</td>
<td>2010</td>
<td>SVM, LDA, KNN, NB</td>
<td>LOO</td>
<td>AUC: 0.89</td>
<td>T2W</td>
<td>DCE</td>
<td>ADC</td>
<td>23</td>
<td>1.5</td>
<td>PZ</td>
<td>M, B</td>
<td>WM</td>
<td></td>
</tr>
<tr>
<td>Ozer [109]</td>
<td>2010</td>
<td>SVM, MRF</td>
<td>LOO</td>
<td>Sen: 0.78, Spe: 0.74</td>
<td>T2W</td>
<td>DCE</td>
<td>ADC</td>
<td>20</td>
<td>1.5</td>
<td>PZ</td>
<td>M, B</td>
<td>WM</td>
<td></td>
</tr>
<tr>
<td>Ozer [110]</td>
<td>2009</td>
<td>SVM</td>
<td>LOO</td>
<td>Sen: 0.76, Spe: 0.75</td>
<td>T2W</td>
<td>DCE</td>
<td>ADC</td>
<td>20</td>
<td>1.5</td>
<td>PZ</td>
<td>M, B</td>
<td>WM</td>
<td></td>
</tr>
<tr>
<td>Parfait [111]</td>
<td>2012</td>
<td>SVM, ANN</td>
<td>k-fold</td>
<td>Sen: 0.84, Spe: 0.97</td>
<td>MRSI</td>
<td>22</td>
<td>3</td>
<td>PZ</td>
<td>M, B</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peng [112]</td>
<td>2013</td>
<td>SVM</td>
<td>LOO</td>
<td>Sen: 0.82, Spe: 0.95</td>
<td>T2W</td>
<td>DCE</td>
<td>ADC</td>
<td>48</td>
<td>3</td>
<td>PZ, CG</td>
<td>M, B, GG</td>
<td>WM</td>
<td></td>
</tr>
<tr>
<td>Puech [113]</td>
<td>2009</td>
<td>LDA, RS</td>
<td>LOO</td>
<td>Sen: 1, Spe: 0.43</td>
<td>DCE</td>
<td>100</td>
<td>1.5</td>
<td>PZ, CG</td>
<td>M, B</td>
<td>WM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rampun [114]</td>
<td>2016</td>
<td>BN, RF, KNN</td>
<td>k-fold</td>
<td>AUC: 0.93</td>
<td>T2W</td>
<td>45</td>
<td>1.5</td>
<td>PZ</td>
<td>M, B</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shah [115]</td>
<td>2012</td>
<td>SVM</td>
<td>LOO</td>
<td>f-meas: 0.85</td>
<td>T2W</td>
<td>DCE</td>
<td>ADC</td>
<td>31</td>
<td>3</td>
<td>PZ</td>
<td>M, B</td>
<td>WM</td>
<td></td>
</tr>
<tr>
<td>Sung [116]</td>
<td>2011</td>
<td>SVM</td>
<td>LOO</td>
<td>Sen: 0.90, Spe: 0.77</td>
<td>DCE</td>
<td>42</td>
<td>3</td>
<td>PZ, CG</td>
<td>M, B</td>
<td>WM</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
There are five general categories for reference standards. (1) Standard clinical practice for biopsy uses TRUS targeting and TRUS guidance to deliver needles in a sextant arrangement through the prostate [132]. This is the least accurate reference standard because ultrasound is inferior to MRI for targeting and guidance [133]. (2) The second category of reference standards uses MR targeting and TRUS guidance to aim
needles at suspicious areas [134-136]. This technique can be improved using fusion guided biopsy, a technique in which pre-biopsy MRI are fused with real-time TRUS images [137]. This method improves accuracy but does not allow for confirmation of needle placement because the biopsy is taken prior to the scan to confirm placement [137]. (3) MR guidance can be used with MR targeting to improve biopsy needle placement further as MR guidance allows for confirmation of needle position. (4) Transperineal mapping (TPM) is the most accurate biopsy-based reference standard. TPM takes many samples from the prostate that are projected into a single 2-dimensional plane to get an accurate representation of the prostate [138]. Ideally, each sample is processed and stored separately to allow for accurate location information of any tumour(s). Biopsies are small cylindrical samples of tissue (~18 mm in length and 1 mm in diameter) [139]. These tissue samples are used to determine if cancer is present. (5) Whole-mount histological sections is the final and most accurate reference standard. A histological examination is a readily accepted reference standard used to verify the presence, location and grade of prostate tumours. This examination refers to a pathologist’s inspection of tissue on a microscopic level to identify tissue properties from a ~4-µm-thick section. Whole-mount sections are typically registered to mpMRI by radiologist/pathologist consensus mapping.

1.2.9 Image preprocessing

CAD systems generally include a preprocessing step to normalize the MR data or to transform the data into a specific domain. Inter-image variability is large, even when the same protocol and same scanner is used. Noise filtering, bias correction and signal intensity standardization are the three main groups of pre-processing methods for prostate
Noise filtering is used to correct for noise is caused by inhomogeneity in the main magnetic field during an MRI acquisition [88]. The simplest type of noise filtering is median filtering [109]. However, since the noise follows a Rician distribution, median filters are not well formalized to address the noise distribution in MR images [88]. Wavelet-based filtering is a common approach used to address this limitation [6] [109]. Bias correction techniques are used to correct the inhomogeneity of the bias field. Parametric and non-parametric methods are used to correct for the smooth variation of the signal intensity throughout the image [7] [124]. When using an endorectal coil, a hyper-intense signal around the coil appears. This can make segmentation and registration more challenging [140] [141].

1.3 Feature extraction

Discriminative features can be used to identify PCa. Choosing the right features for a classifier is more important than choosing the classifier itself [142]. First order features include basic statistics measures such as mean, median and percentiles. These features, illustrated in Figure 1.4, are good for identifying basic differences such as hypo-intense
regions; however these features do not take into consideration spatial distributions.

**Figure 1-4:** Histogram showing some common first order texture features.

Second order texture features look at a neighbourhood of pixels in order to measure differences in spatial distributions of pixels. These features can identify differences first order features miss, as illustrated in Figure 1.5.

**Figure 1-5:** Two images with the same mean intensity of 0.5 but second order energy of 0.309 and 0.5 for a) and b) respectively.
1.3.1.1 Statistics-based features

The most common approach can be classified as statistical methods, or first order texture features [88]. A feature map is first computed for the whole image. Statistics are extracted from regions of interest (ROIs).

1.3.2 Intensity-based features

PCa can be detected using MRI signal intensity data. They are one of the most common features used in CAD systems. In order to mitigate inter-scan variability in MR images, signal intensity standardization is often done. Litjens et al. [4] used the MR signal equation, the proton density image and a reference tissue to estimate a T2-estimate map.

1.3.3 Edge-based features

Edge based features can be created by convolving the original image with an edge operator. There are three main types of edge operators: Prewitt, Sobel, and Kirsch [4]. Prewitt and Sobel operators are linear edge detectors that compute a discrete differentiation operator to find an approximation of the gradient of the image intensity function. Kirsch operators are a non-linear edge detector that finds the maximum edge strength in a few predetermined directions. A Gabor filter is another linear filter that uses frequency and orientation representations that are similar to those of the human visual system.

1.3.4 Texture-based features

Radiomics is emerging as an area of study that looks to extract additional information from medical images [143, 144]. Radiomics aims to enable more personalized treatment
for patients by tailoring their treatment based on predicted response on pre-treatment imaging features [145]. Radiomics depends on quantitative image feature analysis, which involves the extraction of measurable texture information from an image. Second-order and third-order features can be extracted in many ways, including statistical methods, structural methods, model-based methods and transform-based methods [146]. Statistical texture analysis uses high-order statistics of an image intensity histogram. A grey-level co-occurrence matrix (GLCM) can be used to calculate GLCM texture features by analyzing neighbouring pixel pairs. The GLCM $C$, defined over an $n \times m$ image $I$, parameterized by an offset $(\Delta x, \Delta y)$ can be defined as

$$C_{\Delta x, \Delta y}(i, j) = \sum_{p=1}^{n} \sum_{q=1}^{m} \left\{ 1, \text{if } I(p, q) = i \text{ and } I(p + \Delta x, q + \Delta y) = j \right\} 0, \text{otherwise}$$

(1)

1.3.5 Haralick features

The GLCM is typically large and sparse, and therefore various metrics of the matrix can be taken to generate a more useful set of features [147]. These features include contrast, correlation, sum of squares, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, measure of correlation 1 and 2 and maximum correlation coefficient. Contrast is a commonly used texture feature that is measured by

$$\sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} C(i, j) \right\} \{i - j\} = n, (2)$$

where $N_g$ is the number of gray levels in the image. Correlation and homogeneity are also commonly used in CAD systems. Correlation can be calculated as
where $\mu_x$, $\mu_y$, $\sigma_x$, and $\sigma_x$ are the means and standard deviations of $p_x$ and $p_y$, the partial probability density functions.

1.3.5.1 Fractal analysis

Cancerous tissue shows a higher fractal dimension than healthy tissue. A fractal dimension is a ratio providing a statistical index of complexity comparing how detail in a pattern changes with the scale to which it is measured. To estimate the fractal dimension, a wavelet-based method in a multi-resolution framework can be used as reported by Lopes et al [6]. Lv et al. [7] calculate texture fractal dimension (TFD) as a geometric description of image textures. TFD is computed using

$$
\text{TFD} = \lim_{s \to 0} \frac{\log N_s}{\log (\frac{1}{s})},
$$

where $s$ is the box size and $N_s$ is the number of the boxes with size scale $s$.

1.3.5.2 Wavelet-based features

Wavelet filter decomposition approaches are used to extract fine structural image detail at different orientations and scales [123]. This may prove useful in quantitatively characterizing micro and macroscopic visual cues radiologist use when identifying PCa [123]. The integral wavelet transform is defined as

$$
[W\psi f](a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} \psi(\frac{x-b}{a}) f(x) dx.
$$

From this, the wavelet coefficients $c_{jk}$ are given by
where $a = 2^j$ is called the dyadic or binary dilation and $b = k2^{-j}$ is the dyadic or binary position.

### 1.3.5.3 Anatomical-based

In order to account for anatomical features, some studies use the relative distance to the prostate boundary as a feature. The relative position features in $x$, $y$, and $z$ directions are also used as features. The relative position features are calculated as

$$B(x) = \min_{y \in P_b} d(x, y), \quad (7)$$

where $x$ is the position of a voxel, $d$ is the Euclidean distance operator, $P$ is the set of prostate voxels and $P_b$ is the set of prostate boundary voxels. Litjens et al. [4] also used PZ probability features including intensity, texture and anatomical features to generate a likelihood map for each voxel belonging to the PZ [4]. The relative distance is computed as

$$\text{RD}(x) = \frac{B(x)}{\max_{y \in P_b} B(x)} \cdot \quad (8)$$

They also measure the relative position using

$$\text{RP}(x_i) = \frac{x_i - \min_{y \in P} x_i}{\max_{y \in P} x_i - \min_{y \in P} x_i} \quad (9)$$
1.3.5.4 Percentiles

Percentiles are the most widely used statistic [88]. Percentiles can be manually or automatically set by observing the distribution to find the best discriminant value for differentiating malignant vs. healthy tissue [88].

1.3.5.5 Statistical moments

Statistical-moments are basic statistics such as mean, standard deviation, kurtosis and skewness. These features are included in many studies.

1.3.5.6 Histogram-based

Liu et al. introduced four types of histogram-based features [148]. The first corresponds to the histogram of the signal intensity. The second is the Histogram of Oriented Gradient (HOG) [148]. The HOG uses the distribution of gradient directions to describe local shape. The third type is shape context which is a way to describe the shape of an object [149]. The last type use the Fourier transform of the histogram created via Local Binary Pattern (LBP) [150]. Kwak et al. [97] use LBP by generating a binary pattern code

\[
LBP_{p,R} = \sum_{p=0}^{P-1} s(G_p - G_c)2^p, \quad (10)
\]

where \(c\) is a given center pixel, \(s(x)\) is 1 if \(x \geq 0\) and 0 if \(x < 0\), \(g_c\) and \(g_p\) represent the gray level of the centre pixel and its neighbourhood pixels \(p=0,…,P-1\), and \(R\) is the radius of the neighbourhood.
1.3.6 Feature selection

Feature selection is used in CAD models because exhaustive search of high-dimensional feature spaces is infeasible due to computational limitations. There have been several feature selection methods used in CAD systems. Feature selection techniques can be categorized into feature selection and feature extraction. There are three main types of feature selection categories: filters, wrappers and embedded methods [151].

1.3.6.1 Filters

Filters type methods select variables using criteria that are independent of the model. They use general features such as correlation and mutual information. Filters are fast but they ignore feature dependencies. Niaf et al. [107] use a filter method that uses a two-sample t test to reduce the feature space. Mutual information was introduced by [107] to select a subset of features. Maximum relevance and minimum redundancy were introduced by Peng et al. [112], as another approach to find independent features.

1.3.6.2 Principal component analysis

Principal component analysis (PCA) is a statistical technique used to reduce dimensionality [152]. PCA computes the covariance matrix for the full data set to reduce the data while preserving relevant information. The input image, X, can be transformed using a set of p-dimensional vectors of weights

\[ w_{(k)} = (w_1, ..., w_p)(k) \] (11)

where p is the number of columns. The first component can be computed by
\[ w_{(1)} = \arg \max \left\{ \frac{w^T X^T X w}{w^T w} \right\}. \quad (12) \]

Further components can be found by subtracting the first \( k-1 \) principal components from \( X \) and then finding the loading vector which extracts the maximum variance from the new data matrix

\[ w_{(k)} = \arg \max \left\{ \frac{w^T \tilde{X}_k^T \tilde{X}_k w}{w^T w} \right\}. \quad (13) \]

Multiple linear regression methods such as ridge regression, canonical correlation analysis and principal component regression are also used in CAD systems.

### 1.3.6.3 Laplacian Eigenmaps

Laplacian Eigenmaps is a non-linear method that builds a graph from neighbourhood information. These eigenmaps try to find a low-dimensional space in which the proximity of the data should be preserved from the high-dimensional space.

### 1.3.6.4 Local linear embedding

Local linear embedding (LLE) is a widely used technique that attempts to discover nonlinear structure in high dimensional data. LLE maps its inputs into a lower dimensional single global coordinate system. Tiwari et al. [9] used a bagging approach with multiple neighbourhood sizes for detection of PCa.
1.3.7 Classifier Performance

There are several machine learning algorithms used in CAD systems. See Table 1.1 for an overview of CAD system performance and methodologies.

Figure 1-6: SVM (left), and KNN (right) decision boundaries.

1.3.7.1 Rule-based methods

The first group of CAD systems use rule-based methods with labels of cancerous or benign to classify malignant vs. healthy tissue [7]. With unsupervised classification, clustering methods have been used to classify tissue without using labels [107].

1.3.7.2 Linear methods

Linear model classifiers such as linear discriminant analysis (LDA) optimize linear separation between two classes [153]. A popular choice among studies is the use a SVM. A SVM is desirable because it maximizes the margin between classes, improving classification performance (Figure 1.6 left). A separating boundary takes the form

\[ \{ x \in \chi | w^T x + b = 0 \} \quad (14) \]

where given a training dataset of n samples \( \{(x_1, y_1, ..., (x_n, y_n))\} \subset X \times Y \), where X is the feature space and y is the two-class labelling, Y = {-1, +1}. To achieve a separation boundary an optimization problem is created by combining a minimum norm objective function and classification constraints. The optimization problem is used to classify a test
vector by computing $\text{sign}(f(x))$ where $f(x)$ represents the signed distance to the margin
and can be expressed as

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i x_i^T x + b. \ (15)$$

1.3.7.3 Non-linear methods

Non-linear models are commonly used to improve accuracy at the cost of additional time and complexity. K-nearest neighbours (KNN) is a non-parametric method that uses the votes of its neighbours for classification, where the classification is done by choosing the result as the majority of the neighbours (Figure 1.6 right). Quadratic discriminant analysis is non-linear and can be used instead of LDA [154]. SVMs can be nonlinear by using a nonlinear mapping function to map the data to a higher dimensional space, where the data is linearly separable. Probabilistic classifiers have also been investigated as a nonlinear method of classification. Probabilistic classifiers include the naïve Bayes model, which is commonly used [106] [155].

1.3.7.4 Artificial neural networks

An artificial neural network (ANN) is a network inspired by biological neural networks. ANNs are an interconnected group of nodes with an input, hidden and output layer [131]. Probabilistic neural networks are another type of network that can be modelled by changing the activation function of the hidden layer. Probabilistic feed-forward neural-networks have also been investigated [111].

1.3.7.5 Ensemble classifiers

Ensembles combine many weak classifiers that perform slightly better than random to produce one strong classifier. Ensemble learning classifiers include AdaBoost, random forest and probabilistic boosting tree [6], [156], [117]. Graphical model
classifiers such as Markov random fields have been used for classification and to segment lesions in mpMRI [102], [157].

1.4 Thesis hypothesis and objectives

To take a meaningful step toward addressing the unmet clinical need to provide accurate and personalized assessment, the main objective of this thesis is to create a computer-aided system for classification of malignancy and Gleason score using a highly accurate imaging-histology registration method to generate a reference standard. The central hypothesis of the thesis is that a CAD system that uses a highly accurate reference standard for training can provide accurate and reliable classification and diagnosis of localized prostate cancer. This system will use SVM and KNN machine learning algorithms to classify regions using texture features derived from mpMRI. These algorithms were chosen based on their compatibility with smaller sample sizes, ease of interpretation and performance among relevant literature for PCa classification. The two primary objectives of this thesis are to:

1) Develop and validate a method for classification of malignant, benign and confounding tissue on mpMRI using first and second order texture features with a SVM and KNN machine learning algorithm.

2) Develop and validate a method for classification of high and low Gleason grade tumours on mpMRI using first and second order texture features with a SVM and KNN machine learning algorithm.
References


72. van Dorsten, F.A., M. van der Graaf, M.R. Engelbrecht, G.J. van Leenders, A. Verhofstad, M. Rijpkema, J.J. de la Rosette, J.O. Barentsz, and A. Heerschap,


2 Chapter 2.

Computer-Assisted Characterization of Malignancy and
Gleason Grade of Prostate Cancer on Multi-Parametric MRI

The contents of this chapter are in preparation for submission to the Medical Physics
journal: D. Soetemans, G. S. Bauman, E. Gibson, M. Gaed, J. A. Gomez, M. Moussa, J.
L. Chin, S. Pautler, A. D. Ward. Characterization of Malignancy and Gleason Grade of
Prostate Cancer on Multi-Parametric MRI.

2.1 Introduction

Prostate cancer (PCa) is one of the most prevalent cancers among men [1].
Treatment selection depends on risk stratification, which in turn requires an estimate of
tumour stage and grade. In the hands of highly-skilled readers, multi-parametric magnetic
resonance imaging (mpMRI) has shown promise for detection of PCa, [2-6] but
substantial inter-observer variability in lesion contouring has been observed [7].
Improving lesion classification performance may lead to a reduction in overtreatment of
PCa [8] [9, 10] by improving differentiation between indolent and aggressive tumours [11].

Computer-assisted diagnosis (CAD) models have been investigated to improve
classification of malignancy and Gleason grade. These models use a radiomics-based
approach, which refers to using a comprehensive set of quantitative image features to
determine clinical parameters of interest based on imaging features [12]. This approach
may enable more personalized treatment for patients, as radiomics have been investigated
as a method to determine a relationship between imaging appearance and tumour
phenotypes [13]. Clinical evaluation of a CAD model has shown that differentiation between benign and aggressive PCa improves when combining the CAD prediction with a radiologist’s Prostate Imaging – Reporting and Data System (PIRADS) score, especially among novice radiologists [14].

Many models have been validated using histology obtained from biopsies that are preferentially guided toward suspicious regions identified on mpMRI, [15], [16], [17], [18], [19]. Such validation may be impacted by sparse and spatially-biased sampling and biopsy localization uncertainty in image registration techniques. Spatial biases may be the result of false-positive low-intensity benign regions, as low-intensity regions are indicative of prostate cancer (although not uniquely so) on T2-weighted (T2W) MRI and apparent diffusion coefficient (ADC) maps derived from diffusion-weighted MRI, by the PIRADS guidelines [2]. Cancers are also often under-graded based on transrectal ultrasound (TRUS)-guided biopsies, [20] with the rate of discrepancy between the Gleason grade at biopsy and the true grade at prostatectomy estimated to be 40% [21].

Many recent models use post-radical prostatectomy (RP) histologically verified lesions mapped to mpMRI by radiologist-pathologist consensus [22], [6]. This consensus mapping approach is most straightforward when identifying true-positive mpMRI lesions corresponding to histologic foci, but identifying mpMRI false-positive and false-negative lesions is more challenging and the attendant uncertainty in the error in spatial mapping remains. There have been several reports of previous work suggesting that CAD models may be promising for PCa classification on mpMRI. CAD models have achieved an area under the receiver operator curve (AUC) of 0.97 for the differentiation of PCa from benign tissue [23] [24]. Many recent studies have focused on a more difficult
classification task involving differentiation of prostate cancer from confounding abnormalities such as radiological false positives, prostatic inter-epithelial neoplasia (PIN), atrophy, inflammation, and benign prostatic hyperplasia [25], [22], [26]. However, these studies have limitations which have made widespread clinical implementation of computer-assisted PCa lesion classification elusive, including small sample sizes and registration errors in reference standards due to the aforementioned spatial biases and localization errors.

*Critically needed is a model that is validated against a histologic reference standard that is densely sampled in an unbiased fashion.* We address this using our technique for highly accurate fusion of multi-parametric MRI (mpMRI) with whole-mount digitized histology of the surgical specimen [27]. This study investigates the use of supervised machine learning for 1) classification of malignant and benign tissue, 2) classification of malignant and confounding tissue (i.e. false positive regions identified by radiologists), and 3) classification of high (Gleason grade > 3) and low (Gleason grade 3) tumours, using first and second order texture features extracted from mpMRI.

### 2.2 Materials and methods

**Patient selection and image acquisition**

As part of a clinical trial we obtained prostate specimens from 22 subjects who underwent radical prostatectomy with the following inclusion criteria: (1) male, (2) age 18 years or older and (3) clinical prostate cancer stage T1 or T2 with histological confirmation from biopsy. The exclusion criteria were (1) prior therapy for prostate cancer, (2) use of 5-alpha reductase inhibitors within 6 months of the study start, (3)
inability to comply with pre-operative imaging, (4) allergy to contrast agents, (5) sickle cell or other anemias, (6) hip prosthesis, (7) sources of artifact within the pelvis and (8) contraindications to MR imaging. The research was approved by our institutional human subjects research ethics board and informed consent was obtained from each subject.

**Imaging**

Prostate mpMRI were obtained 19 ± 8 weeks (mean ± SD) after biopsy and 2 ± 1 weeks (mean ± SD) before surgery using a Discovery MR750 (GE Healthcare, Waukesha, WI, USA) at 3 Tesla with an endorectal coil (Prostate eCoil; Medrad, Warrendale, PA). Clinical T2W protocol used repetition time (TR) 4000-13000 ms, echo time (TE) 156-164 ms, bandwidth 31.25 kHz, 2 averages, field of view (FOV) 14 cm x 14 cm x 6.2 cm, slice thickness 3.3-3.6 mm, slice spacing 3.3-3.6 mm, 128 x 256 matrix size, 20-43 slices, flip angle 90°. Clinical DWI protocol used repetition time (TR) 4000 ms, echo time (TE) 70-77 ms, bandwidth 125 kHz, 3 averages, field of view (FOV) 14 cm x 14 cm x 6.2 cm, slice thickness 2.2 mm, slice spacing 2.2 mm, 320 x 192 matrix size, 40 slices, flip angle 90°. After prostatectomy, two sets of strand-shaped fiducials were added, running from base to apex [7]. Internal fiducials were made from cotton thread infused with a 1:40 solution of gadopentetate dimeglumine and blue Tissue Marking Dye (Triangle Biomedical Sciences, Durham, NC). Surface-mounted fiducials were made using animal tissue infused with a 1:40 solution of gadopentetate dimeglumine and 10% buffered formalin. This process resulted in seven surface-mounted fiducials and three internal fiducials [7]. The specimen was then immersed in Christo-Lube (Lubrication Technology Inc., Franklin Furnace, OH, USA) to provide a black background and minimize boundary artifacts on imaging. Ex vivo imaging used a T1-weighted protocol
3D spoiled gradient recalled sequence, TR 6.5 ms, TE 2.5 ms, bandwidth ± 31.25 kHz, eight averages, FOV 14 cm × 14 cm × 6.2 cm, slice thickness 0.4 mm, 256 × 192 matrix, 312 slices, flip angle 15°, duration 25 min] and a T2-weighted protocol (3D fast spin echo sequence, TR 2000 ms, TE 151.5 ms, bandwidth ± 125 kHz, three averages, FOV 14 cm × 14 cm × 6.2 cm, slice thickness 0.4 mm, 320 × 192 matrix, 312 slices, duration 25 min). Histology images and contours were registered to mpMRI by registering the histology to ex-vivo 3D MRI using 10 fiducial markers [28, 29] followed by semi-automatic registration of the ex-vivo MRI scan and the MRI scans to a high resolution in vivo 3D T2W MRI. The histology to ex-vivo registration was performed using an affine transformation utilizing the spatial information provided by the 10 fiducial markers [7]. The ex vivo-to-in vivo MRI registration was performed using a thin-plate-spline warp technique aligning anatomic landmarks [7]. Landmarks were identified along the prostate border, peripheral zone, urethra and verumontanum, as well as corresponding benign prostatic hyperplasia nodules and cystic spaces to warp each digital histopathologic image to its corresponding in vivo MR imaging plane. Average 3D registration error, measured using intrinsic landmarks, was ≤ 2 mm [7].

**Image pre-processing**

Spatial T2W signal variation due to endorectal coil sensitivity was mitigated using the N4ITK image bias correction algorithm [30] implemented in Slicer 4.2 (Surgical Planning Lab, Harvard Medical School, Boston, USA). Inter-patient differences in T2W MR image intensity were accounted for by scaling the intensities such that the mean signal in manually segmented periprostatic fat regions, as indicated by a graduate student, matched an arbitrary reference value across patients.
Pathologist Contouring

After standard whole-mount paraffin embedding, a 4-μm-thick section was cut from each midgland tissue block and stained with hematoxylin and eosin. The resulting slides were digitized on a ScanScope GL (Aperio Technologies, Vista, CA, USA) bright field slide scanning system with a 0.5 μm pixel size. A pathologist contoured and graded all cancerous regions, as well as a prostatic intraepithelial neoplasia (PIN). The images and contours were then downsamped to 30 μm/pixel. The Gleason grade is used to measure the aggressiveness of cancer [31].

Radiologist Contouring

Four observers read the mpMRI scans independently and blinded to histopathology. These observers included one radiology resident and three radiologists, with 5, 6, 2.5 and 2.5 years of prostate MRI assessment experience, respectively. Their performance is illustrated in Table 2.2. Observers used the PIRADS guidelines to identify equivocally, likely or highly likely cancerous lesions (i.e. PIRADS scores 3-5) on mpMRI and then delineated them on the unregistered T2W and ADC images separately.
Table 2.2: Radiologist true positives and false positives at different PIRADS thresholds. R1-R4 indicates the four radiologists.

<table>
<thead>
<tr>
<th>Radiologist</th>
<th>ADC TPs</th>
<th>ADC FPs</th>
<th>T2W TPs</th>
<th>T2W FPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1 PIRADS 5</td>
<td>11</td>
<td>4</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>R1 PIRADS ≥ 4</td>
<td>22</td>
<td>24</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>R1 PIRADS ≥ 3</td>
<td>25</td>
<td>35</td>
<td>19</td>
<td>32</td>
</tr>
<tr>
<td>R2 PIRADS 5</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>26</td>
</tr>
<tr>
<td>R2 PIRADS ≥ 4</td>
<td>31</td>
<td>31</td>
<td>33</td>
<td>38</td>
</tr>
<tr>
<td>R2 PIRADS ≥ 3</td>
<td>36</td>
<td>65</td>
<td>36</td>
<td>52</td>
</tr>
<tr>
<td>R3 PIRADS 5</td>
<td>11</td>
<td>11</td>
<td>14</td>
<td>20</td>
</tr>
<tr>
<td>R3 PIRADS ≥ 4</td>
<td>17</td>
<td>27</td>
<td>21</td>
<td>33</td>
</tr>
<tr>
<td>R3 PIRADS ≥ 3</td>
<td>25</td>
<td>39</td>
<td>28</td>
<td>50</td>
</tr>
<tr>
<td>R4 PIRADS 5</td>
<td>2</td>
<td>2</td>
<td>13</td>
<td>22</td>
</tr>
<tr>
<td>R4 PIRADS ≥ 4</td>
<td>22</td>
<td>32</td>
<td>14</td>
<td>36</td>
</tr>
<tr>
<td>R4 PIRADS ≥ 3</td>
<td>22</td>
<td>34</td>
<td>15</td>
<td>38</td>
</tr>
</tbody>
</table>

Zonal Segmentation

The peripheral zone (PZ) and central gland (CG) were delineated manually by a graduate student using the T2W MR images. Separating the prostate into the PZ and CG has been previously shown to increase classification accuracy [3] [32].

Tumour Size Filtering

Some of the tumour regions on histology were separated by a small amount, creating small islands of cancer foci. To connect these islands, the histology contours were first rasterized to binary images with pixel size matching that of the mpMRI (0.2743 mm/pixel). Next, a disk-shaped structuring element with a radius of 2 pixels (0.55 mm) was used to expand all the tumour regions by morphological dilation on the mpMRI images. The islands were then eroded using the same structuring element. Analysis was performed on all remaining tumours having a longest diameter of ≥ 1.5 mm.
Feature extraction

We calculated 22 first order statistical features to analyze image intensity. See Table 2.3 for the full list of texture features, where \( g(i,j) \) refers to the GLCM, \( H_X \) and \( H_Y \) are entropies of \( g_x \) and \( g_y \), \( H_{XY} = - \sum_{i} g_{ij} \log_2 g_{ij} \), \( H_{X}= - \sum_{ij} g_{ij} \log_2 \{ g_x(i) g_y (j) \} \), \( H_{Y} = - \sum_{i} \sum_{j} g_x(i) g_y (j) \log_2 \{ g_x(i) g_y (j) \} \). \( \mu \) refers to the mean, \( \bar{x} \) refers to the median, \( \sigma \) refers to the standard deviation, and \( N \) refers to the number of pixels. The GLCM neighbourhood voxel distance was 0.2743 \( \mu m \) (1 voxel), and texture features were averaged over all four directions.

Figure 2.6: Illustration of extracted ADC texture features showing: a) energy, b) entropy, correlation and contrast calculated using a square window with a length and width of 2mm centered on each voxel.
### Table 2.3: First and second order texture features used in this study.

<table>
<thead>
<tr>
<th>First-Order Features</th>
<th>Second-Order Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ( \frac{1}{N} \sum_i x_i )</td>
<td>Variance ( \frac{1}{N-1} \sum (x_i - \mu)^2 )</td>
</tr>
<tr>
<td>Standard Deviation ( \sqrt{\frac{\sum (x_i - \mu)^2}{N-1}} )</td>
<td>Entropy ( -\sum_i IP \log_2 IP )</td>
</tr>
<tr>
<td>Energy ( \sum_i x_i^2 )</td>
<td>Uniformity ( \sum_i IP^2 )</td>
</tr>
<tr>
<td>Median</td>
<td>1 Percentile</td>
</tr>
<tr>
<td>Maximum</td>
<td>10 Percentile</td>
</tr>
<tr>
<td>Minimum</td>
<td>90 Percentile</td>
</tr>
<tr>
<td>Range</td>
<td>99 Percentile</td>
</tr>
<tr>
<td>Mean Absolute Deviation ( \frac{1}{N} \sum_i</td>
<td>x_i - \mu</td>
</tr>
<tr>
<td>Root Mean Square ( \sqrt{\frac{1}{N} \sum_i x_i^2} )</td>
<td>Variance of Gradient ( \sum_{i,j} (i-\mu_i + (j-\mu_j))^4 g(i,j) )</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>Bimodality Coefficient</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
GLCM features can characterize anatomical structures, homogeneity, and grey-level transitions. During feature selection, training set positive or negative samples were duplicated to ensure the ratio of positive to negative samples was 1:1. This is done to discourage the classifier from choosing a decision boundary with an unbalanced FPR and FNR. Features were rescaled to range from 0-1.

**Feature selection**

Each region of interest in our dataset is characterized by an 88-dimensional point in texture feature space. Therefore, exhaustive search of the feature space to find the optimal feature subset is not feasible due to computational limitations. A forward feature selection method was used with 10 repetitions of randomized cross validation to sequentially add the feature that increased classification accuracy the most, using the same classifier as used in subsequent the machine learning step.

**Machine Learning**

A support vector machine (SVM) classifier was used for each experiment. A SVM maps vectors from the input space into a high dimensional feature space. A kernel function is used to determine the mapping. It finds a linear decision rule in the form of an
optimal separating boundary that has the widest margin between the decision boundary and the input vector. A decision boundary takes the form

\[ \{ x \in \chi | w^T x + b = 0 \} \]  

(16)

where given a training dataset of n samples \( \{(x, y_1, \ldots, x_n, y_n)\} \subseteq \chi \times Y \), and where \( \chi \) is the feature space and \( Y \) is the class labelling, \( Y = \{-1, +1\} \). An objective function takes the form

\[ \min_{w \in \mathbb{R}^d, b, \beta_i \in \mathbb{R}} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \beta_i, \]  

(17)

which is constrained by

\[ y_i(w^T x_i + b) \geq 1 - \beta_i, \quad i = 1, \ldots, n \]  

(18)

\[ 0 \leq \beta_i, \quad i = 1, \ldots, n \]

where \( C \) is the cost coefficient that weights the classification error and \( \beta_i \) correspond to the distance to the margin for possibly misclassified samples \( x_i \). This optimization problem can be expressed as

\[ \max_{\alpha_i \in \mathbb{R}} -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^{n} \alpha_i, \]  

(19)

where \( \alpha_i \) are the Lagrange multiplier coefficients and which is constrained by

\[ \sum_{i=1}^{n} \alpha_i y_i = 0. \]  

(20)

\[ 0 \leq \alpha_i \leq C \quad i = 1, \ldots, n \]

A test vector is then classified by computing \( \text{sign}(f(x)) \), where \( f(x) \) represents the signed distance to the margin and is expressed as

\[ f(x) = \sum_{i=1}^{n} \alpha_i y_i x_i^T x + b, \]  

(21)

where the sign of \( f(x) \) is used to determine the classification and the distance to the margin is used as the classifier’s confidence. All results with the SVM were obtained
using the PRTools library [33]. A support vector machine maximizes a margin for classification.

KNN is a non-parametric method that uses the votes of its neighbours for classification, where the classification is done by choosing the result as the majority of the neighbours. Default parameters were used for each classifier. For the SVM the default kernel was linear. K is the number of neighbours voting and is optimized with respect to the leave-one-out error on the training dataset, which is performed during each fold of cross validation.

2.3 Results

This section presents the performance of the system for the SVM and KNN algorithms. For each experiment, 4-fold cross validation was used to measure the AUC, FPR and FNR.

Classification of Tumours and Benign Tissue

Our first experiment uses a design similar to several studies that classify malignant and benign tissue [34, 35]. Table 2.4 shows the results of the experiment. In total, 69 tumours and 44 benign tissue regions in the PZ, and 23 tumours and 22 benign tissue regions in the CG were used. Benign tissue regions were created by selecting a region from the contralateral side of the prostate when possible for each prostate slice. The imbalance of tumours to benign tissue is due to prostate slices where there was little or no benign tissue region available for a benign sample. Note that we were able to achieve an AUC of 0.96 in the PZ using a SVM, which is competitive with the highest reported among the
literature. Also note that performance in the CG was lower than the PZ, which agrees with evidence presented in relevant literature [3] [32].

Table 2.4: Classification of malignant and benign tissue.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Zone</th>
<th>AUC</th>
<th>FPR</th>
<th>FNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>PZ</td>
<td>0.96</td>
<td>0.05</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>0.75</td>
<td>0.24</td>
<td>0.43</td>
</tr>
<tr>
<td>KNN</td>
<td>PZ</td>
<td>0.94</td>
<td>0.13</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>0.75</td>
<td>0.23</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Classification of Tumours and Radiologist False Positives

The second experiment focused on classification of malignant and confounding tissue, in the form of radiologist false positives. The results are illustrated in Table 2.5. Post-RP histologically verified tumours were used as positive samples and radiologist false positives were used as negative samples. Sixty-nine tumours and 102 radiologist false positive regions in the PZ and 23 tumours and 52 radiologist false positive regions in the CG were used in this experiment. We were able to achieve an AUC of 0.89 and 0.88 in the PZ using the SVM and KNN algorithms, respectively. These accuracies are comparable to other studies that use a similar setup to classify malignant tissue and confounding pathologies such as prostatic intraepithelial neoplasia, atrophy, benign prostatic hyperplasia and inflammation.
Table 2.5: Classification of malignant and confounding tissue in the form of radiologist false positives.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Zone</th>
<th>AUC</th>
<th>FPR</th>
<th>FNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>PZ</td>
<td>0.89</td>
<td>0.10</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>0.84</td>
<td>0.21</td>
<td>0.23</td>
</tr>
<tr>
<td>KNN</td>
<td>PZ</td>
<td>0.88</td>
<td>0.05</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>CG</td>
<td>0.71</td>
<td>0.22</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Classification of High and Low Gleason Grade

Our third experiment involved characterizing tumours with a Gleason grade of 3 and tumours with a Gleason grade >= 3+4. Table 2.6 illustrates the results for this experiment. 29 malignant regions had a Gleason grade >= 3+4 and 49 malignant regions had a Gleason grade of 3. We were unable to achieve high accuracies for this experiment, highlighted by an AUC of just 0.48 for the KNN in the PZ. Other studies have reported success using the ADC maps to classify high and low Gleason scores [36] [35].

Table 2.6: Classification of high and low Gleason grade.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Zone</th>
<th>AUC</th>
<th>FPR</th>
<th>FNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>PZ</td>
<td>0.60</td>
<td>0.44</td>
<td>0.49</td>
</tr>
<tr>
<td>KNN</td>
<td>PZ</td>
<td>0.48</td>
<td>0.35</td>
<td>0.77</td>
</tr>
</tbody>
</table>
2.4 Discussion

Our results demonstrate that the CAD system can classify malignancy of PCa with an AUC accuracy of 0.96. Comparing our system to previous systems suggests that a highly-accurate fusion of histology with mpMRI is important for classification accuracy, as this system enables clean samples for both malignant and benign regions of interest for model training. Using our system for highly-accurate registration we were able to achieve an AUC of 0.96 using a SVM algorithm for classification of tumours and benign tissue in the PZ. This performance is competitive with the top performing models that have an AUC of 0.97. Our results for classification in the PZ and CG agree with current published literature which indicates that classification in the CG is more difficult than the PZ. [32]

Many recent studies have focused on classification of tumours and confounding structures. Our results agree with the current literature which has found that this is a much more difficult task compared to classification of malignant and benign tissue. Litjens et al. [22] achieved AUCs of 0.73, 0.75, 0.63, and 0.69 for differentiation of tumours vs. PIN, atrophy, inflammation and BPH, respectively. Some studies also look at classifying tumours vs. radiologist false positives [24] [36]. Vos et al. were able to achieve an AUC of 0.82. Using our approach for classification of malignant and confounding tissue in the form of radiologist false positives, we were able to achieve an AUC of 0.89 in the PZ using the KNN algorithm. To our knowledge, this is the highest reported accuracy in the literature for this type of classification. Table 2.2 illustrates the false negatives and false positives of the radiologists for contouring PCa at different PIRADS thresholds. Radiologist performance was moderate, suggesting the model may improve classification accuracy when used as a secondary observer. Accuracy results for
classification of high and low Gleason grade were lower than seen in the literature. We suspect this is due to two main reasons. The first is that our sample size of high Gleason score tumours was small as there were very few prostates in our dataset with a Gleason score > 3+4. The other reason we suspect our accuracy was lower is due to spatial biases. Unlike many studies, our dataset was not retrospectively annotated after analyzing histology results. Looking at our radiologist annotations on the ADC maps shows spatial bias in the form of incorrectly extending tumour boundaries to include hypo-intense benign tissue in some cases. This may explain why our accuracy for high and low grade was lower than other studies in the literature.

Our study has some limitations. First, our study consisted of patients who were treated with radical prostatectomy, inducing a selection bias. Although prostatectomy is reserved for patients with aggressive PCa, our data included a large number of Gleason score 6 regions and few with a Gleason score >= 4. This may explain why our performance for detecting aggressive prostate cancer was low. Second, a small sample size of 22 patients was used. Further validation in a larger dataset with more high grade cancer is required for advances in accuracy.

2.5 Conclusion

We have developed a CAD system for classification of suspicious prostate tissue. The systems performance for classification of malignant and benign tissue in the PZ is competitive with the highest reported among the literature. Our system also achieved a high accuracy rate when characterizing malignant and confounding tissue in the form of radiologist false positives. This research supports the development of a clinical CAD
system in the form of a radiological tool for improved diagnosis and patient treatment planning.

There are several potential usage scenarios for our system. The first would be to use the system as a second observer for radiologists during screening. This could reduce the workload of radiologists as multiple opinions are often given during screening. This system could also help with patient risk stratification by suggesting whether a suspicious area is high or low grade cancer. This allows for more informed physician decisions and may allow for reduced overtreatment.
References


3 Chapter 3.

Discussion and conclusions

The final chapter of this thesis reexamines the research objectives and provides the key findings and conclusions. This chapter also addresses limitations of the current study and the field, and the potential solutions for both. Finally, future directions for PCa detection and diagnosis are presented.

3.1 Contributions of the thesis

This thesis contributes advances in methods, concepts and knowledge in several key areas.

Methods: CAD systems can be difficult to compare due to differences in datasets and design. Many of the top performing models focus on classification of tumours vs. healthy tissue ([1] [2]), however, a more clinically relevant CAD system would classify tumours vs. confounding regions. Some recent studies have investigated classification of these confounding pathologies, including radiologist false positives, prostatic inter-epithelial neoplasia (PIN), atrophy, inflammation, and benign prostatic hyperplasia [3], [4], [5]. Our study measured the performance for both classification of benign tissue and confounding tissue in order to enable a more complete breakdown when comparing models. The approach presented in this thesis is the first ever to classify malignant and confounding tissue using first and second order texture features derives from mpMRI, validated using a highly accurate registration algorithm for fusion of histology to mpMRI.
3.2 Validation

Previous CAD systems used radiological consensus for registration of histology to mpMRI for validation. The consensus technique can lead to spatial biases in the form of false-positive low-intensity benign regions, as low-intensity regions are indicative of prostate cancer on T2W and ADC MRI. Our system is the first to use a densely-sampled unbiased reference standard derived from a registration of histology images to MR images with < 2 mm target registration error [6]. Rather than relying on biases in radiologist consensus mapping, this reference standard enables us to generate high-quality unbiased contours for our samples when training our model.

Knowledge: Using a SVM for classification of malignant and benign tissue in the PZ we were able to achieve an AUC of 0.96. This is competitive with the highest reported AUC among the literature. Using a SVM for classification of malignant tissue and confounding tissue in the form of radiologist false positives we were able to achieve an AUC of 0.89. To our knowledge, this is the highest reported accuracy among the literature. Our system had poor performance for classification of high and low Gleason grade. Although other models have been shown to have moderate success for this type of classification, our results suggest there is a significant amount of room for improvement for this type of classification.

3.3 Limitations

Our study has several limitations. First, our study consisted of patients who were treated with radical prostatectomy, inducing a selection bias. In a clinical setting, patients would have a range of outcomes from completely healthy to having an advanced form of prostate cancer. Our study focused only on patients who were scheduled for radical
prostatectomy, ensuring at least one tumour was present per patient. However, even with this selection bias, many of our patients had a low Gleason grade cancer. This distribution of aggressiveness plays into our favour, as many of the patients we analyzed could have been candidates for a focal therapy. A second limitation of our study is a small sample size of 22 patients that was used. Further validation in a larger dataset is required because prostate cancer will have a different appearance based on a number of factors such as zone of origin, geographical location, age, and genomic makeup. In order to properly classify each type of tumour, two approaches may be appropriate. The first approach to this problem would be to build a global dataset with a large number of samples that is available to validate models with. However, this approach may be incredibly expensive and time consuming due to all of the preparation work and collaboration that would be required. A second approach to fix a small sample size would be to investigate ways of extracting and registering information from previously processed prostates. There are large repositories of prostates and prostate cancer that could be tapped into in order to expand the size of the dataset. A major limitation inherent to all CAD systems is the lack of an understanding of what causes the expression of texture features. Lalonde et al. have looked at analyzing a genomic classifier in order to understand which genes can be used to predict biochemical recurrence and metastasis [7]. An understanding of these mechanisms is critical in order for CAD system to proceed to clinic, as it would eliminate the black box approach currently taken. This in turn may increase the confidence and optimism of physicians and radiologist to use such a system.
3.4 Future Directions

The methods developed in this thesis support many research initiatives in several directions. This section addresses remaining gaps in knowledge that exist and potential applications of this system on future directions.

3.4.1 External validation

CAD systems have been shown to improve radiologist performance with detection and diagnosis. However, prior to clinical validation, multiple large independent datasets must be used to validate such tools. External validation is important because it ensures models are not overfitted to their data. In most previous publications, CAD systems were validated using cross validation. Although this is seen an appropriate approach, a more rigorous approach would be to have a training set, a testing set and validation set. A validation set is not common with CAD models because it requires further division of the samples which are usually very limited. Such validation may provide additional information to improve assessment and provide personalized treatment plans for patients.

3.4.2 Incorporation of additional knowledge

This thesis has suggested that an accurate histology to MR registration system for detection and diagnosis of PCa may improve accuracy. There remains room for improvement for both PCa detection and diagnosis. Our models have focused exclusively on using imaging features to classify tumours and suspicious regions. Incorporating patient characteristics such as age, family history, PSA level, etc. may improve
performance and better inform physicians for treatment planning, although this presents various challenges. These challenges include maintaining patient confidentiality, obtaining the data, and quantifying the data. Radiomic features are quantitative and continuous, but clinical data can be qualitative or discontinuous. Data can take the form of binary decisions (i.e., does the patient have a history of PCa) or a scale (i.e., a rating of 1-5 stars). Data in this form may influence classifiers to behave in an unpredictable fashion, particularly when the sample size is low, as each additional training sample could maximally move the decision boundary in one direction. Characteristics such as patient genomics and family history also may be inaccessible or challenging to obtain.

3.4.3 Decision Support

The impact of using a CAD system to aid a physician with detection and diagnosis may be in the ability to provide more personalized treatment plans for patients. Litjens et al. have shown that when a CAD system is used as a secondary observer, radiologist classification performance improves, especially among novice radiologists [8]. Completely automatic CAD systems have also been investigated [9] [10]. Various approaches can be taken to generate contours around suspicious regions. Giannini et al. created a fully automatic CAD system for peripheral zone lesions that used a voxel-wise malignancy probability map to create regions by keeping pixels with a probability higher than 60% to be malignant. Litjens et al. have used a Hessian blob detector approach to generate initial regions of interest [9]. Additional research on automatically generating regions of interest that are suspicious for PCa may further improve the performance of these systems.
3.2 Conclusions

Although CAD models have made significant improvements, tough challenges still remain.

Working in a high dimensional feature space requires the use of feature selection techniques in order to determine the optimal feature set. There are several feature selection methods used in the literature, such as forward feature selection, reverse feature selection and branch and bound feature selection. There are several reasons different techniques are employed including time taken and complexity of the method. Further research is required in order to compare and contrast the different techniques to determine their optimal use.

Several machine learning classifiers have been used to characterize and classify tumours. Each of these classifiers have advantages and disadvantages associated with them. For example, a support vector machine can be advantageous for classification as it maximizes the margin between two classes of points. This improves accuracy but is much slower, and thus is only beneficial with a small number of samples.

Confounding regions have overlapping appearance to prostate cancer and can mask the appearance of cancer. Texture feature extraction may allow for better assessment of cancer from confounding regions to assess the subtle differences. Studies have looked at characterizing tumours vs. radiological false positives, prostatic inter-epithelial neoplasia (PIN), atrophy, inflammation, and benign prostatic hyperplasia [3], [4], [5]. Further investigation of classification of confounding regions may improve overall performance.

In conclusion, we created a CAD system for classification of malignant, benign
and confounding tissue as well as low and high Gleason grade, validated using our method for highly accurate fusion of mpMRI with whole-mount digitized histology of the surgical specimen. Our system is competitive with the highest reported accuracy among the literature for classification of malignant and benign tissue. Our system exceeded reported accuracies for classification of malignant and confounding tissue in the form of radiologist false positives. In order to further improve CAD systems, classification of confounding regions should be a primary area of focus.
References


DEREK J. SOETEMANS
CURRICULUM VITAE

EDUCATION

M.E.Sc. in Biomedical Engineering (in progress)
The University of Western Ontario
Thesis: Computer-assisted characterization of prostate cancer on MRI

Supervisor: Dr. Aaron Ward
Baines Centre, Victoria Hospital, London, Ontario

B.E.Sc. in Electrical Engineering with Professional Internship (2009-2014)
The University of Western Ontario
Internship: Facilities engineering intern at General Motors CAMI Assembly

AWARDS

2016 IMNO 2nd place poster award - $200
2014 Computer-Assisted Medical Intervention NSERC CREATE Program - $12000/ann
2014 NSERC USRA - $4500/month for 4 months
2013 General Motors “Challenge the Status Quo” Award
2011 IEEE Programming Competition Winner
2010 International Plowing Match Scholarship - $1200
2009, 2013 Dean Honour List
2009, Queen Elizabeth II Aiming for the Top Scholarship - $2500
2009, Brooke Telecom Scholarship - $1000
2009, East Lambton Secondary School Foundation Scholarship - $2500
2009, Colin Hood award of Excellence
2009, UWO Continuing Admission Scholarship - $2500/annum

Publications

Poster presentations
